



Quantities vs. prices: monopoly regulation without transfer under asymmetric demand information

Dan Wang¹ · Peng Hao² · Jiancheng Wang³

Received: 17 October 2022 / Accepted: 5 February 2023 / Published online: 5 April 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

This paper examines the optimal monopoly regulation without transfer based on Basso, Figueroa and Vásquez (Rand J Econ 48(3):557–578, 2017), which compare the quantity-based and price-based instruments to regulate a monopoly that has better information concerning its market demand than the regulator. The optimal screening mechanisms, which offer multiple menus of contracts for the regulated firm to select, and pooling mechanisms, which only provide a uniform contract, are characterized for each instrument. Furthermore, the corresponding performances of the regulator's social welfare are ranked. Results show that, with non-increasing marginal costs of the regulated firm, the screening price mechanism would strictly dominate the screening quantity mechanism. The pooling price mechanism is always preferred to the pooling quantity mechanism when the slope of marginal costs is negative or slightly positive. Otherwise, the pooling quantity mechanism may be superior depending on the relative magnitude of the slope of marginal costs and demand function.

Keywords Optimal delegation · Quantities versus prices · Monopoly regulation · Asymmetric demand information · Infeasible transfer

✉ Peng Hao
haoopeng@ahu.edu.cn

Dan Wang
dwang919@ccnu.edu.cn

Jiancheng Wang
wangjc1987@foxmail.com

¹ School of Information Management, Central China Normal University, Wuhan 430079, China

² School of Business, Anhui University, Hefei 230039, China

³ School of Economics, Wuhan University of Technology, Wuhan 430072, China

1 Introduction

The optimal design of a regulatory mechanism for a monopolist, such as firms that provide power, natural gas, water, and other energy, differs from one domain to another, not only because of the possibility of private information but also the feasibility and efficiency of transfers. On one hand, the theoretical literature (Baron and Myerson 1982; Laffont and Tirole 1986, 1993; Amador and Bagwell 2022) on monopoly regulation typically assumes that the regulated firm probably has private information about its costs. The study in a setting where the regulated firm has the information advantage on its demand, however, is also important and relevant to explore. As Riordan (1984) and Lewis and Sappington (1988a, 1988b) and Basso et al. (2017) argue, a regulated firm is likely to have more information about the demand for its product than the regulator, not only because the firm generally has better knowledge concerning its product's quality and reliability, but also because they usually devote significant resources to acquire the knowledge of its markets that is hardly shared with the regulator. On the other hand, monetary transfers to the regulated firm are used by the regulator to establish the contracts of the product in most of the literature (Lewis and Sappington 1988a, 1988b; Laffont and Tirole 1993; Aguirre and Beitia 2004; Basso et al. 2017). Nevertheless, there are circumstances in which the monetary transfer instruments may be limited. Some scholars (Baron 1989; Laffont and Tirole 1993; Armstrong and Sappington 2007) argue that the regulators often do not have the authority to explicitly employ public funds to compensate the regulated firm or tax the regulated firm. For example, the "Regulations on Electric Power Supervision" promulgated by the State Council of the People's Republic of China clearly determined that the competent pricing department and the power regulator shall supervise the price of electricity by laws, administrative regulations and the provisions of the State Council. The electric power regulator shall, in accordance with the law, supervise the fair and non-discriminatory opening of the electricity market to entities engaged in electricity trading and the fair opening of the power grid by transmission enterprises. However, the regulatory authority does not have the ability to explicitly tax or pay subsidies. Also, even though the monetary transfer can be realized through positive or negative access fees, the scope for such transfers may be limited by universal service or consumers' arbitrage behaviors in practice.

In a seminal paper, Basso et al. (2017) compare price and quantity instruments to regulate a monopoly that has private information about its demand or costs. They contribute to studying an alternative option of quantity-based mechanisms for monopoly regulation. The monetary transfers, however, are used by the regulator to the regulated firm in their paper. Based on the model of Basso et al. (2017), we study the optimal monopoly regulation under the restriction that direct monetary transfers are infeasible. We characterize the optimal screening mechanisms that provide multiple menus for the regulated firm to select and pooling mechanisms that only offer one unique contract for quantity and price instruments, and compare performances of the regulator's social welfare. The pooling

mechanisms deserve discussion for two reasons. First, the pooling mechanisms would be preferred when screening the demand information of a regulated firm is not valuable for the regulator. Second, the pooling mechanisms are easier to understand and implement for regulators in practice even though less welfare would be induced.

When the regulator is allowed to offer multiple menus for the regulated firm to self-select and the costs of the regulated firm are restricted to be linear, the quantity-floor regulation emerges as the optimal quantity regulatory policy. Sufficient conditions are provided for the optimality of this quantity-floor mechanism among all feasible delegation mechanisms. We show that this quantity-floor mechanism would be optimal when the relative concavity parameter of the regulator and utility's welfare is large and the density of the demand information is non-increasing. This optimal quantity mechanism would never attain the first-best level, while the quantity mechanism in Basso et al. (2017) can attain the first-best level when the consumers' welfare and the firm's utility are equally important for the regulator. The optimal quantity delegation results from the following trade-off: the firm can tailor its flexible policy to the demand information, but the implemented quantity is too low for the regulator. Hence, the first-best level that depends on the firm's type cannot be attained, since there is no way to utilize all information without any bias. For price regulation, when marginal costs are constant, the first-best price level that equals fixed marginal cost, however, can be achieved due to the fact that it is not dependent on the firm's private information. When marginal costs are decreasing, the optimal screening quantity depends on the floor. The screening price involves bunching. When the realized demand is not the lowest type, the screening price mechanism creates fewer losses caused by underproduction, which leads to the preference for screening price mechanism. Therefore, we find that, with the linear costs of the regulated firm, the screening price mechanism would strictly dominate the screening quantity mechanism.

When the regulator is restricted to pooling mechanisms, both price and quantity regulation switch from menus to fixed values. Compared with the transfer-based case in Basso et al. (2017), the firm's individual rationality constraint may not bind in our paper. With linear and decreasing marginal costs, the optimal quantity and price are both optimally set to leave the lowest firm's demand type with zero profits due to the binding individual rationality constraint. We find that the pooling price mechanism is always preferred to the pooling quantity mechanism. For non-decreasing marginal costs, there are two cases in consideration and the equal weights to consumers' surplus and regulated firm's profits for the regulator are assumed for clear and convenient analysis. First, when the slope of marginal costs is slightly positive or no larger than a positive threshold, the optimal quantity is still derived by the binding individual rationality constraint for the lowest type. The optimal price, however, is optimally chosen when the price equals marginal cost with expected firm's type since the firm's individual rationality constraint does not bind. Hence, it shows that the optimal pooling price mechanism is still superior in this circumstance. Second, when the slope of marginal costs is large exceeding a positive threshold, we demonstrate that the individual rationality constraints in quantity and price regulation both do not bind, and thus, the optimal pooling quantity and price are both

determined based on the intersection of the demand with expected firm's type and marginal costs. In this case, we show the pooling quantity will dominate if the slope of marginal costs is great than the absolute value of the demand slope. Therefore, the pooling quantity mechanism may be suitable for public monopoly cases, such as transportation, electricity production and transmission, etc., where the demand is rather inelastic, and marginal cost is increasing because of the high volumes of traffic or electricity. High volumes exceeding the capacity would result in system congestion. In a real-world setting, the level of departure and arrival flights for each time at the airport is determined to be four slots in Chicago O'Hare and Washington National, which provides evidence that the regulator uses the pooling quantity mechanism rather than the price mechanism.

Some articles have respectively focused on the issues of asymmetric cost or demand information and infeasible transfers. For example, for the literature on asymmetric cost information, Baron and Myerson (1982) first apply mechanism design to monopoly service regulation, considering how regulators should regulate a monopoly with private cost information. In optimal regulatory policies, prices and monetary transfers as subsidies are designed as functions of firm cost reporting, thus maximizing expected social welfare under the constraint that firms have non-negative profits and no incentives to distort their costs. For the literature on asymmetric demand information, Riordan (1984) and Lewis and Sappington (1988a) first focus on transfer-based price monopoly regulation when there is asymmetric information on demand. Compared with Riordan's (1984) case of the constant marginal cost of production, Lewis and Sappington (1988a) focus on the case where marginal cost of production varies with output. Lewis and Sappington (1988b) study the optimal price regulatory policies when regulators do not fully observe both the cost function and the demand function of the firm. The transfer payments are not feasible in some circumstances, which is a key determinant in designing the optimal regulatory policy. For the literature on infeasible transfers, Alonso and Matouschek (2008) first regard the monopoly regulation problem as an optimal delegation problem, which is first introduced by Holmström (1984), when the costs are private observed by the regulated firm and the transfer instrument is infeasible, while the participation constraint is not included in their analysis. Then, Amador and Bagwell (2022) analyze the monopoly regulation with asymmetric cost information under the restriction that transfers are infeasible by extending the Lagrangian approach to delegation problems of Amador and Bagwell (2013) to include ex-post participation constraint. It is worth mentioning that the price-cap regulation and the quantity-floor regulation are the same with asymmetric cost information in the research of Amador and Bagwell (2022) since the realized quantity induced by the optimal price is equivalent to the optimal quantity. As Basso et al. (2017) state, the quantity and price regulations are not equivalent to asymmetric demand information, although these two instruments lead to the same level of welfare with asymmetric cost information.

The pooling mechanisms of our paper are also related to the literature (Williams 2002; Moledina et al. 2003; Krysiak 2008; Krysiak and Oberauner 2010) on the comparison of price and quantity instruments under asymmetric demand or cost information that is first conducted by Weitzman (1974). The key differences in our paper are twofold: first, the firms in these papers are price takers and choose a quantity of production based on their cost functions, disregarding the demand, while the firm in our paper

is a monopolist who has the market power and is faced with the whole demand, and therefore the quantity induced for the price will not be offered by profit maximization but the demand; second, the direct transfers in these papers are feasible. Therefore, the results in these papers cannot directly extend to our paper. The significant differences between several key studies are summarized in Table 1.

The rest of the paper is organized as follows. Section 2 provides the basic theoretical model. Section 3 studies the optimal screening monopoly regulation mechanism in quantity and price instruments and the comparative welfare analysis is derived. In Sect. 4, the optimal pooling quantity and price mechanisms are characterized and compared. Section 5 briefly summarizes the results and suggests opportunities for further work. Formal proofs are in the Appendix.

2 The basic model

We consider the relationship between a regulator and a regulated monopoly firm, such as firms that provide power, natural gas, water resources, and other energy. This model framework has a broad appeal and can be applied to various monopoly regulations, such as electricity, telecommunication, transportation regulation, etc. The specific regulatory environment under consideration is the following.

