Regulating a Monopolist With Uncertain Costs Without Transfers*

Manuel Amador
Federal Reserve Bank of Minneapolis

Kyle Bagwell†
Stanford University

November 15, 2016

Abstract

We analyze the Baron and Myerson (1982) model of regulation under the restriction that transfers are infeasible. Extending the Lagrangian approach to delegation problems of Amador and Bagwell (2013b) to include an ex post participation constraint, we report sufficient conditions under which optimal regulation takes the simple and common form of price-cap regulation. We also identify families of demand and distribution functions and welfare weights that are sure to satisfy our sufficient conditions. We illustrate our sufficient conditions using examples with log demand, linear demand, constant elasticity demand and exponential demand, respectively. We also establish our findings for two representations of the ex post participation constraint, where the representations differ regarding the feasibility of exclusion for some types.

1 Introduction

The optimal regulatory policy for a monopolist is influenced by many considerations, including the possibility of private information, the objective of the regulator, and the feasibility and efficiency of transfers. Simple solutions obtain in some settings. For example, in the textbook case of a single-product monopolist with constant marginal cost and a positive fixed cost, with all costs commonly known, a regulator that maximizes aggregate social surplus obtains the "first-best" ("second-best") solution by setting price equal to marginal (average) cost when transfers are feasible and efficient (are infeasible). In other settings,

*The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

†We would like to thank Mark Armstrong, Joe Harrington, Roger Noll, Alessandro Pavan, Peter Troyan, Robert Wilson, Frank Wolak and seminar participants at Nottingham, Penn, Ryerson, SMU, Vanderbilt and USC for helpful discussions. Manuel Amador acknowledges NSF support under award number 0952816. Kyle Bagwell thanks the Center for Advanced Studies in the Behavioral Sciences at Stanford for hospitality and support as a Fellow during 2014-15. Corresponding author, email: amador.manuel@gmail.com; fax: 612-204-5515.

however, optimal regulation can take more subtle forms. Armstrong and Sappington (2007) survey the nature of optimal regulation in different settings and discuss as well the design of practical policies, such as price-cap regulation, that are frequently observed in practice. As they emphasize, an important question is whether practical policies perform well in realistic settings where private information may be present and transfer instruments may be limited.

In a seminal paper, Baron and Myerson (1982) consider the optimal regulation of a single-product monopolist with private information about its costs of production. In their model, a regulatory policy indicates, for every possible cost type, a price and a transfer from consumers to the monopolist, and a regulatory policy is feasible if it is incentive compatible and satisfies an ex post participation constraint. The regulator chooses over feasible regulatory policies to maximize a weighted social welfare function that weighs consumer surplus no less heavily than producer surplus.[1] In a standard representation of their model, the monopolist incurs a commonly known and non-negative fixed cost and is privately informed as to the level of its constant marginal cost, where the monopolist's marginal cost has a continuum of possible types drawn from a commonly known distribution function. If the regulator gives greater welfare weight to consumer surplus and the distribution function is log concave, then the optimal regulatory policy defines a non-decreasing price schedule with a positive mark up for all but the lowest cost type. By comparison, if the regulator were to maximize aggregate social surplus, then as Loeb and Magat (1979) observe the optimal regulatory policy would achieve a first-best outcome, with price equal to marginal cost and transfers set so that the monopolist receives all consumer surplus.

In this paper, we characterize optimal regulatory policy in the Baron-Myerson model with constant marginal costs when transfers are infeasible. Our no-transfers assumption contrasts sharply with Baron and Myerson's assumption that all (positive and negative) transfers are available. We motivate our no-transfers assumption in three ways. First, as is commonly observed, regulators often do not have the authority to explicitly tax or pay subsidies.[2] Second, while transfers from consumers to firms may also be achieved via access fees in two-part tariff schemes, the scope for such transfers may be limited in practice, particularly when universal service is sought and consumers are heterogeneous.[3] Finally, in other settings, the

[1] An alternative approach is developed by Laffont and Tirole (1993, 1986). They assume that the regulator maximizes aggregate social surplus and that transfers are inefficient (i.e., transfers entail a social cost of funds). Under this approach, consumers incur a cost in excess of one dollar for every dollar that is received as a transfer by the monopolist.

[2] For further discussion, see, e. g., (Armstrong and Sappington, 2007, p. 1607), (Baron, 1989, p. 1351), (Church and Ware, 2000, p. 840), (Joskow and Schmalensee, 1986, p. 5), (Laffont and Tirole, 1993, p. 130) and (Schmalensee, 1989, p. 418).

[3] As Laffont and Tirole (1993, p. 151) explain, "optimal linear pricing is a good approximation to optimal two-part pricing when there is concern that a nonnegligible fixed premium would exclude either too many customers or customers with low incomes whose welfare is given substantial weight in the social welfare

scope for a positive access fee may be limited by the possibility of consumer arbitrage, while the scope for a negative access fee may be limited by the prospect of strategic consumer behavior designed to capture "sign-up" bonuses. In view of these considerations, we remove the traditional assumption that all transfers are available and consider the opposite case in which all transfers are infeasible. Specifically, we assume that the regulated firm is restricted to a uniform price (i.e., linear pricing).[4] As our main finding, we report sufficient conditions under which price-cap regulation emerges as the optimal regulatory policy.

Our sufficient conditions take two forms. We first offer general sufficient conditions for the optimality of price-cap regulation. These conditions are defined in terms of general relationships between functions that describe the regulator's welfare, the monopolist's profit and the distribution of cost types, respectively. Once a specific regulatory setting is proposed, these general relationships may be checked. Building on this logic, our second set of sufficient conditions is stronger and identifies families of demand and distribution functions and welfare weights that are sure to satisfy the general sufficient conditions. We develop this approach in two ways. First, we examine the case of a log demand function and establish that our general sufficient conditions are then satisfied under a simple restriction on the distribution function and welfare weight. The restriction is sure to hold if the density is non-decreasing but can also hold for distributions that are non-monotonic or even decreasing. Second, we show that our general sufficient conditions are satisfied if the density is non-decreasing and an easy-to-check inequality holds. The inequality captures a relationship between properties of the demand function and the welfare weight. Using this inequality, we show that price-cap regulation is optimal if the density is non-decreasing and (i) demand is linear or exponential and the regulator maximizes aggregate social surplus, or (ii) demand takes a constant elasticity or log form. The inequality fails to hold, however, if the inverse demand function is strictly concave. We note that the case in which the regulator maximizes aggregate social surplus is of particular interest from a normative standpoint.

As mentioned above, price-cap regulation is a common form of regulation. The appeal of price-cap regulation is often associated with the incentive that it gives to the regulated firm to invest in endogenous cost reduction.[5] By contrast, we establish conditions for the optimality of price-cap regulation in a model in which costs are private and exogenous. We note further that our no-transfers assumption is critical: price-cap regulation is not optimal

function."

[4]In this respect, we follow the lead of Schmalensee (1989), who also examines a regulatory model with linear pricing schemes. (Schmalensee, 1989, p. 418) provides additional motivation for the practical relevance of linear pricing schemes in regulatory settings.

[5]For further discussion, see, for example, (Armstrong and Sappington, 2007, p. 1608) and the references cited therein.

in the standard Baron-Myerson model with transfers. Our findings thus indicate that this practical regulatory policy may perform not just well but optimally when a regulator faces a privately informed monopolist and transfers are infeasible.

We develop our findings for two different representations of ex post participation constraints. In the first representation, the regulator is constrained to ensure that the monopolist earns non-negative profit while providing positive output under all cost realizations. This no-shutdown representation may be motivated with reference to settings in which the monopolist provides essential services with poor substitution alternatives. The second representation expands the choice set for the regulator, by allowing that the regulator may choose a menu of permissible outputs such that some types produce zero output, incur no fixed cost, and thus earn a profit of zero. This representation thus allows that the regulator may "exclude" some types and may be motivated with reference to settings in which the monopolist provides an inessential service. We begin by establishing conditions for the optimality of price-cap regulation in the first setting, where exclusion is not allowed. We then show, perhaps surprisingly, that analogous characterizations for the optimality of price-cap regulation can then be directly obtained when exclusion is allowed, with the main difference being that the price cap may then be set at a lower level.

Our work is related to research on optimal delegation. The delegation literature begins with Holmstrom (1977), who considered a setting in which a principal faces a privately informed and biased agent and in which contingent transfers are infeasible. The principal then selects a set of permissible actions from the real line, and the agent selects his preferred action from that set after privately observing the state of nature.[6] A key goal in this literature has been to identify general conditions under which the principal optimally defines the permissible set as an interval. Alonso and Matouschek (2008) consider a setting with quadratic utility functions and provide necessary and sufficient conditions for interval delegation to be optimal. Extending the Lagrangian techniques of Amador et al. (2006), Amador and Bagwell (2013b) consider a general representation of the delegation problem and establish necessary and sufficient conditions for the optimality of interval delegation. Our analysis here builds on the Lagrangian methods used by Amador and Bagwell. A novel feature of the current paper is that the analysis is extended to include an ex post participation constraint.[7]

[6]A large literature follows Holmstrom's original work. See, for example, Amador et al. (2006), Ambrus and Egorov (2009), Armstrong and Vickers (2010), Frankel (2010), Martimort and Semenov (2006), Melumad and Shibano (1991) and Mylovanov (2008). Related themes also arise in repeated games with private information; see Athey et al. (2004) and Athey et al. (2005).

[7]Amador and Bagwell (2013a) also build on the Lagrangian methods used by Amador and Bagwell (2013b). The focus of Amador and Bagwell (2013a), however, is very different from that of the current paper. Amador and Bagwell (2013a) consider an optimal delegation problem with a two-dimensional action set, where one of the actions corresponds to "money burning," and they provide sufficient conditions under

We expect our methods will facilitate the application of optimal delegation theory to other settings in which participation constraints play an important role.

As Alonso and Matouschek (2008) explain, the monopoly regulation problem can be understood as an optimal delegation problem. In this context, when the regulator (i.e., the principal) uses price-cap regulation, the monopolist (i.e., the agent) is subjected to a rule, since prices above the cap are not allowed, but is also granted some discretion, since any price below the cap is permitted. Price-cap regulation can be understood as a form of interval delegation, where the maximal price is defined by the cap and the minimum price is defined by the monopoly price of the lowest-cost firm. As an application of their theoretical analysis of optimal delegation, Alonso and Matouschek (2008) study optimal regulation when costs are privately observed by the regulated firm and transfers are infeasible, and they also report conditions under which price-cap regulation is optimal. Our analysis differs from their interesting analysis in two main respects. First, Alonso and Matouschek assume that the monopolist produces regardless of its cost type, and they do not include a participation constraint in their analysis. Indeed, the price-cap solution that they derive would violate an ex post participation constraint, since the cap is below the marginal cost of the highest-cost firm. Second, Alonso and Matouschek assume that demand is linear and the regulator maximizes aggregate social surplus. Our sufficient conditions include this case, while also allowing for a participation constraint, both with and without the possibility of exclusion, and include as well more general demand functions and regulator objectives.[8] For the case without exclusion, we also note that the cap in our model is placed at a higher level and generates zero profit for the highest-cost firm.

The remainder of the paper is organized as follows. Section 2 sets up the regulator's problem, and Section 3 characterizes the optimal regulatory policy when attention is restricted to allocations that can be induced by caps. Section 4 then develops general sufficient conditions for the optimality of a cap allocation in the set of all allocations that satisfy incentive com-

which money burning expenditures are used in an optimal delegation contract. Building on work by Ambrus and Egorov (2009), Amador and Bagwell (2013a) also consider an application with an ex ante participation constraint under the assumption that ex ante (non-contingent) transfers are feasible. The participation constraint can then be addressed using standard methods. In the present paper, by contrast, the participation constraint must hold ex post and cannot be addressed using standard methods since transfers are infeasible.

[8] Alonso and Matouschek (2008) also argue that price cap regulation is optimal when demand exhibits constant elasticity and the regulator maximizes aggregate social welfare. As they explain, however, profits and welfare functions are no longer quadratic functions when demand exhibits constant elasticity, and so their results here apply only to the extent that the welfare and profit functions can be reasonably approximated using second-order Taylor series expansions. By constrast, an advantage of our Lagrangian approach is that it is not restricted to quadratic payoff functions. We thus directly analyze the model with constant elasticity demand, and indeed the sufficient conditions that we report for the optimality of price cap regulation allow for a wider range of regulator preferences when demand exhibits constant elasticity than when demand is linear. Using our sufficient conditions, other demand specifications may be considered as well.

patibility and participation constraints. Sections 5 and 6 consider settings with log demand and non-decreasing densities, respectively, and provide conditions under which the general sufficient conditions are satisfied. Section 6 examines the linear demand, constant elasticity demand and exponential demand examples in detail. Section 7 considers the case where exclusion of some types is feasible. Section 8 concludes.

2 The Regulator's Problem

In this section, we present our basic model and formally define the problem that confronts the regulator. We also identify the bias in the monopolist's unrestricted output choice.

Let $P(z)$ denote the inverse-demand function where z is the quantity demanded. We assume that the marginal cost of production is constant and given by γ. Let π be the quantity produced. The monopolist's profits are then given by

$$P(\pi)\pi - \gamma\pi - \sigma,$$

where $\sigma \geq 0$ is the fixed cost of production for the monopolist. We next define consumer surplus by

$$\int_0^{\pi} P(z)dz - P(\pi)\pi$$

Aggregate social surplus, for a given γ, is then given by the following:

$$\int_0^{\pi} P(z)dz - \gamma\pi - \sigma.$$

The marginal cost γ is private information to the monopolist and is distributed over the support $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ where $\overline{\gamma} > \underline{\gamma} \geq 0$ with a differentiable cumulative distribution function $F(\gamma)$.[9] The associated density, $f(\gamma) \equiv F'(\gamma)$, is strictly positive and differentiable. The production quantity π resides in the set Π, which is an interval of the real line with non-empty interior. The function $P(\pi)$ is well-defined and finite for all $\pi \in \Pi$. We assume that $\inf \Pi = 0$ and define $\overline{\pi}$ to be in the extended reals and such that $\overline{\pi} = \sup \Pi$.

We assume that the regulator has no access to transfers or taxes, and can only impose restrictions on the quantity produced by the monopolist. As discussed above, our no-transfers assumption means that the regulator cannot impose taxes or subsidies, and it implicitly

[9] We assume that the monopolist's fixed cost is commonly known. For the setting where the monopolist provides an essential service, so that the regulator must ensure a positive output for all types, all of our results hold when the monopolist is also privately informed about its fixed cost if σ is defined as the highest possible fixed cost level.

implies as well that the monopolist cannot use an access fee. We thus assume that the monopolist selects a uniform price, with the regulator determining the feasible menu of such prices through the selection of a feasible menu of quantities. We allow that the regulator's objective is to maximize a weighted social welfare function in which profits receive weight $\alpha \in (0,1]$. The regulator maximizes aggregate social surplus when $\alpha = 1$ and gives greater weight to consumer interests when $\alpha < 1$.

We envision the regulator as choosing a menu of permissible outputs, with the understanding that a monopolist with cost type γ selects its preferred output from this menu. Thus, if the regulator seeks to assign an output $\pi(\gamma)$ to a monopolist with type γ, then an incentive compatibility constraint must be satisfied. As well, if the regulator seeks a positive output from a monopolist with type γ, then type γ must earn more by producing $\pi(\gamma) > 0$ than by shutting down and avoiding the fixed cost of production, σ.

We consider two settings. In the first setting, the regulator is constrained to ensure that the monopolist provides positive output under all cost realizations. This no-shutdown setting may be motivated with reference to applications where the monopolist provides essential services with poor substitution alternatives.[10] The second setting, by contrast, expands the choice set for the regulator, by allowing that the regulator may choose a menu of permissible outputs such that some types produce zero output, incur no fixed cost, and thus earn a profit of zero. The second setting thus allows that the regulator may "exclude" some types. To motivate this case, we may consider applications in which the service that the monopolist provides is not essential. Our approach is to begin with the first setting, where exclusion is infeasible. After characterizing optimal regulation for this setting, we turn in Section 7 to consider optimal regulation when exclusion is feasible.