2.1 Cost, demand and information

The regulated firm's cost is known to be given by the function

$$C(q) = c_0q + c_1q^2/2,$$

with marginal cost $C'(q) = c_0 + c_1q > 0$, which can be increasing, decreasing and constant with the output. We assume that any fixed cost is normalized to zero for ease of presentation and without loss of generality. It is also common knowledge that the unit price p and quantity q follow the inverse demand function

$$p(q, \theta) = P(q) + \theta,$$

where the parameter θ captures the firm's private information about the demand, and can be regarded as the regulated firm's type. The regulator is not informed about θ when contracting but knows that θ is the realization of a random variable on $\Theta \in [\underline{\theta}, \bar{\theta}]$ where $\bar{\theta} > \underline{\theta} > 0$ with a cumulative distribution function $F(\theta)$, a density $f(\theta)$ and the monotone hazard rate $d((1 - F(\theta))/f(\theta))/d\theta \leq 0$ for most distributions, such as uniform distribution, normal distribution, logarithmic distribution, exponential distribution, and Laplace distribution. It assumes a linear demand function which means that $p_q(q, \theta) = P'(q) < 0 < p_\theta(q, \theta) = 1$, $P''(q) = 0$ where subscripts denote partial derivatives. For every price $p \geq 0$ and type θ , the direct demand function $Q(p, \theta)$ obeys $p \equiv P(Q(p, \theta)) + \theta$. Following the derivatives with respect to p and θ , we find that $Q_p(p, \theta) = 1/P'(q) < 0 < Q_\theta(p, \theta) = -1/P'(q)$, that is higher the realizations of θ imply a larger quantity of output consumers demand at any nonnegative price. It

Table 1 Significant differences between key studies

	Baron and Myerson (1982)	Riordan (1984)	Lewis and Sappington (1988a)	Lewis and Sappington (1988b)	Amador and Bagwell (2022)	Basso et al. (2017)	Weitzman (1974)	This paper
Asymmetry cost information	•			•	•			
Asymmetry demand information		•	•	•		•		•
Feasible transfers	•	•	•	•		•		
Infeasible transfers					•			•
Price regulation	•	•	•	•	•	•	•	•
Quantity regulation								
Price-maker firm	•	•	•	•	•	•	•	•
Price-taker firm								
Constant marginal cost	•	•			•			
Variable marginal cost			•			•		•

assumes that $C''(q) = c_1 > P'(q)$ to ensure a concave optimization problem and non-negative results when the firm's optimal policy is executed. In words, the demand is steeper than marginal costs when the latter is downward sloping.

2.2 Contracts

For screening mechanisms, the regulator designs an effective incentive regulation to foster the firm to tell the truth about its demand type. We assume that the regulator has no access to monetary transfer payments, only restrictions on the price or quantity can be imposed. Therefore, following Holmström's (1984) delegation mechanism, an assignment menu $x(\theta)$ that can be a quantity $q(\theta)$ or price $p(\theta)$ can be stipulated by the regulator. The menu emphasizes the role that the regulator plays in limiting ex ante the possible options chosen by the firm. It is no loss of generality in restricting the analysis to such direct communication and truthful mechanism from the Revelation Principle. For pooling regulatory mechanisms without considering the incentive of a firm's truth-telling, a unique quantity q^u or price p^u for all firm's demand types is stipulated.

2.3 Objectives

It is straightforward to write the respective firm's profits for quantity and price regulation as

$$u(q, \theta) = (P(q) + \theta - c_0 - c_1 q / 2)q \text{ and } u(p, \theta) = (p - c_0 - c_1 Q(p, \theta) / 2)Q(p, \theta).$$

To ensure it is a concave optimization problem, we assume $u_{qq}(q, \theta) < 0$ which means $c_1 > 2P'(q)$ and $u_{pp}(p, \theta) < 0$ which means $c_1 Q_p(p, \theta) < 2$.

The consumer surplus for quantity and price regulation is given by

$$v(q, \theta) = \int_0^q (P(\tilde{q}) + \theta)d\tilde{q} - (P(q) + \theta)q \text{ and } v(p, \theta) = \int_p^\infty Q(\tilde{p}, \theta)d\tilde{p}.$$

The regulator's objective is to maximize a weighted sum of consumer surplus and the firm's profit in which the latter receives a weight $\alpha \in [0, 1]$. When $\alpha < 1$, the consumer interest is a greater concern for the regulator. The social welfare function for quantity and price regulation can be written as

$$E_\theta w(q, \theta) = E_\theta \left[\int_0^q (P(\tilde{q}) + \theta)d\tilde{q} - (P(q) + \theta)q + \alpha(P(q) + \theta - c_0 - c_1 q / 2)q \right] \text{ and } E_\theta w(p, \theta) = E_\theta \left[\int_p^\infty Q(\tilde{p}, \theta)d\tilde{p} + \alpha(p - c_0 - c_1 Q(p, \theta) / 2)Q(p, \theta) \right].$$

To ensure it is a concave optimization problem, we also assume $w_{qq}(q, \theta) < 0$ which means $c_1 > (2 - 1/\alpha)P'(q)$ and $w_{pp}(p, \theta) < 0$ which means $c_1 Q_p(p, \theta) < 2 - 1/\alpha$.

3 Screening quantity and price regulatory mechanisms

In this section, we characterize the optimal screening monopoly incentive regulation in both quantities and prices, and make a comparative analysis of welfare.

3.1 Screening quantity regulation

For quantity regulation, we first define the first-best quantity mechanism under complete information as a benchmark. Under the individual rational (IR) constraint, the regulator’s problem can be written as

$$\begin{aligned} \max_q \int_0^q (P(\tilde{q}) + \theta)d\tilde{q} - (P(q) + \theta)q + \alpha(P(q) + \theta - c_0 - c_1q/2)q \\ \text{s.t.} (P(q) + \theta - c_0 - c_1q/2)q > 0. \end{aligned} \tag{1}$$

Note that the value of an outside option is normalized to be zero. To figure out this problem, the first step is to define the optimal quantity when the IR constraint is ignored. Then, check whether the optimal quantity satisfies the IR constraint. The regulator’s optimal quantity without considering the IR constraint.

From the first-order condition of the objective function, we find that

$$P(q^o) + \theta - c_0 - c_1q^o/2 = [c_1/2 - (1/\alpha - 1)]P'(q^o)q^o.$$

We assume $c_1 \leq -2(1/\alpha - 1)P'(q)$ for simplicity. Thus, the IR constraint will bind, since the quantity $q^o(\theta)$ violates the firm’s IR constraint except when $\alpha = 1, c_1 = 0$. We can deduce that the first-best quantity would be $q^*(\theta)$ that depends on the firm’s type θ , such that $P(q^*) + \theta - c_0 - c_1q^*/2 = 0$. We find that $q^{*'}(\theta) = 2/(c_1 - 2P'(q)) > 0$.

We now consider the case of incomplete information on demand. First, we need to identify the firm’s quantity flexible allocation. Specifically, if the firm is flexible in choosing the quantity and unrestricted by the regulator, the regulated firm’s optimal quantity policy will be obtained as

$$q^f(\theta) = \arg \max_q (P(q) + \theta - c_0 - c_1q/2)q.$$

The first-order condition is given by $P'(q^f(\theta))q^f(\theta) + \theta - c_0 + P(q^f(\theta)) - c_1q^f(\theta) = 0$. The quantity flexible allocation is the firm’s quantity as a function of its type. Following the derivatives with respect to θ , we find that $q^{f'}(\theta) = -1/u_{qq}(q, \theta) > 0$. Notice as well that $P(q^f(\theta)) + \theta - c_0 - c_1q^f(\theta)/2 = (c_1/2 - P'(q^f(\theta)))q^f(\theta) > 0$ and thus $u(q_f(\theta), \theta) > 0$ for all $\theta \in \Theta$. Therefore, we can utilize this first-order condition in the regulator’s objective and obtain the direction of the bias of the firm with type θ :

$$w_q(q^f(\theta), \theta) = -P'(q^f(\theta))q^f(\theta) > 0 \text{ for all } \theta \in \Theta.$$

Therefore, the firm would prefer lower output from the perspective of the regulator. To limit this downward bias and ensure the truth-telling of firm, the regulator would stipulate a menu of quantities $\{q(\hat{\theta})\}_{\hat{\theta} \in \Theta}$ contingent on the firm’s type announcement. The incentive compatibility (IC) constraint is given as follows

$$\theta = \arg \max_{\hat{\theta}} (P(q(\hat{\theta})) + \theta - c_0 - c_1 q(\hat{\theta})/2) q(\hat{\theta}).$$

Standard revealed preferences arguments (Samuelson 1938) indicate that $q(\theta)$ is weakly increasing in demand information θ and almost everywhere differentiable. The first-order condition for truth-telling can be obtained as

$$q'(\theta)[P'(q(\theta))q(\theta) + \theta - c_0 - c_1 q(\theta) + P(q(\theta))] = 0. \tag{2}$$

From this, the quantity $q(\theta)$ is either the firm’s quantity flexible policy or independent of the information on demand. The regulator may set a floor on the permissible output satisfying this IR constraint. Thus, we can get the regulator’s optimization problem as:

$$\max_{q(\theta)} E \left[\int_0^{q(\theta)} (P(\bar{q}) + \theta) d\bar{q} - (P(q(\theta)) + \theta)q(\theta) + \alpha(P(q(\theta)) + \theta - c_0 - c_1 q(\theta)/2)q(\theta) \right] \text{ s.t. (1) and (2).}$$

To figure out this problem, we need to define an optimal quantity-floor allocation as Lemma 1 when the IR constraint is ignored, which is similar to Alonso and Matouschek (2008) with cost asymmetric information. The proof appears in the Appendix.

Lemma 1 *Ignoring the individual rationality constraint, the optimal quantity-floor allocation is*

$$q^c(\theta) = \begin{cases} q^f(\theta_c) & ; \theta \in [\underline{\theta}, \theta_c] \\ q^f(\theta) & ; \theta \in (\theta_c, \bar{\theta}] \end{cases} \text{ where the cut-off type } \theta_c \text{ satisfies}$$

$$\int_{\underline{\theta}}^{\theta_c} w_q(q^f(\theta_c), \theta) dF(\theta) = 0.$$

The cut-off condition implies that the average bias among the pooled types is zero. Note that $q^f(\theta) < q^*(\theta) \leq q^o(\theta)$. Then, we need to verify whether the participation constraint holds. Under this optimal quantity-floor allocation, the IR constraint can be written as $u^c(\theta) = (P(q^c(\theta)) + \theta - c_0 - c_1 q^c(\theta)/2)q^c(\theta) \geq 0$ for all $\theta \in \Theta$. Notice that

$$u^{c'}(\theta) = \begin{cases} q^f(\theta_c) > 0 & ; \theta \in [\underline{\theta}, \theta_c] \\ q^f(\theta) > 0 & ; \theta \in (\theta_c, \bar{\theta}] \end{cases}.$$

Therefore, it suffices to check whether the IR constraint for the lowest demand type is satisfied, due to the strictly increasing property of the firm’s profits with the firm’s type. Then, we get the Lemma 2 which is similar to Amador and Bagwell (2022) with cost asymmetric information. See the proof in the appendix.

Lemma 2 *The individual rationality constraint is not satisfied under the optimal quantity-floor allocation.*

To satisfy the IR constraint, a sufficient flexible allocation is chosen so that all types would produce non-negative profit. There would exist a cut-off demand type $\theta_r \in (\underline{\theta}, \theta_c)$ such that the IR constraint for the lowest demand type binds (see Fig. 1). The result is summarized in the following proposition 1, which is also similar to Amador and Bagwell (2022) with cost asymmetric information.