Rescaling the regulator's objective by the factor $\frac{1}{\alpha}$, we thus represent the *regulator's problem* as follows:

$$\max_{\pi:\Gamma\to\Pi} \int_\Gamma \left(-\gamma\pi(\gamma) + P(\pi(\gamma))\pi(\gamma) - \sigma + \frac{1}{\alpha}\left(\int_0^{\pi(\gamma)} P(z)dz - P(\pi(\gamma))\pi(\gamma) \right) \right) dF(\gamma)$$

subject to:

$$\gamma \in \arg\max_{\tilde{\gamma}\in\Gamma} -\gamma\pi(\tilde{\gamma}) + P(\pi(\tilde{\gamma}))\pi(\tilde{\gamma}), \text{ for all } \gamma \in \Gamma$$

$$\sigma \leq -\gamma\pi(\gamma) + P(\pi(\gamma))\pi(\gamma), \text{ for all } \gamma \in \Gamma$$

[10]See (Laffont and Tirole, 1993, pp. 62-3, 493-4) for related discussion. In addition, even if shutdown were allowed, optimal regulation would never result in shut down if social surplus were sufficiently high even for the highest cost firm, as Baron and Myerson (1982) establish for a model with transfers and as we confirm in Section 7 for our model without tranfers.

The first constraint is the incentive compatibility constraint, while the second constraint is the ex post participation or individual rationality (IR) constraint. Notice that when $\sigma > 0$ the IR constraint requires that every $\gamma \in \Gamma$ produces a positive output, $\pi(\gamma) > 0$.[11] An allocation is feasible if it satisfies both of these constraints.

Ignoring the participation constraint, this application fits into the framework developed by Amador and Bagwell (2013b).[12] In particular, using the notation of that paper, we may define the regulator's problem as

$$\max_{\pi:\Gamma\to\Pi} \int_\Gamma w(\gamma, \pi(\gamma))dF(\gamma) \qquad \text{subject to:}$$
$$\gamma \in \arg\max_{\tilde{\gamma}\in\Gamma} -\gamma\pi(\tilde{\gamma}) + b(\pi(\tilde{\gamma})), \text{ for all } \gamma \in \Gamma$$
$$\sigma \leq -\gamma\pi(\gamma) + b(\pi(\gamma)), \text{ for all } \gamma \in \Gamma$$

where

$$\begin{aligned} w(\gamma, \pi) &= -\gamma\pi + b(\pi) - \sigma + \frac{1}{\alpha}v(\pi), \\ b(\pi) &= P(\pi)\pi, \\ v(\pi) &= \int_0^\pi P(z)dz - P(\pi)\pi, \end{aligned}$$

Notice that, in the current application, $b(\pi)$ defines the total revenue for the monopolist, $v(\pi)$ represents consumer surplus, and $w(\gamma, \pi)$ is the regulator's welfare function.

We impose the following assumptions:

Assumption 1. *The inverse demand function is such that $P(\pi)$ is twice-continuously differentiable with $P'(\pi) < 0 < P(\pi)$ for all $\pi \in \Pi$. We assume that $b''(\pi) < 0$ and $w_{\pi\pi}(\gamma, \pi) \leq 0$ for all $\pi \in \Pi$ and $\gamma \in \Gamma$. We assume that $P(0) > \overline{\gamma}$ when $0 \in \Pi$, and that $\lim_{\pi\to 0} P(\pi) > \overline{\gamma}$ when $0 \notin \Pi$.*

[11] The assumption of positive output is also embedded in our representation of the objective function, which specifies that the avoidable fixed cost σ is always incurred. In the special case where $\sigma = 0$, this specification is immaterial, and the regulator's problem is consistent with no production for some types. For this special case, our results in Sections 2-6 directly establish sufficient conditions under which shutdown is not optimal even when feasible. (Consideration of shutdown for this special case also requires that $0 \in \Pi$.) For the general case where $\sigma > 0$, we postpone consideration of feasible shutdown until Section 7.

[12] One further difference is that the flexible allocation (i.e., the ideal allocation for the monopolist or agent) is upward sloping in the framework of Amador and Bagwell (2013b) while as we discuss below the flexible allocation is downward sloping in the current setting. This difference can be easily addressed with a straightforward notational modification, in which π is re-defined as the extent to which actual output falls short of some upper bound.

Under Assumption 1, we obtain that

$$\begin{aligned} b'(\pi) &= P(\pi) + \pi P'(\pi) \\ v'(\pi) &= -\pi P'(\pi) > 0 \text{ for all } \pi > 0 \\ w_\pi(\gamma, \pi) &= -\gamma + b'(\pi) + \frac{1}{\alpha} v'(\pi) \\ &= -\gamma + P(\pi) + \pi P'(\pi) - \frac{1}{\alpha} \pi P'(\pi). \end{aligned}$$

Similarly, using Assumption 1, second derivatives take the following forms and signs:

$$\begin{aligned} b''(\pi) &= P''(\pi)\pi + 2P'(\pi) < 0 \text{ for all } \pi \in \Pi \\ v''(\pi) &= -[P''(\pi)\pi + P'(\pi)] \\ w_{\pi\pi}(\gamma, \pi) &= b''(\pi) + \frac{1}{\alpha} v''(\pi) \\ &= P''(\pi)\pi + 2P'(\pi) - \frac{1}{\alpha}[P''(\pi)\pi + P'(\pi)] \leq 0 \text{ for all } \pi \in \Pi. \end{aligned}$$

Notice that $P'(\pi) < 0$ implies that w is strictly concave when $\alpha = 1$. Importantly, we make no assumption as regards the sign of $v''(\pi)$. If marginal revenue is steeper than demand (i.e., $b''(\pi) < P'(\pi)$), then $v''(\pi) > 0$. This condition holds if the demand function $z(P)$ is log-concave but fails otherwise. For example, as we discuss in greater detail below, $v''(\pi) > 0$ when demand is linear, and $v''(\pi) < 0$ when demand exhibits constant elasticity.

The flexible allocation. Assuming that the participation constraint is satisfied, we let $\pi_f(\gamma)$ denote the allocation that a monopolist would choose were it unrestricted by a regulator. The monopolist's flexible allocation is thus defined as

$$\pi_f(\gamma) = \arg\max_{\pi \in \Pi} \big\{ -\gamma\pi + b(\pi) \big\}.$$

The flexible allocation is simply the monopoly output as a function of the monopolist's cost type. The associated first-order condition is given by

$$b'(\pi) - \gamma = 0.$$

Notice that $b'(0) = P(0) > \overline{\gamma}$, and so $\pi_f(\overline{\gamma}) > 0$. We assume further that $\pi_f(\underline{\gamma}) < \overline{\pi}$ so that $\pi_f(\gamma)$ is interior for all $\gamma \in \Gamma$. It follows that $\pi_f(\gamma)$ is differentiable, with $\pi_f'(\gamma) = 1/b''(\pi_f(\gamma)) < 0$. Note as well that $P(\pi_f(\gamma)) > \gamma$ and thus $-\gamma\pi_f(\gamma) + b(\pi_f(\gamma)) > 0$ for all $\gamma \in \Gamma$.

Given interiority, we also have the following relationships:

$$\begin{aligned} w_\pi(\gamma, \pi_f(\gamma)) &= \frac{1}{\alpha} v'(\pi_f(\gamma)) \\ &= -\frac{1}{\alpha} P'(\pi_f(\gamma))\pi_f(\gamma) \\ &= \frac{1}{\alpha}[P(\pi_f(\gamma)) - \gamma] > 0 \text{ for all } \gamma \in \Gamma. \end{aligned}$$

Thus, the regulator model is characterized by downward or *negative bias*: the agent's (i.e., the monopolist's) preferred π is too low from the principal's (i.e., the regulator's) perspective. This suggests the possibility of a solution that imposes a lower bound on π for higher types (or equivalently a cap on the price for higher types). We explore this possibility below.

3 Optimality Within the Set of Cap Allocations

In this section, we solve the regulator's problem under the further restriction that the regulator is restricted to choose among cap allocations. Paying particular attention to the participation constraint, we then identify a candidate for an optimal solution among all feasible allocations.

Let us define a cap allocation as follows:

Definition 1. A *cap allocation* indexed by γ_c is an allocation π_c given by

$$\pi_c(\gamma) = \begin{cases} \pi_f(\gamma) & ; \gamma \in [\underline{\gamma}, \gamma_c] \\ \pi_f(\gamma_c) & ; \gamma \in (\gamma_c, \overline{\gamma}] \end{cases}$$

It is straightforward to confirm that a cap allocation is incentive compatible. Notice also that the allocation $\pi_c(\gamma)$ actually defines a floor rather than a cap. We refer to this allocation as a cap allocation, since it corresponds to a cap on permissible prices and links thereby with the literature on price-cap regulation.