Proposition 1 *The optimal IR-quantity-floor mechanism obeys:*

$$q_r(\theta) = \begin{cases} q^f(\theta_r) & ; \theta \in [\underline{\theta}, \theta_r] \\ q^f(\theta) & ; \theta \in (\theta_r, \bar{\theta}] \end{cases}$$

where the cut-off type θ_r satisfies

$$u(q^f(\theta_r), \underline{\theta}) = (P(q^f(\theta_r)) + \underline{\theta} - c_0 - c_1 q^f(\theta_r) / 2) q^f(\theta_r) = 0. \text{ Note that}$$

if $\theta_r \geq \bar{\theta}$, $q_r(\theta) = q^f(\theta_r)$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$; if $\theta_r \leq \underline{\theta}$, $q_r(\theta) = q^f(\theta)$, for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

Next, we proceed to determine the sufficient conditions for this IR-quantity-floor mechanism to be optimal. As the theory of optimal delegation (Holmström, 1984; Alonso and Matouschek 2008; Amador and Bagwell 2013) states, if the interval delegation is to be optimal among all feasible delegation, it must be superior for the regulator to the alternative delegation with jump discontinuities, such as the alternative allocation in Fig. 2. We adopt the “guess-and-verify” approach of Amador and Bagwell (2013) in which extended Lagrangian methods is used to verify the optimality of the guessed solution. For simplicity, we take the case $c_1 = 0$ as an example. See more details on the research of Anton and Andriy (2019), and focus on the comparison of the quantity and price instruments.

The relative concavity parameter of the regulator and firm’s objective functions, which is the key to determining the optimality of the IR-quantity-floor solution, is defined as follows:

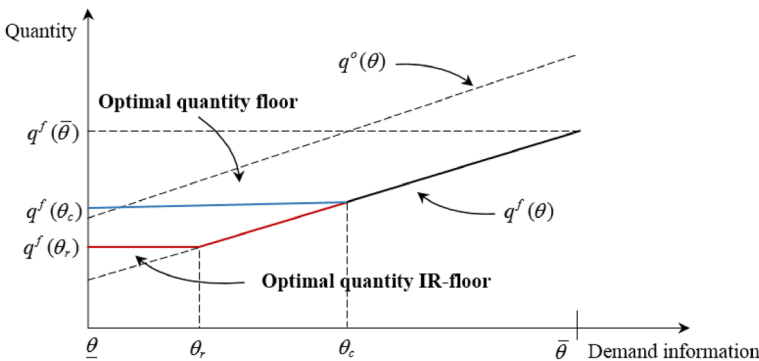


Fig. 1 Optimal quantity-floor mechanism where for all $\theta \in [\underline{\theta}, \theta_c]$ the blue line is the optimal quantity-floor allocation without considering the individual rationality constraint and the red line is the optimal IR-quantity-floor allocation

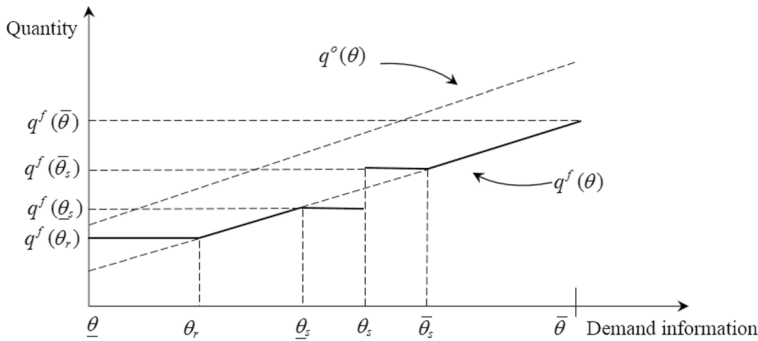


Fig. 2 The alternative allocation with one jump discontinuity in which a “step” at θ_s , that is indifferent between $q^f(\theta_{s-})$ and $q^f(\theta_s)$, is induced

$$\kappa = \min_{q, \theta} \left\{ \frac{w_{qq}(q, \theta)}{u_{qq}(q, \theta)} \right\}.$$

The resulting sufficient conditions can be summarized as follows. The proof can be seen in the appendix.

Proposition 2 *If the following conditions are satisfied, the optimal solution to the regulator’s problem will be realized with the IR-quantity-floor mechanism:*

- (i) $\frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})f(\tilde{\theta})d\tilde{\theta} \geq 0,$
- (ii) $\frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \hat{\theta})f(\hat{\theta})d\hat{\theta} - \kappa F(\theta) \leq \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})f(\tilde{\theta})d\tilde{\theta}$
for all $\theta \in [\underline{\theta}, \theta_r],$ and
- (iii) $\kappa F(\theta) - w_q(q^f(\theta), \theta)f(\theta)$ is non – decreasing for all $\theta \in [\theta_r, \bar{\theta}].$

We now check the property of sufficient conditions. The left side of condition (i) shows the average quantity bias among the pooled firm which reflects the first-order condition for the optimal cut-off type. Without considering the IR constraint, the average quantity bias is equal to zero which is defined in Lemma 1. When the focus is turned to the optimal quantity allocation satisfying the IR constraint, the average quantity bias is restricted to be non-negative, which shows a lower cut-off value is feasible through the concavity of the problem. The sufficient condition (ii) and (iii) would be more easily satisfied when the relative concavity parameter of the regulator and firm’s objective functions κ is large, since the distribution $F(\theta)$ is non-negative and increasing for all $\theta \in [\underline{\theta}, \bar{\theta}].$ We can also observe that the condition (iii)

is more easily satisfied when the density $f(\theta)$ is non-increasing, because $w_q(q_f(\theta), \theta)$ is positive.

To provide some intuition to interpret why the relative concavity of the regulator and firm’s objective functions and the density slope are critical, the alternative allocation with one jump discontinuity in Fig. 2 is considered. The alternative allocation has both good side and bad side as compared with the optimal IR-quantity-floor allocation. In terms of the discussion of the density slope, the higher quantity allocation for all $\theta \in [\underline{\theta}_s, \bar{\theta}_s]$ closer to the regulator’s preferred efficient quantity and lower quantity allocation for all $\theta \in [\underline{\theta}_s, \theta_s]$ with a larger bias are induced in the alternative allocation. To make the optimal IR-quantity-floor allocation superior, the non-increasing density can induce the effect of bad side dominates the effect of good side in the regulator’s expected social welfare. Then, for the discussion of the relative concavity of the regulator and firm’s objective functions, we find that the variance of the allocation around $q^f(\theta)$ for $\theta \in [\underline{\theta}_s, \bar{\theta}_s]$ is increased in alternative allocation as compared with the optimal IR-quantity-floor allocation. When the regulator’s objective is more concave than the firm’s objective in which the relative concavity is large, the regulator would not benefit from the increase in variance afforded by the alternative allocation. Thus, the IR-quantity-floor allocation remains superior. Building on the above intuitive explanations, the following corollary offers conditions on the density and the relative concavity of the objective functions for Proposition 2. The proof can be seen in the appendix.

Corollary 1 *If $f'(\theta) \leq 0$ and $\kappa \geq \frac{1}{2}\alpha$ for all $\theta \in \Theta$, the sufficient conditions in Proposition 2 will be satisfied, then the optimal IR-quantity-floor allocation will be the optimal solution to the regulator’s problem.*

To this end, a numerical example is offered for a clear explanation. Consider a linear demand function $P(q) = 10 - q$ with $q \in [0, 10]$, a linear cost function $C(q) = c_0q$ with $c_0 < 10$ to ensure $q_f(\theta) = (10 - c_0 + \theta)/2 > 0$ and $\theta \in [1, 5]$. For this numerical example, $\kappa = \alpha - 1/2$. $w_{qq}(q, \theta) < 0$ and $w_{pp}(p, \theta) < 0$ mean $\alpha > 1/2$. Following corollary 1, the sufficient conditions in Proposition 2 will be satisfied when $\kappa \geq 1/2\alpha$, which means only when $\alpha = 1$. Therefore, the quantity-floor mechanism, which obeys that

$$\text{if } c_0 > 7, q_r(\theta) = \begin{cases} 11 - c_0 & ; \theta \in [1, 12 - c_0] \\ (10 - c_0 + \theta)/2 & ; \theta \in (12 - c_0, 5] \end{cases}$$

and if $c_0 \leq 7$, $q_r(\theta) = 11 - c_0$, for all $\theta \in [1, 5]$,

can be optimal when the regulator gives equal weight to the consumers’ and firm’s utility. Otherwise, the alternative allocation with jump discontinuities will be optimal.

3.2 Screening price regulation

For price regulation, we also first define the first-best price mechanism under complete information. Under the IR constraint, the regulator’s problem can be written as

$$\begin{aligned} & \max_{p(\theta)} \int_p^\infty Q(\tilde{p}, \theta) d\tilde{p} + \alpha(p - c_0 - c_1 Q(p, \theta) / 2) Q(p, \theta) \\ & \text{s.t. } u(p, \theta) = (p - c_0 - c_1 Q(p, \theta) / 2) Q(p, \theta) \geq 0. \end{aligned} \tag{3}$$

Similar to the quantity regulation, we consider the optimal price when the IR constraint is ignored as follows:

$$p^o(\theta) = \arg \max_p \int_p^\infty Q(\tilde{p}, \theta) d\tilde{p} + \alpha(p - c_0 - c_1 Q(p, \theta) / 2) Q(p, \theta).$$

From the first-order condition and the assumption $c_1 \leq -2(1/\alpha - 1)P'(q) = -2(1/\alpha - 1)/Q_p(p, \theta)$, we find that the price $p^o(\theta)$ violates the firm’s IR constraint except when the weight $\alpha = 1, c_1 = 0$, since

$$p^o(\theta) - c_0 - c_1 Q(p^o(\theta), \theta) / 2 = ((1/\alpha - 1)/Q_p(p^o(\theta), \theta) + 1/2c_1) Q(p^o(\theta), \theta) \leq 0.$$

Thus, through the binding IR constraint, we can deduce that the first-best price would be $p^*(\theta)$ that depends on the firm’s type θ except when $c_1 = 0$, such that $p^*(\theta) - c_0 - c_1 Q(p^*(\theta), \theta) / 2 = 0$. Taking the deviation of θ , we find that the monotonicity of the $p^*(\theta)$ depends on the slope of marginal cost c_1 , such that

$$p^{*'}(\theta) = \frac{c_1 Q_\theta(p^*(\theta), \theta)}{2 - c_1 Q_p(p^*(\theta), \theta)}.$$

Considering the case of incomplete information on demand, we need to identify the firm’s price flexible allocation as well. If the firm is flexible in choosing the price and unrestricted by the regulator, the regulated firm’s optimal price policy will be obtained as

$$p^f(\theta) = \arg \max_p (p - c_0 - c_1 Q(p, \theta) / 2) Q(p, \theta).$$

The first-order condition is given by $[p^f(\theta) - c_0 - c_1 Q(p^f(\theta), \theta) / 2] Q'_p(p^f(\theta), \theta) + (1 - c_1 Q_p(p^f(\theta), \theta) / 2) Q(p^f(\theta), \theta) = 0$. Following the derivatives with respect to θ , we find that $p^{f'}(\theta) = -(1 - c_1 Q_p) / u_{pp} > 0$. Notice as well that $p^f(\theta) - c_0 - c_1 Q(p^f(\theta), \theta) / 2 = -(Q_p(p^f(\theta), \theta) - c_1 / 2) Q(p^f(\theta), \theta) > 0$ and thus $u(p_f(\theta), \theta) > 0$ for all $\theta \in \Theta$. Therefore, we can utilize this first-order condition in the regulator’s objective and obtain the direction of the bias of the firm with type θ :

$$w_p(p^f(\theta), \theta) = -Q(p^f(\theta), \theta) < 0 \text{ for all } \theta \in \Theta.$$

Therefore, the firm would prefer a higher price than the regulator’s first-best price. Compared to quantity instrument, further characterizing the optimal price is not easy. If marginal costs are constant, we can get that the first-best price would be $p^* = c_0$ that depends on the firm’s type θ . Therefore, with incomplete information on demand, the first-best price mechanism is feasible. When marginal costs are decreasing, the optimal price mechanism involves complete bunching, where the

regulator fixes a pooling price independent of the firm’s report. This is because the first-best price mechanism ceases to be incentive compatible. Compared to the complete bunching, using the regulated firm’s private information is not superior. Only if marginal costs are increasing, using the private information may be profitable, as only in that case the price schedule is increasing in θ . However, when $\alpha = 1$, the IR condition may not bind, and the results are unambiguous. For simplification, we focus on the non-increasing marginal cost here. Specifically, the regulator’s optimization problem is:

$$\max_{p(\theta)} E_{\theta} \left[\int_{p(\theta)}^{\infty} Q(\tilde{p}, \theta) d\tilde{p} + \alpha(p(\theta) - c_0 - c_1 Q(p(\theta), \theta) / 2) Q(p(\theta), \theta) \right] \text{ s.t. (3)}$$

The results can be summarized as following. The proof can be seen in the appendix.

Proposition 3 *Assuming non-increasing marginal cost, the optimal screening price p_r*

- (1) *is the first-best price, such that $p^* = c_0$, if marginal costs are constant $c_1 = 0$.*
- (2) *obeys $u(p^u, \underline{\theta}) = (p^u - c_0 - \frac{1}{2}c_1 Q(p^u, \underline{\theta}))Q(p^u, \underline{\theta}) = 0$, if marginal costs are decreasing $c_1 < 0$.*

The result shows that if marginal costs are constant, the first-best price is feasible, since the price is not dependent on the firm’s information. Therefore, leaving flexibility to the firm is not necessary, since the regulator does not need to utilize firm’s private information. if marginal costs are decreasing, the optimal p^u is optimally set to leave the lowest type with zero profits, since p^s violates the IR constraint, and profits are increasing with the firm’s type. Then, it is natural to compare the social welfare of the screening quantity and price mechanism.

Corollary 2 *Assuming that the transfers are not feasible and costs of firm are non-increasing, the price mechanism strictly dominates the quantity mechanism for all weight $\alpha \in [0, 1]$.*

With constant marginal costs, as shown in Proposition 1, the optimal quantity mechanism would never attain the first-best level for all weight $\alpha \in [0, 1]$, while the first-best price mechanism is feasible. Thus, the price mechanism would strictly dominate the quantity mechanism for all $\alpha \in [0, 1]$. The reason is that the information of regulated firm is valuable for the regulator in quantity mechanism, while it is useless in price mechanism. Specifically, for quantity regulation, the first-best quantity depends on the firm’s type, and there exists a downward bias of firm even when the weight $\alpha = 1$. Hence, without using the transfers, the optimal quantity delegation results from the following trade-off: the firm can tailor the flexible quantity policy to its demand information, but the implemented quantity is too low from the perspective of regulator. For price regulation, the first-best price, however, is not dependent on the information. Therefore, leaving flexibility to the firm is not necessary,

since the regulator does not need to utilize firm’s private information. Compared to the result of Basso et al. (2017) who state that the transfer-based price and quantity mechanisms generate the same level of welfare when $\alpha = 1$, we show that the price mechanism still strictly dominates the quantity mechanism when $\alpha = 1$. This is because the transfers are not socially costly anymore when $\alpha = 1$, and they would be the free tools to separate firm’s type. Without using the transfers, the only way to tell the firm’s type is by providing the flexibility of firm to select its preferred price or quantity with any regulator’s restriction. It, however, is costly for the regulator, since a downward bias is left.

With decreasing marginal costs, the optimal screening quantity depends on the threshold θ_r . If $\theta \in (\theta_r, \bar{\theta}]$, the optimal quantity is firm’s flexible policy which is optimally set based on the intersection of $c_0 + (c_1 - P'(q))q$ and the demand of the type θ . If $\theta \in [\underline{\theta}, \theta_r]$, the optimal quantity is the quantity floor which is optimally set based on the intersection of $c_0 + (c_1 - P'(q))q$ and the demand of the threshold type θ_r . The screening price involves bunching, and is optimally set based on the intersection of $c_0 + c_1q/2$ and the demand of the lowest type $\underline{\theta}$. When the realized demand is not the lowest type, both the optimal screening quantity and price mechanisms induce deadweight losses caused by underproduction. Specifically, the first-best quantity of the realized demand would be larger than the realized quantity under the screening quantity and price regulation. More outputs will be induced in price regulation than the outputs in quantity regulation, since the outputs can adjust with the realized demand in price regulation. Thus, the screening price mechanism creates fewer losses, as shown in Fig. 3, which leads to the preference for screening price mechanism.

4 Pooling quantity and price regulatory mechanisms

We now proceed to the pooling quantity and price regulatory mechanisms in which the regulator just utilizes his incomplete information on firm’s demand type to set one unique quantity or price without incentivizing the truth-telling of firm. These pooling mechanisms would be reasonable and possible when a delegation of quantity or price choice is not valuable for the regulator for all $\alpha \in [0, 1]$, which is induced by too

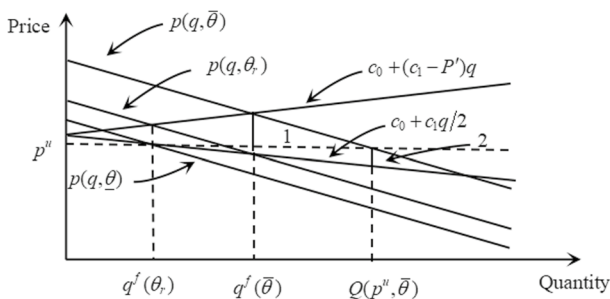


Fig. 3 Deadweight losses of screening quantity and price mechanisms, when decreasing marginal costs are assumed. 2 is the deadweight losses of screening price mechanism, and the sum of 1 and 2 is the deadweight losses of quantity mechanism

large bias of firm. The mechanisms also deserve discussions because they would be easier to be implemented and understood in practice. The problem of further characterizing the optimal pooling quantity and price is also non-trivial, because the IR constraints may not bind in some cases. A linear demand function $P(q) = P(0) + P'(q)q$ is assumed to have a sharper characterization. To ensure non-negative price and quantity and clear results, it assumes that $P(0) + E(\theta) - c_0 > 4(E(\theta) - \underline{\theta})$ and $c_1 \leq \bar{c}_1$ where $\bar{c}_1 = -P'(q)[(P(0) + E(\theta) - c_0)/(E(\theta) - \underline{\theta}) - 1] > 0$. When we consider the case with non-decreasing marginal costs, it is enough to consider the case where the weight is equal to one, in order to make an unambiguous comparison of pooling quantity and price regulation.

First, maximize $w(q(\theta), \theta)$ subject to the IR constraint $u(q(\theta), \theta) \geq 0$ and maximize $w(p(\theta), \theta)$ subject to $u(p(\theta), \theta) \geq 0$. We can define the first-best quantity $q(\theta)$ and price $p(\theta)$ for all types $\theta \in \Theta$ under complete information when marginal costs are linear, such that

$P(q(\theta)) + \theta = c_0 + c_1q(\theta)$ and $p(\theta) = c_0 + c_1Q(p(\theta), \theta)$ if marginal costs are increasing and the weight $\alpha = 1$;

$P(q(\theta)) + \theta = c_0 + \frac{1}{2}c_1(q(\theta))$ and $p(\theta) = c_0 + \frac{1}{2}c_1Q(p(\theta), \theta)$ if marginal costs are decreasing.

Notice that the IR constraints will bind when marginal costs are decreasing, otherwise, the IR constraints will not bind.

4.1 Pooling quantity regulation

Without complete information, for pooling quantity regulation, the regulator’s problem is to choose one single q to maximize $E_\theta w(q, \theta)$ subject to the IR constraint $u(q, \theta) \geq 0$. To solve the problem, we also consider the optimal pooling quantity when the IR constraint is ignored first, as follows:

$$q^s = \arg \max_q E_\theta \left[\int_0^q (P(\tilde{q}) + \theta)d\tilde{q} - (P(q) + \theta)q + \alpha(P(q) + \theta - c_0 - \frac{1}{2}c_1q)q \right].$$

From the first-order condition $-(1 - \alpha)P'(q^s)q^s + \alpha(P(q^s) + E(\theta) - c_0 - c_1q^s) = 0$ and the monotonicity of the firm’s profit with the demand information $u_\theta(q, \theta) = q > 0$, we find that the quantity q^s will violates the firm’s IR constraint if

$$P(q^s) + \underline{\theta} - c_0 - \frac{1}{2}c_1q^s = \left[\frac{1}{2}c_1 - (1 - \frac{1}{\alpha})P'(q^s) \right]q^s - (E(\theta) - \underline{\theta}) < 0.$$

Solving this, we can obtain the following proposition. The proof can be seen in the Appendix.

Proposition 4 *Assuming quadratic costs and linear demand functions, the optimal pooling quantity q^u*

- (1) is q^s in which $P(q^s) + E(\theta) = c_0 + c_1 q^s + \left(\frac{1}{\alpha} - 1\right) P'(q^s) q^s$ and q^s depends on the weight α , if the slope of marginal costs $c_1 \in [c_1^q, \bar{c}_1]$;
- (2) obeys $u(q^u, \underline{\theta}) = q^u \left(P(q^u) + \underline{\theta} - c_0 - \frac{1}{2} c_1 q^u \right) = 0$, in which q^u is independent with α , if the slope of marginal costs $c_1 \in (P'(q), c_1^q)$.

Note that $\bar{c}_1 > c_1^q = 2 \left(1 - \frac{1}{\alpha} - \frac{E(\theta) - \underline{\theta}}{\alpha [P(0) + E(\theta) - c_0 - 2(E(\theta) - \underline{\theta})]} \right) P'(q) > 0$; and if the weight $\alpha = 1$, the optimal quantity is chosen when the price with expected firm's type equals marginal costs such that $P(q^s) + E(\theta) = c_0 - c_1 q^s$, and

$$c_1^q = -P'(q) \frac{2(E(\theta) - \underline{\theta})}{P(0) + E(\theta) - c_0 - 2(E(\theta) - \underline{\theta})}.$$

This proposition shows that, if marginal costs are decreasing or increasing in which the slope is not too large, the binding IR constraint and the increasing profits with the firm's type result that the optimal q^u is optimally set to leave the lowest type with zero profits. Otherwise, the optimal quantity q^s satisfies the IR constraint, and the quantity depends on the weight that the regulator assigns to the firm's profits.

4.2 Pooling price regulation

For price regulation, the regulator's problem is to choose one unique p to maximize $E_\theta w(p, \theta)$ subject to the IR constraint $u(p, \theta) \geq 0$. Characterizing the optimal pooling price p^s when the IR constraint is ignored, we can obtain the price p^s will violates the firm's IR constraint if

$$p^s - c_0 - \frac{1}{2} c_1 Q(p^s, \theta) = \left(1 - \frac{1}{2} c_1 Q_p(p^s, E(\theta)) - \frac{1}{\alpha} \right) \frac{Q(p^s, E(\theta))}{-Q_p(p^s, E(\theta))} - \frac{1}{2} c_1 Q_p(p^s, E(\theta))(E(\theta) - \underline{\theta}) < 0.$$

Whether the optimal pooling price p^s satisfies the IR constraint depends on the value of marginal costs. If the IR constraint binds, we find that the monotone property of the firm's profit with the demand information determines which type is most reluctant to participate and left with zero profits. The solution to this optimization problem can be summarized as following. The proof can be seen in the appendix.