We define an *optimal simple cap allocation* to be an optimal cap allocation when the participation constraint is ignored. The following lemma provides a necessary condition for an optimal simple cap allocation:

Lemma 1. *The cap allocation $\gamma_c < \overline{\gamma}$ is an optimal simple cap allocation only if*

$$\int_{\gamma_c}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_c)) dF(\gamma) = 0$$

We assume that there is a unique $\gamma_c < \overline{\gamma}$ that solves the first order condition in Lemma 1. In the absence of a participation constraint, we could use results from Amador and Bagwell (2013b) and establish a general set of environments under which the optimal simple cap allocation is optimal over the full class of incentive compatible allocations. As we now argue, however, the presence of a participation constraint implies that the optimal simple cap allocation is no longer feasible.

The basic point can be understood using a graphical argument. The graph on the right in Figure 1 illustrates the optimal simple cap allocation in bold. This allocation is illustrated relative to the flexible allocation, $\pi_f(\gamma)$, and the regulator's ideal allocation, $\pi_e(\gamma)$, which we define as the allocation that maximizes $w(\gamma, \pi)$. Notice that $\pi_e(\gamma)$ is downward sloping and that $\pi_e(\gamma) > \pi_f(\gamma)$, where the inequality reflects the aforementioned downward bias. For given γ, $\pi_e(\gamma)$ induces a price equal to marginal cost (i.e., $P(\pi_e(\gamma)) = \gamma$) when $\alpha = 1$. When $\alpha < 1$, the regulator's ideal allocation entails even higher quantities and thus drives price below marginal cost. The optimal simple cap allocation is such that the cap is ideal for the regulator on average for affected types (i.e., for $\gamma \geq \gamma_c$). The graph on the left in Figure 1 illustrates the same information in terms of the induced prices, which are also depicted in bold. As this graph illustrates, the optimal simple cap allocation places the price cap at a level that is ideal for the principal on average for affected types. This graph also suggests that the participation constraint is violated for the highest types when the optimal simple cap allocation is used. For type $\overline{\gamma}$, the optimal price cap lies below the regulator's ideal price, $P(\pi_e(\overline{\gamma}))$, which equals $\overline{\gamma}$ when $\alpha = 1$ and is less than $\overline{\gamma}$ when $\alpha < 1$. The optimal price cap is thus strictly below $\overline{\gamma}$; hence, since the fixed cost σ is non-negative, the participation constraint must fail for the highest-cost type when the optimal simple cap allocation is used.

To develop this point with full details, let $H(\gamma) = b(\pi_c(\gamma)) - \gamma\pi_c(\gamma)$. The participation constraint is equivalent to $H(\gamma) \geq \sigma$ for all $\gamma \in \Gamma$. Note that $H(\gamma)$ is continuous, and that

$$H'(\gamma) = \begin{cases} (b'(\pi_c(\gamma)) - \gamma)\pi_c'(\gamma) - \pi_c(\gamma) = -\pi_f(\gamma) < 0 & ; \gamma \in (\underline{\gamma}, \gamma_c) \\ -\pi_f(\gamma_c) < 0 & ; \gamma \in (\gamma_c, \overline{\gamma}) \end{cases}$$

Hence, H is strictly decreasing. So to check whether the participation constraint holds it suffices to check whether $H(\overline{\gamma}) \geq \sigma$, that is, whether the allocation is individually rational for the highest cost type. We have the following lemma:

Lemma 2. *The optimal simple cap allocation $\gamma_c < \overline{\gamma}$ violates the participation constraint.*

Proof: Note that $\int_{\gamma_c}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_c))dF(\gamma) = 0$. Using $w_{\pi\gamma}(\gamma, \pi) = -1 < 0$ and $\gamma_c < \overline{\gamma}$, it follows that $w_\pi(\overline{\gamma}, \pi_f(\gamma_c)) < 0$. Next, observe that $w_\pi(\overline{\gamma}, \pi_f(\gamma_c)) = -\overline{\gamma} + b'(\pi_f(\gamma_c)) + \frac{1}{\alpha}v'(\pi_f(\gamma_c)) = -\overline{\gamma} + \gamma_c + \frac{1}{\alpha}[P(\pi_f(\gamma_c)) - \gamma_c] = \frac{1}{\alpha}[P(\pi_f(\gamma_c)) - \overline{\gamma}] + \frac{1-\alpha}{\alpha}(\overline{\gamma} - \gamma_c)$. Given $\gamma_c < \overline{\gamma}$

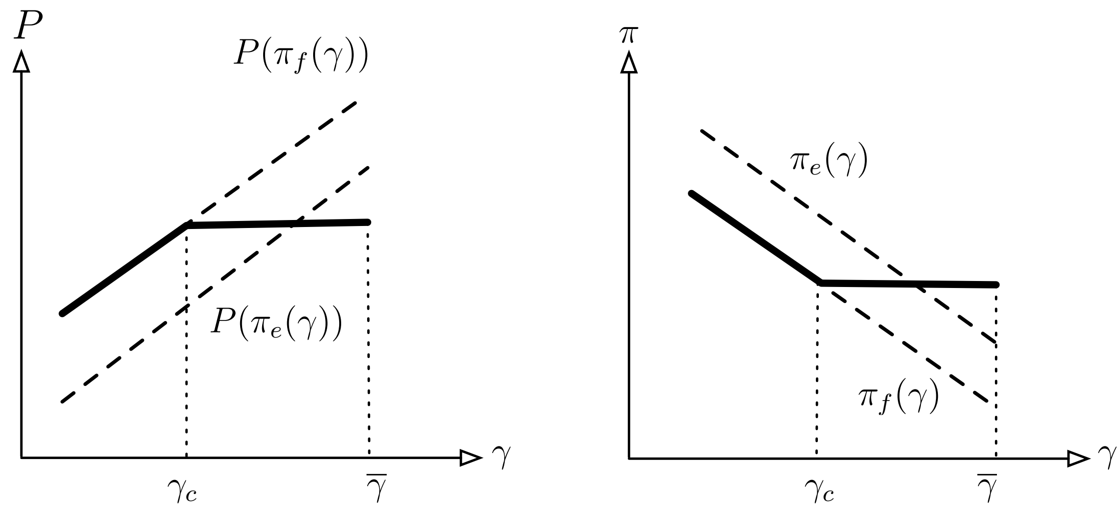

Figure 1: Optimal Simple Cap Allocation Fails IR.

and $\alpha \in (0,1]$, we conclude that $P(\pi_f(\gamma_c)) - \overline{\gamma} < 0$; thus, since $\sigma \geq 0$, we may deduce that the participation constraint is violated for the highest types. □

Based on the above, we are led to consider the "closest" cap allocation to the optimal simple cap allocation that satisfies the participation constraint. To define this allocation, we begin by imposing the following assumption:

Assumption 2. $-\overline{\gamma}\pi_f(\underline{\gamma}) + b(\pi_f(\underline{\gamma})) < \sigma < -\overline{\gamma}\pi_f(\overline{\gamma}) + b(\pi_f(\overline{\gamma}))$.

This assumption implies that the highest-cost monopolist could earn positive profit when selecting its monopoly or flexible output, $\pi_f(\overline{\gamma})$, but would earn negative profit when selecting the higher output that corresponds to the monopoly or flexible output for the lowest-cost monopoly, $\pi_f(\underline{\gamma})$. There must then exist an intermediate cost type, $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ with $\gamma_H > \gamma_c$, such that the highest-cost monopolist would earn zero profit (price at its average cost) when selecting the monopoly or flexible output for this intermediate type, $\pi_f(\gamma_H)$. We thus have the following definition:

Definition 2. Let $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ be such that $-\overline{\gamma}\pi_f(\gamma_H) + b(\pi_f(\gamma_H)) = \sigma$.

Alternatively, γ_H can be defined as the value such that

$$P(\pi_f(\gamma_H)) = \overline{\gamma} + \sigma/\pi_f(\gamma_H).$$

Figure 2 illustrates the demand function, the average cost of the monopolist with cost type $\overline{\gamma}$, and the resulting determination of $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$.

We are now prepared to define the *IR-cap allocation*:

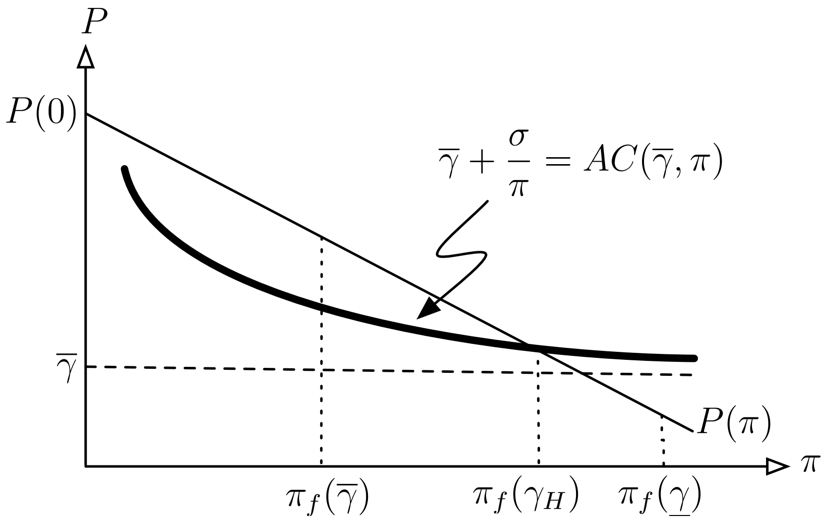

Figure 2: Determination of γ_H.

Definition 3. The *IR-cap allocation* is given by

$$\pi_{IR}(\gamma) = \begin{cases} \pi_f(\gamma) & ; \gamma \in [\underline{\gamma}, \gamma_H] \\ \pi_f(\gamma_H) & ; \gamma \in (\gamma_H, \overline{\gamma}] \end{cases}$$

The IR-cap allocation is illustrated in Figure 3 for the case in which $\sigma > 0$. The graph on the right illustrates the IR-cap allocation, while the graph on the left captures the price allocation that is induced by the IR-cap allocation.

As above, we may establish that under the IR-cap allocation the utility of the firm declines strictly with the firm's cost type. Thus, in the IR-cap allocation, $-\gamma\pi(\gamma) + b(\pi(\gamma)) > \sigma$ for all $\gamma < \overline{\gamma}$. The IR-cap allocation thus satisfies the participation constraint and is the closest cap allocation to the optimal simple allocation that does so. The IR-cap allocation is a candidate for an optimal allocation among all feasible allocations.

4 Sufficient Conditions

We now provide general sufficient conditions for the IR-cap allocation to be optimal among all feasible allocations. We first develop some intuition using a simple graph. We then describe the general approach and finally state and prove our general sufficiency result.

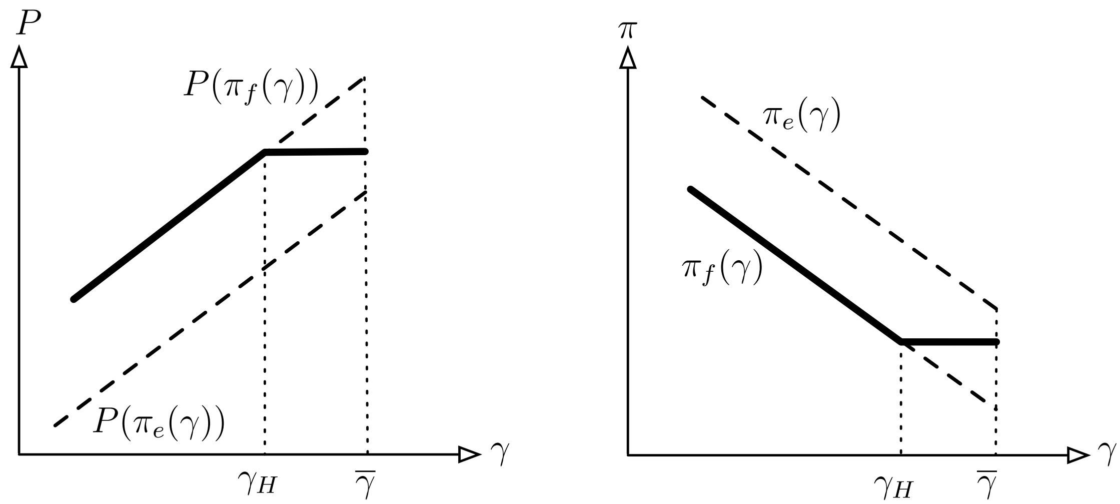

Figure 3: IR-cap allocation (with $\sigma > 0$).

4.1 Intuition

To develop some intuition, we consider alternatives to the IR-cap allocation. If the IR-cap allocation is to be optimal among all feasible allocations, then in particular it must be preferred by the regulator to alternative feasible allocations that are generated by "drilling holes" in the flexible part of the allocation. Figure 4 illustrates one such alternative allocation, in which output levels between $\pi_1 \equiv \pi_f(\gamma_1)$ and $\pi_2 \equiv \pi_f(\gamma_2)$ are prohibited and where $\underline{\gamma} < \gamma_1 < \gamma_2 < \gamma_H$. There then exists a unique type $\widetilde{\gamma} \in (\gamma_1, \gamma_2)$ that is indifferent between π_1 and π_2. The alternative allocation thus induces a "step" at $\widetilde{\gamma}$, with the allocation π_1 selected by $\gamma \in [\gamma_1, \widetilde{\gamma})$ and the allocation π_2 selected by $\gamma \in [\widetilde{\gamma}, \gamma_2]$, where for simplicity we place type $\widetilde{\gamma}$ with the higher types.

In comparison to the IR-cap allocation, the alternative allocation has advantages and disadvantages. First, the alternative allocation generates output choices for $\gamma \in [\gamma_1, \widetilde{\gamma})$ that are closer to the the regulator's ideal choices for such types; however, the alternative allocation also results in output choices for $\gamma \in [\widetilde{\gamma}, \gamma_2]$ that are further from the regulator's ideal choices for such types. These observations suggest that a non-decreasing density should work in favor of the IR-cap allocation, since the disadvantageous features of the alternative allocation then receive greater probability weight in the regulator's expected welfare. Second, the alternative allocation increases the variance of the induced allocation around $\pi_f(\gamma)$ over the interval $[\gamma_1, \gamma_2]$. This effect brings into consideration the relative magnitudes of $\frac{1}{\alpha}v''(\pi)$ and $b''(\pi)$, where the latter determines the slope of $\pi_f(\gamma)$. In particular, if $v(\pi)$ is concave, then we expect that the variance effect works in favor of the IR-cap allocation, since the regulator would then not welcome an increase in variance. If instead $v(\pi)$ is convex, then

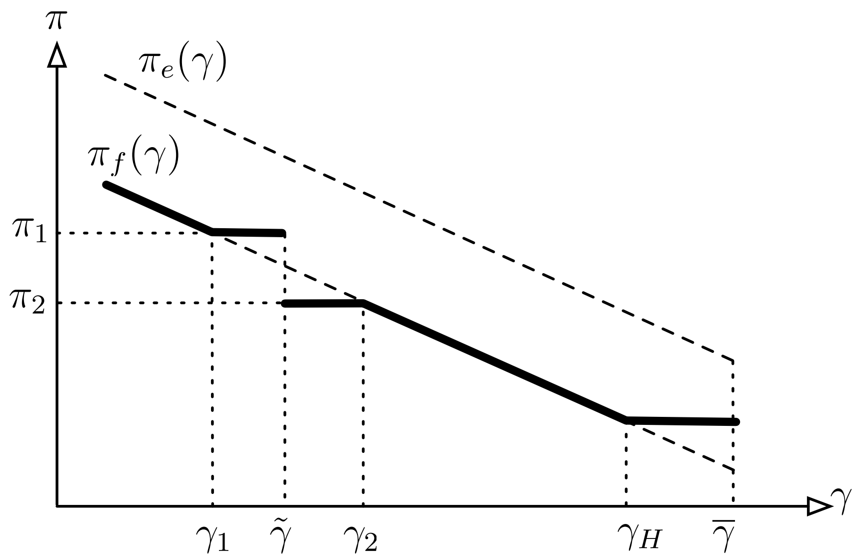

Figure 4: Drilling a hole (with $\sigma > 0$).

the regulator would benefit from the greater variance afforded by the alternative allocation, with the overall benefit to the regulator being larger when α is smaller. We may thus anticipate that the IR-allocation could remain optimal when $v(\pi)$ is convex, provided that the density rises fast enough, α is sufficiently large and/or $b''(\pi)$ is large in absolute value (so that $\pi_f(\gamma)$ is flat, in which case steps add little variation).