Proposition 5 Assuming quadratic costs and linear demand functions, the optimal pooling price p^u

- (1) is p^s in which $p^s = c_0 + c_1 Q(p^s, E(\theta))$, if marginal costs are non-decreasing $c_1 \in [0, \bar{c}_1]$ and the weight $\alpha = 1$;
- (2) obeys $u(p^u, \underline{\theta}) = \left(p^u - c_0 - \frac{1}{2} c_1 Q(p^u, \underline{\theta}) \right) Q(p^u, \underline{\theta}) = 0$, in which p^u is independent with α , if marginal costs are decreasing $c_1 \in (P'(q), 0)$.

The result shows that if marginal costs are decreasing, the optimal p^u is also optimally set to leave the lowest type with zero profits, since p^s violates the IR constraint for all $\alpha \in [0, 1]$, and profits are increasing with the firm's type. When marginal costs are non-decreasing and $\alpha = 1$, the optimal price p^s , which is based on the intersection of price and marginal costs with expected demand, will satisfy the IR constraint. Then, it is natural to compare the social welfare of the pooling quantity and price mechanism. The proof is in the Appendix.

Corollary 3 *When marginal costs are decreasing, $c_1 < 0$, the optimal pooling price mechanism is preferred to the quantity mechanism for all $\alpha \in [0, 1]$. When marginal costs are non-decreasing and the weight $\alpha = 1$, there are two cases: (1) if $c_1 \in (0, c_1^q)$, the optimal pooling price mechanism also dominates; (2) if $c_1 \in [c_1^q, \bar{c}_1]$, the price mechanism is preferred only when marginal costs $c_1 < -P'(q)$.*

With decreasing marginal costs, for all weight $\alpha \in [0, 1]$, both the pooling quantity and price are optimally set based on the intersection of $c_0 + 1/2c_1q$ and the demand of the lowest type $\underline{\theta}$. When the realized demand is not the lowest type, both the optimal pooling quantity and price mechanisms induce deadweight losses caused by underproduction. Specifically, the first-best quantity of the realized demand would be larger than the realized quantity under the pooling quantity and price regulation. More outputs will be induced in price regulation than the outputs in quantity regulation, since the outputs can adjust with the realized demand in pooling price regulation. Thus, the pooling price mechanism creates fewer losses, as shown in Fig. 4, the top panel, which leads to the preference for pooling price mechanism.

If the slope of marginal costs is slightly positive, $c_1 \in [0, c_1^q)$, the optimal pooling quantity is still based on the intersection of $c_0 + 1/2c_1q$ and the demand of the lowest type $\underline{\theta}$, while the pooling price is optimally chosen when the price equals marginal cost with expected firm's type; see Fig. 4, the middle panel. When the realized demand is larger than expected demand, now, both the optimal pooling quantity and price mechanisms create deadweight losses caused by overproduction. Similarly, the pooling price mechanism is preferred due to a smaller overproduction. When the demand turns out to be smaller than expected demand, the deadweight losses of pooling quantity mechanism is still induced by overproduction, while the losses of pooling price mechanism are caused by underproduction. By the assumption, the slope of marginal costs is definitely lower than the absolute value of demand slope in this case. Therefore, the optimal pooling price mechanism always dominates because of lower deadweight losses.

If the slope of marginal costs is extremely positive, $c_1 \in [c_1^q, \bar{c}_1]$, the analysis is similar to the discussion of Basso et al. (2017) with increasing marginal costs. Both the optimal pooling quantity and price are optimally chosen when the price equals marginal cost with expected firm's type. Now, when the realized demand is lower (resp. higher) than expected demand, the deadweight losses comparison of the quantity mechanism created by overproduction (resp. underproduction) and price mechanism caused by underproduction (resp. overproduction) depends on the slope of marginal costs and demand functions, as shown in the bottom panel of Fig. 4.

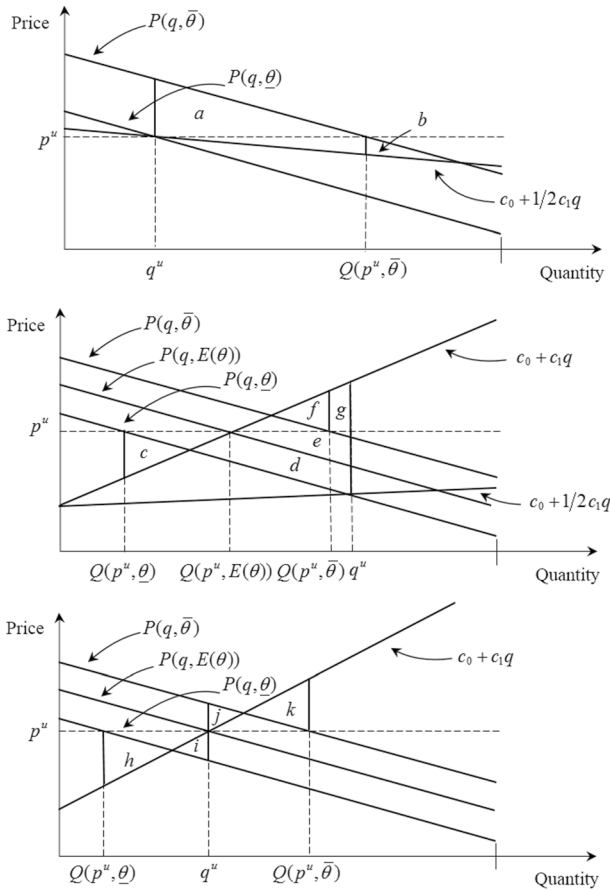


Fig. 4 Deadweight losses of pooling quantity and price mechanisms with the weight $\alpha = 1$. At the top, when decreasing marginal costs are assumed, b is the deadweight losses of pooling price mechanism, and the sum of a and b is the deadweight losses of quantity mechanism. In the middle, the slope of marginal costs is assumed to be slightly positive. When the demand turns out to be $E(\theta)$, the deadweight losses of quantity and price are $e + f + g$ and zero respectively. When the demand turns out to be θ , the deadweight losses of quantity and price are $d + e + f + g$ and c respectively. When the demand turns out to be $\bar{\theta}$, the deadweight losses of quantity and price are $f + g$ and f respectively. At the bottom, the slope of marginal costs is assumed to be heavily positive. The deadweight losses of quantity and price are i and h respectively if the demand θ is realized, and the deadweight losses of quantity and price are j and k respectively if the demand $\bar{\theta}$ is realized

Then, it is natural to rank the social welfare of the screening and pooling mechanisms. We can obtain the following corollary.

Corollary 4 Assuming that the transfers are not feasible and costs of firm are non-increasing, the price pooling mechanism is preferred to the screening and pooling quantity mechanisms for all $\alpha \in [0, 1]$.

5 Conclusion remarks

In practice, direct monetary transfers to the regulated firms are not feasible in some circumstances, since the regulator may have no access to the public funds or inability to tax. Based on the work of Basso et al. (2017) who compare the quantity-based and price-based instruments under asymmetric demand information, this paper examines the optimal design of monopoly regulation under the restriction that monetary transfers are infeasible. Screening mechanisms that provide multiple menus for regulate firm to select and pooling mechanisms that only offer one unique contract are characterized, and the quantity regulation and price regulation performances of regulator’s social welfare are compared.

We find that, with constant marginal costs of the regulated firm, the screening price mechanism would strictly dominate the screening quantity mechanism, since the first-best price is feasible. As compared with Basso et al. (2017), the optimal screening IR-quantity-floor mechanism would never attain the first-best level when the consumers’ welfare and the firm’s utility are equally important to the regulator. With decreasing marginal costs, the optimal screening quantity depends on the floor. The screening price involves bunching. When the realized demand is not the lowest type, the screening price mechanism creates fewer losses caused by underproduction, which leads to the preference for screening price mechanism. With pooling mechanisms, the optimal quantity and price are not always optimally chosen based on the intersection of the price and marginal cost with expected firm’s type, because the firm’s individual rationality constraint may not bind in our paper. We show that the pooling price mechanism is always superior if the slope of marginal costs is negative or slightly positive, otherwise, the ranking of quantity and price instruments would depend on the relative magnitude of the slope of marginal costs and demand functions. Furthermore, we find that the price pooling mechanism is preferred to the screening and pooling quantity mechanisms with firm’s non-increasing marginal costs.

This paper can be extended. In this paper, we normalize the fixed cost is zero. The extension is to study the possibility of shutdown with positive fixed cost, which is kept for further research.

Appendix

Proof of Lemma 1

Under the quantity-floor allocation, the regulator’s welfare objective can be written as

$$E_{\theta}w(\theta_c) = \int_{\underline{\theta}}^{\theta_c} \int_0^{q^f(\theta_c)} (P(\bar{q}) + \theta)d\bar{q} - (P(q^f(\theta_c)) + \theta)q^f(\theta_c) + \alpha(P(q^f(\theta_c)) + \theta - c_0 - c_1q^f(\theta_c)/2)q^f(\theta_c)]dF(\theta) + \int_{\theta_c}^{\bar{\theta}} \int_0^{q^f(\theta)} (P(\bar{q}) + \theta)d\bar{q} - (P(q^f(\theta)) + \theta)q^f(\theta) + \alpha(P(q^f(\theta)) + \theta - c_0 - c_1q^f(\theta)/2)q^f(\theta)]dF(\theta).$$

With the respect of θ_c , the first-order condition of this function can be obtained as

$$q^f(\theta_c) \int_{\underline{\theta}}^{\theta_c} [-(1-\alpha)P'(q^f(\theta_c))q^f(\theta_c) + \alpha(P(q^f(\theta_c)) + \theta - c_0 - c_1q^f(\theta_c)/2)]dF(\theta) = 0,$$

where $q^f(\theta_c) = -1/u_{qq}(q^f(\theta_c), \theta_c) > 0$. Thus, we can obtain $\int_{\underline{\theta}}^{\theta_c} w_q(q^f(\theta_c), \theta)dF(\theta) = 0$ as requested by Lemma 1.

Proof of Lemma 2

Notice that $w_{q\theta}(q^f(\theta_c), \theta) = \alpha > 0$ and thus,

$$w_q(q^f(\theta_c), \underline{\theta}) = -(1-\alpha)P'(q^f(\theta_c))q^f(\theta_c) + \alpha(P(q^f(\theta_c)) + \underline{\theta} - c_0 - c_1q) < 0.$$

We can conclude

$$u(q^f(\theta_c), \underline{\theta}) = (P(q^f(\theta_c)) + \underline{\theta} - c_0 - c_1q/2)q^f(\theta_c) < [(1/\alpha - 1)P'(q^f(\theta_c)) + c_1/2]q^{f^2}(\theta_c) < 0.$$

Proof of Proposition 2

First, by expressing the IC constraint in usual integral form plus a monotonicity requirement, the regulator’s problem can be rewritten as:

$$\max_q \int_{\underline{\theta}}^{\bar{\theta}} w(q(\theta), \theta)dF(\theta)$$

subject to:

$$(P(q(\theta)) + \theta - c_0)q(\theta) = \int_{\underline{\theta}}^{\theta} q(\tilde{\theta})d\tilde{\theta} + \underline{U}, \text{ for all } \theta \in \Theta$$

$$q(\theta) \text{ non-decreasing, for all } \theta \in \Theta$$

$$(P(q(\theta)) + \theta - c_0)q(\theta) \geq 0, \text{ for all } \theta \in \Theta \tag{A1}$$

where $\underline{U} = (P(q(\underline{\theta})) + \underline{\theta} - c_0)q(\underline{\theta})$. Following Amador and Bagwell (2013), the incentive compatible constraints can be re-stated as a monotonicity requirement and two inequalities which can be shown below:

$$\int_{\underline{\theta}}^{\theta} q(\tilde{\theta})d\tilde{\theta} + \underline{U} - (P(q(\theta)) + \theta - c_0)q(\theta) \leq 0 \text{ for all } \theta \in \Theta, \tag{A2}$$

$$-\int_{\underline{\theta}}^{\theta} q(\tilde{\theta})d\tilde{\theta} - \underline{U} + (P(q(\theta)) + \theta - c_0)q(\theta) \leq 0 \text{ for all } \theta \in \Theta. \tag{A3}$$

By assigning two non-decreasing cumulative Lagrange multiplier functions $\lambda_1(\theta)$ and $\lambda_2(\theta)$ associated with the two inequalities (A2) and (A3), and denoting a non-decreasing multiplier function $\mu(\theta)$ of the participation constraint that satisfy complementary slackness, the Lagrangian for the problem writes as:

$$L = \int_{\underline{\theta}}^{\bar{\theta}} w(q(\theta), \theta)dF(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \left[\int_{\underline{\theta}}^{\theta} q(\tilde{\theta})d\tilde{\theta} + \underline{U} - (P(q(\theta)) + \theta - c_0)q(\theta) \right] d(\lambda_1(\theta) - \lambda_2(\theta)) + \int_{\underline{\theta}}^{\bar{\theta}} (P(q(\theta)) + \theta - c_0)q(\theta)d\mu(\theta).$$

Let us propose some non-decreasing multiplier functions to satisfy

$$\lambda(\theta) = \lambda_1(\theta) - \lambda_2(\theta) = \begin{cases} 0 & ; \theta = \bar{\theta}, \\ -w_q(q^f(\theta), \theta)f(\theta) & ; \theta \in (\theta_r, \bar{\theta}), \\ -\frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})dF(\tilde{\theta}) - \kappa F(\theta) & ; \theta \in [\underline{\theta}, \theta_r], \end{cases}$$

and $\mu(\theta) = \begin{cases} 0 & ; \theta \in (\underline{\theta}, \bar{\theta}] \\ -\frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})dF(\tilde{\theta}) & ; \theta = \underline{\theta} \end{cases}$, where κ is the relative

concavity parameter. The proposition 2 ensures that $\kappa F(\theta) + \lambda(\theta)$ is non-decreasing. We can propose $\lambda_1(\theta) = \kappa F(\theta) + \lambda(\theta)$ and $\lambda_2(\theta) = \kappa F(\theta)$. To satisfy $\mu(\theta)$ is non-decreasing, we require $\frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})dF(\tilde{\theta}) \geq 0$.

To check whether the proposed allocation maximizes the resulting Lagrangian with proposed Lagrange multipliers, it is particularly useful to verify that the resulting Lagrangian is concave in q and the first-order conditions are satisfied. First, we now check the concavity of the Lagrangian. Using these proposed multiplier functions and integrating by parts the Lagrangian, we can obtain

$$L = \int_{\underline{\theta}}^{\bar{\theta}} w(q(\theta), \theta)dF(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \left[\int_{\underline{\theta}}^{\theta} q(\tilde{\theta})d\tilde{\theta} + \underline{U} - (P(q(\theta)) + \theta - c_0)q(\theta) \right] d\lambda(\theta) + \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})dF(\tilde{\theta})(P(q(\underline{\theta})) + \underline{\theta} - c_0)q(\underline{\theta})).$$

Using $\underline{U} = (P(q(\underline{\theta})) + \underline{\theta} - c_0)q(\underline{\theta})$, $\lambda(\underline{\theta}) = -\frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})dF(\tilde{\theta})$ and $\lambda(\bar{\theta}) = 0$, the Lagrangian can be written as

$$L = \int_{\underline{\theta}}^{\bar{\theta}} w(q(\theta), \theta)dF(\theta) - \int_{\underline{\theta}}^{\bar{\theta}} \left[\int_{\underline{\theta}}^{\theta} q(\tilde{\theta})d\tilde{\theta} - (P(q(\theta)) + \theta - c_0)q(\theta) \right] d\lambda(\theta).$$

Integrating by parts, it can be rewritten as

$$L = \int_{\underline{\theta}}^{\bar{\theta}} (w(q(\theta), \theta)f(\theta) + \lambda(\theta)q(\theta))d\theta + \int_{\underline{\theta}}^{\bar{\theta}} P(q(\theta)) + \theta - c_0)q(\theta)d\lambda(\theta).$$

Adding and deducting $\kappa(P(q(\theta)) + \theta - c_0)q(\theta)f(\theta)$, we get.

$L = \int_{\underline{\theta}}^{\bar{\theta}} \{ [w(q(\theta), \theta) - \kappa(P(q(\theta)) + \theta - c_0)q(\theta))]f(\theta) + \lambda(\theta)q(\theta) \}d\theta + \int_{\underline{\theta}}^{\bar{\theta}} (P(q(\theta)) + \theta - c_0)q(\theta) d(\kappa F(\theta) + \lambda(\theta))$. Recalling the relative concavity parameter, we get $w_{qq}(q, \theta) - \kappa u_{qq}(q, \theta) \leq 0$. Therefore, the Lagrangian is concave in $q(\theta)$ if $\kappa F(\theta) + \lambda(\theta)$ is non-decreasing for all $\theta \in \Theta$. The conditions in Proposition 2 show that $\kappa F(\theta) - w_q(q^f(\theta), \theta)f(\theta)$ is non-decreasing for all $\theta \in (\theta_r, \bar{\theta}]$. Thus, we only need to check that the jumps at $\bar{\theta}$ and θ_r are nonnegative. The jumps are

$$-w_q(q^f(\bar{\theta}), \bar{\theta})f(\bar{\theta}) \leq 0, \kappa F(\theta_r) - w_q(q^f(\theta_r), \theta_r)f(\theta_r) \geq -\frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \bar{\theta})dF(\bar{\theta}).$$

The former is satisfied by the utility’s direction of bias, and the latter is satisfied by the condition (ii) in Proposition 2. Therefore, the Lagrangian is concave at the proposed multiplier functions.

We now proceed to show the quantity IR-floor allocation maximizes the Lagrangian. Following Amador and Bagwell (2013), the entire positive ray of the real line for concave function w and u and a convex cone of choice set $\Omega = \{ q|q : \Theta \rightarrow \mathfrak{R}_+ \text{ and } q \text{ nondecreasing} \}$ are extended. Maximizing the concave functions on a convex cone needs the Lagrangian is a concave functional and the following first-order conditions are satisfied in terms of Gateaux differentials:

$$\partial L(q^r, q^r) = 0$$

$$\partial L(q^r, x) \leq 0 \text{ for all } x \in \Omega.$$

Taking Gateaux differential in direction x , using $(P'(q^f(\theta) + \theta - c_0)q^f(\theta) + P(q^f(\theta))) = 0$ and the constructed multiplier functions, we have

$$\partial L(q^r, x) = \int_{\underline{\theta}}^{\theta_r} [w_q(q^f(\theta_r), \theta)f(\theta) - \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \bar{\theta})dF(\bar{\theta}) - \kappa F(\theta) + \kappa(\theta_r - \theta)f(\theta)]x(\theta)d\theta$$

which can be rewritten through integrating by parts as

$$\begin{aligned} \partial L(q^r, x) = & \int_{\underline{\theta}}^{\theta_r} [w_q(q^f(\theta_r), \theta)f(\theta) - \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \bar{\theta})dF(\bar{\theta}) - \kappa F(\theta) + \kappa(\theta_r - \theta)f(\theta)]d\theta x(\theta) \\ & + \int_{\underline{\theta}}^{\theta_r} \{ \int_{\theta}^{\theta_r} [w_q(q^f(\theta_r), \bar{\theta})f(\bar{\theta}) - \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \bar{\theta})dF(\bar{\theta}) - \kappa F(\bar{\theta}) + \kappa(\theta_r - \bar{\theta})f(\bar{\theta})]d\bar{\theta} \} dx(\theta). \end{aligned}$$

By $\partial L(q^r, q^r) = 0$, we have

$$\int_{\underline{\theta}}^{\theta_r} [w_q(q^f(\theta_r), \theta)f(\theta) - \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \bar{\theta})dF(\bar{\theta}) - \kappa F(\theta) + \kappa(\theta_r - \theta)f(\theta)]d\theta x(\theta) = 0.$$

So, we need to satisfy the following inequality

$$\int_{\underline{\theta}}^{\theta_r} \left\{ \int_{\underline{\theta}}^{\theta_r} \left[w_q(q^f(\theta_r), \tilde{\theta})f(\tilde{\theta}) - \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})dF(\tilde{\theta}) - \kappa F(\tilde{\theta}) + \kappa(\theta_r - \tilde{\theta})f(\tilde{\theta}) \right] d\tilde{\theta} \right\} dx(\theta) \leq 0.$$

Integrating by parts, we can get the following inequality holds under the condition (ii) in Proposition 2:

$$\int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})f(\tilde{\theta})d\tilde{\theta} - \frac{\theta_r - \underline{\theta}}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})f(\tilde{\theta})d\tilde{\theta} - \kappa(\theta_r - \underline{\theta})F(\theta) \leq 0.$$

And the condition (i) and (iii) are needed to satisfy the non-decreasing property of the proposed multiplier functions.

We now complete the proof to apply the modified version of Luenberger’s Sufficiency Theorem (1969) in Amador and Bagwell (2013). Setting

- (1) $x_0 = q^r$;
- (2) $X = \{q|q : \Theta \rightarrow \mathbf{O}\}$;
- (3) $\Omega = \{q|q : \Theta \rightarrow \mathfrak{R}_+ \text{ and } q \text{ nondecreasing}\}$;
- (4) f , as a real valued functional of $q \in X$, is the negative of the objective function $Ew(q(\theta), \theta)$;
- (5) $Z = \{(z_1, z_2, z_3)|z_1 : \Theta \rightarrow \mathfrak{R}, z_2 : \Theta \rightarrow \mathfrak{R}, z_3 : \Theta \rightarrow \mathfrak{R}\}$;
- (6) $P = \{(z_1, z_2, z_3)|(z_1, z_2, z_3) \in Z \text{ such that } z_1(\theta) \geq 0, z_2(\theta) \geq 0, z_3(\theta) \geq 0 \text{ for all } \theta \in \Theta\}$;
- (7) The mapping G from Ω to Z is given by the left sides of inequalities (A1), (A2) and (A3);
- (8) The linear mapping T is given by $T((z_1, z_2, z_3)) = \int_{\underline{\theta}}^{\theta_r} z_1 d\lambda_1(\theta) + \int_{\underline{\theta}}^{\theta_r} z_2 d\lambda_1(\theta) + \int_{\underline{\theta}}^{\theta_r} z_3 d\mu(\theta)$,

where non-decreasing multiplier functions $\lambda_1(\theta)$, $\lambda_2(\theta)$ and $\mu(\theta)$, and $\mu(\theta)$ imply $T(z) \geq 0$ for $z \in P$. When (A1), (A2) and (A3) bind under the q^r allocation and the proposed multiplier $\mu(\theta)$ is considered, we get

$$T(G(x_0)) = \int_{\underline{\theta}}^{\theta_r} \left[\int_{\underline{\theta}}^{\theta_r} q^r(\tilde{\theta})d\tilde{\theta} + \underline{U} - (P(q^r(\theta)) + \theta - c_0)q^r(\theta) \right] d(\lambda_1(\theta) - \lambda_2(\theta)) = 0.$$

Therefore, we have found the conditions in Proposition 2 under which the proposed the q^r allocation solves the minimization problem of $f(x)$ for $x \in \Omega$ subject to $-G(x) \in P$.