The intuitive discussion presented here considers only a subset of feasible alternative allocations. A feasible allocation must satisfy incentive compatibility and IR constraints. In our no-transfer setting, the incentive compatibility constraint implies that an allocation must be given by the flexible allocation over any interval for which the allocation is continuous and strictly decreasing; however, an incentive compatible allocation may include many points of discontinuity (steps), where any such point hurdles the flexible allocation as illustrated in Figure 4.[13] For type $\overline{\gamma}$, the IR constraint holds with equality at allocations $\pi_f(\gamma_H)$ and π', where $\pi' < \pi_f(\overline{\gamma})$ is defined so that type $\overline{\gamma}$ is indifferent between $\pi_f(\gamma_H)$ and π'; thus, the IR constraint for type $\overline{\gamma}$ is satisfied provided that the allocation for this type resides in the interval $[\pi', \pi_f(\gamma_H)]$. As a general matter, it is not obvious that the IR constraint must bind for type $\overline{\gamma}$ in our no-transfer setting, since the allocation for this type may be positioned so as to favorably affect allocations for lower types. If the IR-cap allocation is to solve the regulator's problem, it must be superior to all feasible alternative allocations. To develop a formal counterpart to the intuitive discussion above and consider the full set of feasible alternative allocations, we move next to our formal analysis.

[13]For further discussion, see Melumad and Shibano (1991).

4.2 General Approach

As our preceding discussion clarifies, the problem of finding a solution to the regulator's problem is non-trivial, due in part to the prevalence of discontinuous allocations in the feasible choice set. Most of the preceding delegation literature has thus added structure to the problem by assuming quadratic payoff functions and (often) uniform distributions. Our approach here is instead to follow the "guess-and-verify" approach of Amador and Bagwell (2013b). As they argue, once a candidate solution is identified, powerful Lagrangrian methods can be applied to establish the optimality of the candidate in general settings. We must extend the Amador-Bagwell analysis in an important way, however, since the present problem includes an ex post participation constraint.

We proceed as follows. First, we re-state the regulator's problem by expressing the incentive compatibility constraints in their standard form as an integral equation and a monotonicity requirement:[14]

$$\max_{\pi:\Gamma\to\Pi}\int_\Gamma w(\gamma,\pi(\gamma))dF(\gamma) \qquad \text{subject to:}$$
$$-\gamma\pi(\gamma)+b(\pi(\gamma))-\sigma-\int_\gamma^{\overline{\gamma}}\pi(\tilde{\gamma})d\tilde{\gamma}=\overline{U},\ \text{ for all } \gamma\in\Gamma$$
$$\pi(\gamma) \text{ non-increasing, for all } \gamma\in\Gamma$$
$$\sigma\leq -\gamma\pi(\gamma)+b(\pi(\gamma)),\ \text{ for all } \gamma\in\Gamma$$

where $\overline{U}\equiv -\overline{\gamma}\pi(\overline{\gamma})+b(\pi(\overline{\gamma}))-\sigma$ is the profit enjoyed by the monopolist with the highest possible cost type.[15]

Next, we follow Amador and Bagwell (2013b) and re-write the incentive constraints as a set of two inequalities and embed the monotonicity constraint in the choice set of $\pi(\gamma)$. With the choice set for $\pi(\gamma)$ defined as $\Phi\equiv\{\pi|\pi:\Gamma\to\Pi;\text{ and } \pi \text{ non-increasing}\}$, the regulator's

[14] See, for example, Milgrom and Segal (2002).

[15] It is instructive here to compare our regulator's problem, in which transfers are unavailable, with the standard (Baron-Myerson) framework in which transfers are available. In the solution approach for the standard framework, the integral equation is substituted into the objective, the IR constraint is shown to bind for the highest type, the IR constraint for the highest type is substituted into the objective, and the resulting objective is then maximized pointwise. If the solution satisfies the monotonicity constraint, then the problem is solved. By constrast, in our no-transfers setting, we cannot substitute the integral equation into the objective, since we do not have a remaining transfer instrument with which to ensure that the solution of the resulting optimization problem satisfies the integral equation. For the same reason, we cannot substitute the IR constraint for the highest type into the objective. Indeed, as a general matter, when transfers are unavailable it is no longer obvious that the IR constraint for the highest type must bind.

problem may now be stated in final form as follows:

$$\max_{\pi\in\Phi}\int_\Gamma w(\gamma,\pi(\gamma))dF(\gamma) \qquad \text{subject to:} \tag{P}$$

$$\gamma\pi(\gamma)-b(\pi(\gamma))+\sigma+\int_\gamma^{\overline{\gamma}}\pi(\tilde{\gamma})d\tilde{\gamma}+\overline{U}\leq 0, \text{ for all } \gamma\in\Gamma \tag{1}$$

$$-\gamma\pi(\gamma)+b(\pi(\gamma))-\sigma-\int_\gamma^{\overline{\gamma}}\pi(\tilde{\gamma})d\tilde{\gamma}-\overline{U}\leq 0, \text{ for all } \gamma\in\Gamma \tag{2}$$

$$\gamma\pi(\gamma)-b(\pi(\gamma))+\sigma\leq 0, \text{ for all } \gamma\in\Gamma \tag{3}$$

We now describe the general approach of the proof. The proof employs Theorem 1 in Appendix B of Amador and Bagwell (2013b), which utilizes a Lagrangian method. To utilize this theorem, we must construct non-decreasing Lagrangian multiplier functions for the program's constraints such that the IR-cap allocation maximizes the resulting Lagrangian over the choice set Φ and the IR-cap allocation and constructed multipliers together satisfy complementary slackness. To verify that the IR-cap allocation indeed maximizes the resulting Lagrangian over $\pi\in\Phi$, we build on Amador et al. (2006) and Amador and Bagwell (2013b) and express the first order conditions for maximizing the Lagrangian over the set of non-increasing functions, Φ. As in the problem considered by Amador and Bagwell (2013b), a difficulty is that the Lagrangian is not necessarily concave in π. We thus choose our Lagrangian multiplier functions carefully, so that the resulting Lagrangian is concave in π and the first order conditions are sufficient for maximizing the resulting Lagrangian. Differently than Amador and Bagwell (2013b), our present problem includes participation constraints, which as discussed affects the proposed solution candidate and also requires the construction of an additional non-decreasing multiplier function. The conditions under which we can achieve all of these steps then determine the sufficient conditions for the IR-cap allocation to solve the regulator's problem. As we will see in the next section, these conditions can be interpreted in terms of the intuition presented at the start of this section.

4.3 Result and Proof

To present our result, we require a couple of definitions. Let

$$G(\gamma)\equiv\frac{1}{\gamma-\gamma_H}\int_{\gamma_H}^{\gamma}w_\pi(\tilde{\gamma},\pi_f(\gamma_H))f(\tilde{\gamma})d\tilde{\gamma}-\kappa(1-F(\gamma)), \tag{4}$$

where following Amador and Bagwell (2013b) κ is a realtive concavity parameter defined as

$$\kappa = \min_{(\gamma,\pi)\in\Gamma\times\Pi} \left\{ \frac{w_{\pi\pi}(\gamma,\pi)}{b''(\pi)} \right\}.$$

We may now state our general sufficiency result as follows:

Proposition 1. *(Sufficient Conditions) Let $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ be defined as in Definition 2. If*

(i) $G(\gamma) \leq G(\overline{\gamma})$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$,

(ii) $G(\overline{\gamma}) \geq 0$, and

(iii) $\kappa F(\gamma) + w_\pi(\gamma, \pi_f(\gamma))f(\gamma)$ is non-decreasing, for all $\gamma \in [\underline{\gamma}, \gamma_H]$,

for G as given by (4), then an optimal solution to the regulator's problem is achieved with the IR-cap allocation $\pi_{IR}(\gamma)$ such that $\pi_{IR}(\gamma) = \pi_f(\gamma)$ for all $\gamma \in [\underline{\gamma}, \gamma_H)$ and $\pi_{IR}(\gamma) = \pi_f(\gamma_H)$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$.

Proof: Let $\Lambda_1(\gamma)$ and $\Lambda_2(\gamma)$ denote the (cumulative) multiplier functions associated with the two inequalities that define the incentive compatibility constraints in the final form of the regulator's problem. The multiplier functions $\Lambda_1(\gamma)$ and $\Lambda_2(\gamma)$ are restricted to be non-decreasing. Letting $\Lambda(\gamma) \equiv \Lambda_1(\gamma) - \Lambda_2(\gamma)$, we can write the Lagrangian as

$$\mathcal{L} = \int_\Gamma w(\gamma, \pi(\gamma))dF(\gamma) - \int_\Gamma \left(\int_\gamma^{\overline{\gamma}} \pi(\tilde{\gamma})d\tilde{\gamma} + \overline{U} + \gamma\pi(\gamma) - b(\pi(\gamma)) + \sigma \right) d\Lambda(\gamma) + \int_\Gamma \Big(-\gamma\pi(\gamma) + b(\pi(\gamma)) - \sigma \Big) d\Psi(\gamma),$$

where $\Psi(\gamma)$ is the multiplier for the ex post participation constraints. $\Psi(\gamma)$ is also restricted to be non-decreasing.

Our proposed multipliers take the following specific forms:

$$\Lambda(\gamma) = \begin{cases} 0 & ; \gamma = \underline{\gamma} \\ w_\pi(\gamma, \pi_f(\gamma))f(\gamma) & ; \gamma \in (\underline{\gamma}, \gamma_H) \\ A + \kappa(1 - F(\gamma)) & ; \gamma \in [\gamma_H, \overline{\gamma}] \end{cases}$$

and

$$\Psi(\gamma) = \begin{cases} 0 & ; \gamma \in [\underline{\gamma}, \overline{\gamma}) \\ A & ; \gamma = \overline{\gamma} \end{cases}$$

where

$$A = \frac{1}{\overline{\gamma} - \gamma_H} \int_{\gamma_H}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_H)) f(\gamma) d\gamma. \tag{5}$$

We show below that the hypothesis of Proposition 1 guarantees that $R(\gamma) \equiv \kappa F(\gamma) + \Lambda(\gamma)$ is non-decreasing; thus, we may write $\Lambda(\gamma)$ as the difference between two non-decreasing functions, $\Lambda_1(\gamma) = R(\gamma)$ and $\Lambda_2(\gamma) = \kappa F(\gamma)$.[16] We also require that $A \geq 0$ as Φ must be non-decreasing. We establish this inequality below.

We note that the IR-cap allocation together with the proposed multipliers satisfy complementary slackness. The incentive compatibility constraints bind under the IR-cap allocation, and $\Psi(\gamma)$ is constructed to be zero whenever the participation constraint holds with slack.

When these multipliers are used, the resulting Lagrangian is

$$\mathcal{L} = \int_\Gamma w(\gamma, \pi(\gamma)) dF(\gamma) - \int_\Gamma \left(\int_\gamma^{\overline{\gamma}} \pi(\tilde{\gamma}) d\tilde{\gamma} + \overline{U} + \gamma\pi(\gamma) - b(\pi(\gamma)) + \sigma \right) d\Lambda(\gamma) + \left(-\overline{\gamma}\pi(\overline{\gamma}) + b(\pi(\overline{\gamma})) - \sigma \right) A$$

Recalling the definition of $\overline{U}$ and using $\Lambda(\underline{\gamma}) = 0$ and $\Lambda(\overline{\gamma}) = A$, we can then write the Lagrangian as

$$\mathcal{L} = \int_\Gamma w(\gamma, \pi(\gamma)) dF(\gamma) - \int_\Gamma \left(\int_\gamma^{\overline{\gamma}} \pi(\tilde{\gamma}) d\tilde{\gamma} + \gamma\pi(\gamma) - b(\pi(\gamma)) + \sigma \right) d\Lambda(\gamma)$$

Integrating the Lagrangian by parts we get[17]

$$\mathcal{L} = \int_\Gamma \Big(w(\gamma, \pi(\gamma)) f(\gamma) - \Lambda(\gamma)\pi(\gamma) \Big) d\gamma + \int_\Gamma \Big(-\gamma\pi(\gamma) + b(\pi(\gamma)) - \sigma \Big) d\Lambda(\gamma) \tag{6}$$

Let us now consider the concavity of the Lagrangian. Using (6), we may re-write the

[16] For our analysis, only the difference between $\Lambda_1(\gamma)$ and $\Lambda_2(\gamma)$ matters, and so we need only show that there exists two non-decreasing functions, $\Lambda_1(\gamma)$ and $\Lambda_2(\gamma)$, whose difference delivers $\Lambda(\gamma)$.

[17] Observe that $h(\gamma) \equiv \int_\gamma^{\overline{\gamma}} \pi(\tilde{\gamma}) d\tilde{\gamma}$ exists (as π is bounded and measurable by monotonicity) and is absolutely continuous. Observe as well that $\Lambda(\gamma) \equiv \Lambda_1(\gamma) - \Lambda_2(\gamma)$ is a function of bounded variation, as it is the difference between two non-decreasing and bounded functions. We may thus conclude that $\int_{\underline{\gamma}}^{\overline{\gamma}} h(\gamma) d\Lambda(\gamma)$ exists (it is the Riemman-Stieltjes integral), and integration by parts can be done as follows: $\int_{\underline{\gamma}}^{\overline{\gamma}} h(\gamma) d\Lambda(\gamma) = h(\overline{\gamma})\Lambda(\overline{\gamma}) - h(\underline{\gamma})\Lambda(\underline{\gamma}) - \int_{\underline{\gamma}}^{\overline{\gamma}} \Lambda(\gamma) dh(\gamma)$. Given that $h(\gamma)$ is absolutely continuous, we can replace $dh(\gamma)$ with $-\pi(\gamma) d\gamma$.

Lagrangian as

$$\begin{aligned}\mathcal{L} &= \int_\Gamma \Big(w(\gamma,\pi(\gamma)) - \kappa(-\gamma\pi(\gamma) + b(\pi(\gamma)) - \sigma)\Big) f(\gamma)d\gamma - \int_\Gamma \Lambda(\gamma)\pi(\gamma)d\gamma \\ &\quad + \int_\Gamma \Big(-\gamma\pi(\gamma) + b(\pi(\gamma)) - \sigma\Big) d(\kappa F(\gamma) + \Lambda(\gamma))\end{aligned}$$

From the definition of κ, $w(\gamma,\pi(\gamma)) - \kappa b(\pi(\gamma))$ is concave in $\pi(\gamma)$. We may thus conclude that the Lagrangian is concave in $\pi(\gamma)$ if

$$\kappa F(\gamma) + \Lambda(\gamma)$$

is non-decreasing for all $\gamma \in [\underline{\gamma}, \overline{\gamma})$. Using the constructed $\Lambda(\gamma)$ and referring to part (iii) of Proposition 1, we see that $\kappa F(\gamma) + \Lambda(\gamma)$ is non-decreasing for all $\gamma \in [\underline{\gamma}, \overline{\gamma})$ if the jumps in $\Lambda(\gamma)$ at $\underline{\gamma}$ and γ_H are non-negative. We verify these jumps are indeed non-negative below.

We now show that the IR-cap allocation maximizes the Lagrangian. To this end, we use the sufficiency part of Lemma A.2 in Amador et al. (2006), which concerns the maximization of concave functionals on a convex cone. If $\Pi = [0,\infty)$, then our choice set Φ is a convex cone. Following Amador and Bagwell (2013b), if instead $\Pi = (0,\overline{\pi})$ with $\overline{\pi}$ possibly finite and $\pi_f(\gamma)$ interior (i.e., $0 < \pi_f(\overline{\gamma})$ and $\pi_f(\underline{\gamma}) < \overline{\pi}$), then it is straightforward to extend b and w to the entire non-negative ray of the real line and apply Lemma A.2 to the extended Lagrangian with the choice set $\widehat{\Phi} \equiv \{\pi | \pi : \Gamma \to \Re_+;\text{ and } \pi \text{ non-increasing}\}$. Following the arguments in Amador and Bagwell (2013b), we can then establish that the IR-cap allocation maximizes the Lagrangian if the Lagrangian is concave and the following first order conditions hold:

$$\begin{aligned}\partial\mathcal{L}(\pi_{IR};\pi_{IR}) &= 0 \\ \partial\mathcal{L}(\pi_{IR};x) &\leq 0 \text{ for all } x \in \widehat{\Phi},\end{aligned}$$

where $\partial\mathcal{L}(\pi_{IR};x)$ is the Gateaux differential of the Lagrangian in (6) in the direction x.[18] Importantly, the Lagrangian in (6) is evaluated using our constructed multiplier functions.