Proof of Corollary 1

First, we define $T(\theta) = \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})f(\tilde{\theta})d\tilde{\theta} - \kappa F(\theta)$ for all $\theta \in [\underline{\theta}, \theta_r]$ and $T(\underline{\theta}) = \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} w_q(q^f(\theta_r), \tilde{\theta})f(\tilde{\theta})d\tilde{\theta}$.

Considering that $w_q(q(\theta), \theta) = -(1 - \alpha)P'(q(\theta))q(\theta) + \alpha(P(q(\theta)) + \theta - c_0)$, we have

$$\begin{aligned}
 T(\theta) &= \frac{1}{\theta_r - \theta} \int_{\theta}^{\theta_r} [-(1 - \alpha)P'(q^f(\theta_r))q^f(\theta_r) + \alpha(P(q^f(\theta_r)) + \bar{\theta} - c_0)]f(\bar{\theta})d\bar{\theta} - \kappa F(\theta) \\
 &= \frac{1}{\theta_r - \theta} \int_{\theta}^{\theta_r} \alpha(\bar{\theta} - \underline{\theta})f(\bar{\theta})d\bar{\theta} - \kappa F(\theta) - (1 - \alpha)P'(q^f(\theta_r))q^f(\theta_r) \frac{F(\theta_r) - F(\theta)}{\theta_r - \theta}.
 \end{aligned}$$

First, we find that the condition (i) in Proposition 2 can be automatically satisfied, as following

$$T(\underline{\theta}) = \frac{1}{\theta_r - \underline{\theta}} \int_{\underline{\theta}}^{\theta_r} \alpha(\bar{\theta} - \underline{\theta})f(\bar{\theta})d\bar{\theta} + (\alpha - 1)P'(q^f(\theta_r))q^f(\theta_r) \frac{F(\theta_r)}{\theta_r - \underline{\theta}} \geq 0.$$

Second, for the condition (ii) in Proposition 2, we get

$$G'(\theta) = \frac{1}{(\theta_r - \theta)^2} \int_{\theta}^{\theta_r} \alpha(\bar{\theta} - \underline{\theta})f(\bar{\theta})d\bar{\theta} - \alpha \frac{\theta - \underline{\theta}}{\theta_r - \theta} f'(\theta) - \kappa f(\theta) - \frac{(1 - \alpha)P'(q^f(\theta_r))q^f(\theta_r)}{\theta_r - \theta} \left(\frac{F(\theta_r) - F(\theta)}{\theta_r - \theta} - f(\theta) \right).$$

If $f'(\theta) \leq 0$ and $\kappa \geq \frac{1}{2}\alpha$, we have

$$G'(\theta) \leq \frac{1}{(\theta_r - \theta)^2} \int_{\theta}^{\theta_r} \alpha(\bar{\theta} - \underline{\theta})f(\bar{\theta})d\bar{\theta} - \alpha \frac{\theta - \underline{\theta}}{\theta_r - \theta} f'(\theta) - \kappa f(\theta) = f(\theta) \left(\frac{1}{2}\alpha - \kappa \right) \leq 0.$$

Then, $T(\theta) \leq T(\underline{\theta})$ and the condition (ii) is satisfied.

Finally, for the condition (iii) in Proposition 2, we define $R(\theta) = \kappa F(\theta) - w_q(q^f(\theta), \theta)f(\theta) = \kappa F(\theta) + P'(q^f(\theta))q^f(\theta)f(\theta)$. Taking the derivation with the respect of θ , we have

$$R(\theta) = \kappa F(\theta) - w_q(q^f(\theta), \theta)f(\theta) = \kappa F(\theta) + P'(q^f(\theta))q^f(\theta)f(\theta). \quad \text{Considering}$$

$$\kappa = \min_{\theta, q} \left\{ \frac{w_{qq}(q, \theta)}{u_{qq}(q, \theta)} \right\} \leq \frac{-(1 - \alpha)P''(q)q + (2\alpha - 1)P'(q)}{P''(q)q + 2P'(q)} = \alpha - \frac{P''(q)q + P'(q)}{P''(q)q + 2P'(q)}, \quad \text{we obtain}$$

$P''(q)q + P'(q) \geq -(\kappa - \alpha)(P''(q)q + 2P'(q))$. Thus, if $f'(\theta) \leq 0$ and $\kappa \geq \frac{1}{2}\alpha$, the condition (iii) will hold as following

$$\begin{aligned}
 R'(\theta) &\geq \kappa f(\theta) - (\kappa - \alpha)(P''(q^f(\theta))q^f(\theta) + 2P'(q^f(\theta)))q^f(\theta)f(\theta) \\
 &\quad + P'(q^f(\theta))q^f(\theta)f'(\theta) = (2\kappa - \alpha)f(\theta) + P'(q^f(\theta))q^f(\theta)f'(\theta) \geq 0
 \end{aligned}$$

in which $q^f(\theta) = -\frac{1}{u_{qq}(q^f(\theta), \theta)} = -\frac{1}{P''(q^f(\theta))q^f(\theta) + 2P'(q^f(\theta))}$.

Proof of Proposition 3

With decreasing marginal cost, the regulator’s problem is to choose one unique p to maximize $E_{\theta}w(p, \theta)$ subject to the IR constraint $u(p, \theta) \geq 0$. Characterizing the optimal pooling price p^s when the IR constraint is ignored, we can obtain the price p^s will violates the firm’s IR constraint if

$$p^s - c_0 - \frac{1}{2}c_1Q(p^s, \theta) = \left(1 - \frac{1}{2}c_1Q_p(p^s, E(\theta)) - \frac{1}{\alpha}\right) \frac{Q(p^s, E(\theta))}{-Q_p(p^s, E(\theta))} - \frac{1}{2}c_1Q_p(p^s, E(\theta))(E(\theta) - \theta) < 0.$$

When $c_1 < 0$, the IR constraint will be violated, since $p^s - c_0 - \frac{1}{2}c_1Q(p^s, \theta) < 0$. The IR constraint can be written as

$$u(p, \theta) = (p - c_0 - c_1Q(p, \theta))Q(p, \theta) + \frac{1}{2}c_1Q^2(p, \theta) \geq 0. \text{ Thus, we can obtain}$$

$u_\theta(p, \theta) = (p - c_0 - c_1Q(p, \theta))Q_\theta(p, \theta) \geq 0$, since $c_1 < 0$, $Q_\theta(p, \theta) > 0$. Therefore, the optimal price is determined when the IR constraint binds for lowest type, as $u(p^u, \theta) = p^uQ(p^u, \theta) - c_0Q(p^u, \theta) - \frac{1}{2}c_1Q^2(p^u, \theta) = 0$.

Proof of Proposition 4

From the first-order condition and the assumed linear demand function, we can obtain

$$q^s = \frac{\alpha(P(0) + E(\theta) - c_0)}{-(2\alpha - 1)P'(q^s) + \alpha c_1}.$$

Defining $A = \left[\frac{1}{2}c_1 - \left(1 - \frac{1}{\alpha}\right)P'(q^s)\right]q^s - (E(\theta) - \theta)$ and substituting q^s , we can have

$$A = \left[\frac{1}{2}c_1 - \left(1 - \frac{1}{\alpha}\right)P'(q^s)\right] \frac{\alpha(P(0) + E(\theta) - c_0)}{-(2\alpha - 1)P'(q^s) + \alpha c_1} - (E(\theta) - \theta).$$

Assumed that $P(0) + E(\theta) - c_0 > 4(E(\theta) - \theta)$, it shows if $c_1 \geq 2\left(1 - \frac{1}{\alpha} - \frac{E(\theta) - \theta}{\alpha[P(0) + E(\theta) - c_0 - 2(E(\theta) - \theta)]}\right)P'(q^s) > 0$, $A > 0$, and thus, the optimal quantity q^s satisfies the IR constraint. Otherwise, $A < 0$ and q^s violates the IR constraint and the IR constraint binds for θ if

$$2P'(q^s) < c_1 < 2\left(1 - \frac{1}{\alpha} - \frac{E(\theta) - \theta}{\alpha[P(0) + E(\theta) - c_0 - 2(E(\theta) - \theta)]}\right)P'(q^s).$$

Proof of Proposition 5

When $c_1 < 0$, the IR constraint will be violated, since $p^s - c_0 - \frac{1}{2}c_1Q(p^s, \theta) < 0$. The IR constraint can be written as $u(p, \theta) = (p - c_0 - c_1Q(p, \theta))Q(p, \theta) + \frac{1}{2}c_1Q^2(p, \theta) \geq 0$. Thus, we can obtain $u_\theta(p, \theta) = (p - c_0 - c_1Q(p, \theta))Q_\theta(p, \theta) \geq 0$, since $c_1 < 0$, $Q_\theta(p, \theta) > 0$. Therefore,

the optimal price is determined when the IR constraint binds for lowest type, as $u(p^u, \theta) = p^u Q(p^u, \theta) - c_0 Q(p^u, \theta) - \frac{1}{2} c_1 Q^2(p^u, \theta) = 0$.

When $c_1 \geq 0$, we define

$$B = \left(1 - \frac{1}{2} c_1 Q_p(p^s, E(\theta)) - \frac{1}{\alpha} \right) \frac{Q(p^s, E(\theta))}{-Q_p(p^s, E(\theta))} - \frac{1}{2} c_1 Q_p(p^s, E(\theta))(E(\theta) - \theta).$$

Considering the linear demand and $\alpha = 1$, $B = \frac{1}{2} c_1 Q(p^s, 2E(\theta) - \theta) > 0$. Therefore, the optimal price p^s satisfies the IR constraint.

Proof of Corollary 3

Consider $P(q) = P(0) + P'(q)q$ and $Q(p, \theta) = \frac{P(0)+\theta-p}{-P'(q)}$.

Part 1. $c_1 < 0$. We can obtain $q^u = \frac{2(P(0)+\theta-c_0)}{c_1-2P'(q)}$ and $p^u = \frac{2P'(q)c_0-c_1(P(0)+\theta)}{2P'(q)-c_1}$, that are independent with the weight α . Thus, the social welfare of the two regulatory mechanism is linear and increasing with the weight α .

When $\alpha = 0$, the difference of the regulator’s welfare is

$$\begin{aligned} \Delta &= E_\theta(w(q^u, \theta)) - E_\theta(w(p^u, \theta)) = E_\theta \left[\int_0^{q^u} (P(\tilde{q}) + \theta) d\tilde{q} - (P(q^u) \right. \\ &\quad \left. + \theta)q^u - E_\theta \left[\int_0^{Q(p^u, \theta)} (P(\tilde{q}) + \theta) d\tilde{q} - (P(Q(p^u, \theta)) + \theta)Q(p^u, \theta) \right] \right] \\ &= -\frac{P'(q)}{2} (q^{u2} - E_\theta Q^2(p^u, \theta)). \end{aligned}$$

Considering that $E_\theta Q^2(p^u, \theta) > Q^2(p^u, E(\theta))$ through Jensen’s Inequality, we get $\Delta < -\frac{P'(q)}{2} (q^{u2} - Q^2(p^u, E(\theta))) \leq 0$, since $q^u - Q(p^u, E(\theta)) = \frac{E(\theta)-\theta}{P'(q)} \leq 0$.

When $\alpha = 1$,

$$\Delta = \int_0^{q^u} (P(\tilde{q}) + E(\theta)) d\tilde{q} - \left(c_0 + \frac{1}{2} c_1 q^u \right) q^u - E_\theta \left[\int_0^{Q(p^u, \theta)} (P(\tilde{q}) + \theta) d\tilde{q} - (c_0 + \frac{1}{2} c_1 Q(p^u, \theta)) Q(p^u, \theta) \right].$$

Note that $w_{\theta\theta}(p^u, \theta) = -c_1 Q_\theta^2(p^u, \theta) > 0$. Thus, $E_\theta(w(p^u, \theta)) > w(p^u, E(\theta))$ by Jensen’s Inequality. We have $\Delta < \int_0^{q^u} (P(\tilde{q}) + E(\theta)) d\tilde{q} - \left(c_0 + \frac{1}{2} c_1 q^u \right) q^u - \left[\int_0^{Q(p^u, \theta)} (P(\tilde{q}) + E(\theta)) d\tilde{q} - (c_0 + \frac{1}{2} c_1 Q(p^u, E(\theta))) Q(p^u, E(\theta)) \right]$. Considering $w_q(q, E(\theta)) = P(q) + E(\theta) - c_0 - c_1 q = P(q) + \underline{\theta} - c_0 - \frac{1}{2} c_1 q + E(\theta) - \underline{\theta} - \frac{1}{2} c_1 q > 0$ where $P(q) + \underline{\theta} - c_0 - \frac{1}{2} c_1 q \geq 0$ by IR constraint, it shows $\Delta < 0$. Consequently, $E_\theta(w(q^u, \theta)) < E_\theta(w(p^u, \theta))$ for all $\alpha \in [0, 1]$.

Part 2. $0 \leq c_1 < c_1^q$ in which $c_1^q = -\frac{2(E(\theta)-\theta)}{P(0)+E(\theta)-c_0-2(E(\theta)-\theta)} P'(q)$ when $\alpha = 1$. Now, we have

$$\Delta = E_\theta [\int_0^{q^u} (P(\tilde{q}) + \theta) d\tilde{q} - (c_0 + \frac{1}{2}c_1q^u)q^u] - E_\theta [\int_0^{Q(p^u, \theta)} (P(\tilde{q}) + \theta) d\tilde{q} - (c_0 + \frac{1}{2}c_1Q(p^u, \theta))Q(p^u, \theta)] .$$

Considering $w(p^u, \theta) = \int_0^{q^u} (P(\tilde{q}) + \theta) d\tilde{q} - (c_0 + \frac{1}{2}c_1q^u)q^u + (P(q^u) + \theta - c_0 - c_1q^u)(Q(p^u, \theta) - q^u) + \frac{1}{2}(P'(q^u) - c_1)(Q(p^u, \theta) - q^u)^2$, we have

$$\Delta = E_\theta \left[(P(q^u) + \theta - c_0 - c_1q^u)(q^u - Q(p^u, \theta)) + \frac{1}{2}(c_1 - P''(q))(q^u - Q(p^u, \theta))^2 \right] .$$

Noting $q^u = \frac{2(P(0)+\underline{\theta}-c_0)}{c_1-2P'(q)}$ and $Q(p^u, \theta) = \frac{P(0)+E(\theta)-c_0+(1+\frac{c_1}{-P'(q)})(\theta-E(\theta))}{c_1-P'(q)}$, we get $\Delta = \frac{-c_1^2}{2(c_1-P'(q))(c_1-2P'(q))^2} E_\theta K(\theta)$, where $K(\theta) = m(\theta)n(\theta)$ such that $m(\theta) = P(0) - c_0 + \underline{\theta} + \frac{(2P'(q)-c_1)(P'(q)+c_1)}{c_1P'(q)}(\theta - \underline{\theta}) + \left(\frac{c_1}{P'(q)} - 2\right)(E(\theta) - \underline{\theta})$ and $n(\theta) = P(0) - c_0 + \underline{\theta} + \left(2 - \frac{c_1}{P'(q)}\right)(E(\theta) - \underline{\theta}) + \frac{(2P'(q)-c_1)(P'(q)-c_1)}{c_1P'(q)}(\theta - \underline{\theta})$.

From $c_1^q = -\frac{2(E(\theta)-\underline{\theta})}{P(0)+E(\theta)-c_0-2(E(\theta)-\underline{\theta})} P'(q)$, it shows

$P(0) + \underline{\theta} - c_0 = E(\theta) - \underline{\theta} - \frac{2(E(\theta)-\underline{\theta})}{c_1^q} P'(q)$. Substituting it to $m(\theta)$ and $n(\theta)$, we have

$$m(\theta) = -2P'(q) \left(\frac{1}{c_1} - \frac{1}{c_1^q} \right) (E(\theta) - \underline{\theta}) + \frac{(2P'(q)-c_1)(P'(q)+c_1)}{c_1P'(q)} (\theta - E(\theta))$$

$$n(\theta) = -2P'(q) \left(\frac{1}{c_1} - \frac{1}{c_1^q} \right) (E(\theta) - \underline{\theta}) + \frac{(2P'(q)-c_1)(P'(q)-c_1)}{c_1P'(q)} (\theta - E(\theta)),$$

where $P'(q) + c_1 < P'(q) + c_1^q < 0$ under the assumption of $P(0) + E(\theta) - c_0 > 4(E(\theta) - \underline{\theta})$.

Note that $K''(\theta) = \frac{(2P'(q)-c_1)^2(P'(q)-c_1^2)}{2c_1^2P'^2(q)} > 0$, and thus

$$E_\theta K(\theta) > K(E(\theta)) = 4P'^2(q) \left(\frac{1}{c_1} - \frac{1}{c_1^q} \right)^2 (E(\theta) - \underline{\theta})^2 \geq 0,$$

by Jensen's Inequality. Therefore, it shows $\Delta < 0$.

Part 3. $c_1 \geq c_1^q$. We get $q^u = \frac{P(0)+E(\theta)-c_0}{c_1-P'(q)}$, $p^u = \frac{c_1(P(0)+E(\theta)-c_0)P'(q)}{c_1-P'(q)}$.

Similar to Part 2, we have

$$\Delta = E_\theta \left[(P(q^u) + \theta - c_0 - c_1q^u)(q^u - Q(p^u, \theta)) + \frac{1}{2}(c_1 - P'(q))(q^u - Q(p^u, \theta))^2 \right] .$$

Considering $P(q^u) + \theta - c_0 - c_1q^u = \theta - E(\theta)$, $q^u - Q(p^u, \theta) = \frac{E(\theta)-\underline{\theta}}{-P'(q)}$, we obtain $\Delta = E_\theta \left[\frac{1}{2}(c_1 + P'(q)) \frac{(E(\theta)-\underline{\theta})^2}{P'^2(q)} \right]$. Therefore, $\Delta > 0$ only if $c_1 < -P'(q)$. Note that $c_1^q < -P'(q)$ under the assumption of $P(0) + E(\theta) - c_0 > 4(E(\theta) - \underline{\theta})$. To ensure $Q(p^u, \theta)$ is not negative, we assume that $c_1 \leq -P'(q) \left(\frac{P(0)+E(\theta)-c_0}{E(\theta)-\underline{\theta}} - 1 \right)$.

Acknowledgements This research was financially supported by the Fundamental Research Funds for the Central Universities (Grant No. CCNU22XJ030), the National Natural Science Foundation of China (Grant No. 72204193) and the Foundation Anhui Education Department (Grant No. 2022AH050027). The authors would like to thank the funded project for providing material for this research. We would like to thank our anonymous reviewer for the valuable comments in developing this manuscript.

Author contributions Dan Wang wrote the main manuscript text, Peng Hao and Jiancheng Wang prepared figures and proofread the language. All authors reviewed the manuscript.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Aguirre I, Beitia A (2004) Regulating a monopolist with unknown demand: costly public funds and the value of private information. *J Pub Econ Theory* 6(5):693–706
- Alonso R, Matouschek N (2008) Optimal delegation. *Rev Econ Stud* 75(1):259–293
- Amador M, Bagwell K (2013) The theory of optimal delegation with an application to tariff caps. *Econometrica* 81(4):1541–1599
- Amador M, Bagwell K (2022) Regulating a monopolist with uncertain costs without transfers. *Theor Econ* 17:1719–1760
- Anton K, Andriy Z (2019) Persuasion meets delegation. Working Papers, 1902.02628, arXiv.org.
- Armstrong M, Sappington DEM (2007) Recent developments in the theory of regulation. In: Armstrong M, Porter R (eds) *Handbook of industrial organization*. North-Holland, Amsterdam
- Baron DP (1989) Design of regulatory mechanisms and institutions. In: Schmalensee R, Willig RD (eds) *Handbook of industrial organization*. North-Holland, Amsterdam
- Baron DP, Myerson RB (1982) Regulating a monopolist with unknown costs. *Econometrica* 50(4):911–930
- Basso LJ, Figueroa N, Vásquez J (2017) Monopoly regulation under asymmetric information: prices versus quantities. *Rand J Econ* 48(3):557–578
- Holmström B (1984) *On the theory of delegation* Bayesian models in economic theory. Elsevier Science Ltd, New York
- Krysiak FC (2008) Prices vs. quantities: the effects on technology choice. *J Pub Econ* 92:1275–1287
- Krysiak FC, Oberauner IM (2010) Environmental policy a la carte: letting firms choose their regulation. *J Environ Econ Manag* 60:221–232
- Laffont JJ, Tirole J (1986) Using cost observation to regulate firms. *J Polit Econ* 94:614–641
- Laffont JJ, Tirole J (1993) *A theory of incentives in procurement and regulation*. MIT Press, Cambridge, MA
- Lewis TR, Sappington DE (1988a) Regulating a monopolist with unknown demand. *Am Econ Rev* 78:986–998
- Lewis TR, Sappington DE (1988b) Regulating a monopolist with unknown demand and cost functions. *Rand J Econ* 19(3):438–457
- Moledina AA, Coggings JS, Polasky S, Costello C (2003) Dynamic environmental policy with strategic firms: prices versus quantities. *J Environ Econ Manag* 45:356–376
- Riordan MH (1984) On delegation price authority to a regulated firm. *Rand J Econ* 15:108–115
- Samuelson PA (1938) A note on the pure theory of consumer's behavior. *Economica* 5(17):61–71
- Weitzman ML (1974) Prices vs. quantities. *Rev Econ Stud* 41:477–491
- Williams III RC (2002) Prices vs. quantities vs. tradable quantities. Working paper 9283, NBER

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.