Taking the Gateaux differential of the Lagrangian in (6) in direction $x \in \widehat{\Phi}$, we get[19]

[18] Given a function $T : \Omega \to Y$, where $\Omega \subset X$ and X and Y are normed spaces, if for $x \in \Omega$ and $h \in X$ the limit

$$\lim_{\alpha\downarrow 0} \frac{1}{\alpha}\left[T(x+\alpha h) - T(x)\right]$$

exists, then it is called the Gateaux differential at x with direction h and is denoted by $\partial T(x;h)$.

[19] Existence of the Gateaux differential follows from Lemma A.1 in Amador et al. (2006). See Amador and Bagwell (2013b) for further details concerning the application of this lemma.

$$\partial\mathcal{L}(\pi_{IR};x) = \int_\Gamma \Big(w_\pi(\gamma,\pi_{IR}(\gamma))f(\gamma) - \Lambda(\gamma)\Big) x(\gamma) d\gamma$$
$$+ \int_\Gamma \Big(-\gamma + b'(\pi_{IR}(\gamma))\Big) x(\gamma) d\Lambda(\gamma).$$

Using $b'(\pi_f(\gamma)) = \gamma$ and our knowledge of Λ and Ψ we get that

$$\partial\mathcal{L}(\pi_{IR};x) = \int_{\gamma_H}^{\overline{\gamma}} \Big(w_\pi(\gamma,\pi_f(\gamma_H))f(\gamma) - A - \kappa(1-F(\gamma)) - \kappa\Big(\gamma_H - \gamma\Big)f(\gamma)\Big) x(\gamma) d\gamma$$

Hence, integrating by parts, we get

$$\partial\mathcal{L}(\pi_{IR};x) = \left[\int_{\gamma_H}^{\overline{\gamma}} \Big(w_\pi(\gamma,\pi_f(\gamma_H))f(\gamma) - A - \kappa(1-F(\gamma)) - \kappa\Big(\gamma_H - \gamma\Big)f(\gamma)\Big) d\gamma\right] x(\overline{\gamma})$$
$$- \int_{\gamma_H}^{\overline{\gamma}} \left[\int_{\gamma_H}^{\gamma} \Big(w_\pi(\tilde{\gamma},\pi_f(\gamma_H))f(\tilde{\gamma}) - A - \kappa(1-F(\tilde{\gamma})) - \kappa\Big(\gamma_H - \tilde{\gamma}\Big)f(\tilde{\gamma})\Big) d\tilde{\gamma}\right] dx(\gamma)$$

Noting that $\int_{\gamma_H}^{\gamma} \Big(\kappa(1-F(\tilde{\gamma})) + \kappa\Big(\gamma_H - \tilde{\gamma}\Big)f(\tilde{\gamma})\Big) d\tilde{\gamma} = \kappa(\gamma-\gamma_H)(1-F(\gamma))$, we find that

$$\partial\mathcal{L}(\pi_{IR};x) = (G(\overline{\gamma}) - A)(\overline{\gamma} - \gamma_H)x(\overline{\gamma}) - \int_{\gamma_H}^{\overline{\gamma}} (G(\gamma) - A)(\gamma - \gamma_H) dx(\gamma). \tag{7}$$

Using (4) and (5), we also observe that

$$G(\overline{\gamma}) = A \tag{8}$$

We are now ready to evaluate the first order conditions. Using (8), we observe that the first term on the right-hand side of (7) is equal to zero. Since $\pi_{IR}(\gamma)$ is constant for $\gamma \in [\gamma_H, \overline{\gamma}]$, we now see immediately from (7) that $\partial\mathcal{L}(\pi_{IR};\pi_{IR}) = 0$. It remains to show that $\partial\mathcal{L}(\pi_{IR};x) \leq 0$ for all non-increasing $x \in \widehat{\Phi}$. Given that $A = G(\overline{\gamma})$ by (8), we see that this inequality holds if $G(\gamma) \leq A$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$. Thus, the first order conditions are satisfied if $G(\gamma) \leq G(\overline{\gamma})$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$. This is exactly what part (i) of Proposition 1 provides. Recall also that we require $A \geq 0$, since Φ must be non-decreasing. It is now evident that part (ii) of Proposition 1 provides this inequality.

As discussed above, we now finish the argument that $\kappa F(\gamma) + \Lambda(\gamma)$ is non-decreasing for all $\gamma \in [\underline{\gamma}, \overline{\gamma})$ by showing that jumps in $\Lambda(\gamma)$ at $\underline{\gamma}$ and γ_H are non-negative. To verify that the jump in $\Lambda(\gamma)$ at γ_H is non-negative, we observe that

$$A = G(\overline{\gamma}) \geq w_\pi(\gamma_H, \pi_f(\gamma_H))f(\gamma_H) - \kappa(1 - F(\gamma_H)) = G(\gamma_H)$$

follows from part (i) of Proposition 1. Likewise, we may verify that the jump in $\Lambda(\gamma)$ at $\underline{\gamma}$ is non-negative, since $w_\pi(\underline{\gamma}, \pi_f(\underline{\gamma}))f(\underline{\gamma}) > 0$.

To complete the proof, we use Theorem 1 in Amador and Bagwell (2013b). To apply this theorem, we set (i) $x_0 \equiv \pi_{IR}$; (ii) $X \equiv \{\pi | \pi : \Gamma \to \Pi\}$; (iii) f to be given by the negative of the objective function, $\int_\Gamma w(\gamma, \pi(\gamma))dF(\gamma)$, as a function of $\pi \in X$; (iv) $Z \equiv \{(z_1, z_2, z_3) | z_1 : \Gamma \to \mathbb{R},\ z_2 : \Gamma \to \mathbb{R}$ and $z_3 : \Gamma \to \mathbb{R}$ with z_1, z_2, z_3 integrable $\}$; (v) $\Omega \equiv \Phi$; (vi) $P \equiv \{(z_1, z_2, z_3) | (z_1, z_2, z_3) \in Z$ such that $z_1(\gamma) \geq 0, z_2(\gamma) \geq 0$ and $z_3(\gamma) \geq 0$ for all $\gamma \in \Gamma\}$; (vii) $\hat{G}$ (which is referred to as G in Theorem 1) to be the mapping from Φ to Z given by the left hand sides of inequalities (1), (2) and (3); (viii) T to be the linear mapping:

$$T((z_1, z_2)) \equiv \int_\Gamma z_1(\gamma)d\Lambda_1(\gamma) + \int_\Gamma z_2(\gamma)d\Lambda_2(\gamma) + \int_\Gamma z_3(\gamma)d\Psi(\gamma)$$

where Λ_1, Λ_2 and Ψ being non-decreasing functions implies that $T(z) \geq 0$ for $z \in P$. We have that

$$T(\hat{G}(x_0)) \equiv \int_\Gamma \left(\int_{\underline{\gamma}}^{\gamma} \pi_{IR}(\gamma')d\gamma' + \underline{U} - \gamma\pi_{IR}(\gamma) - b(\pi_{IR}(\gamma)) \right) d(\Lambda_1(\gamma) - \Lambda_2(\gamma)) = 0.$$

where the last equality follows from (1) and (2) binding under the π_{IR} allocation. We have found conditions under which the proposed allocation, π_{IR}, minimizes $f(x) + T(\hat{G}(x))$ for $x \in \Omega$. Given that $T(\hat{G}(x_0)) = 0$, then the conditions of Theorem 1 hold and it follows that π_{IR} solves $\min_{x\in\Omega} f(x)$ subject to $-\hat{G}(x) \in P$, which is Problem P.

□

We may interpret part (iii) of Proposition 1 in terms of Figure 4 and the intuition provided above. Observe that part (iii) is more easily satisfied when κ is big. Since $\frac{w_{\pi\pi}(\gamma,\pi)}{b''(\pi)} = 1 + \frac{\frac{1}{\alpha}v''(\pi)}{b''(\pi)}$, we thus conclude that part (iii) is more easily satisfied when the minimum value for $\frac{\frac{1}{\alpha}v''(\pi)}{b''(\pi)}$ is big. Additionally, since $w_\pi(\gamma, \pi_f(\gamma)) > 0$, we see that part (iii) is also more easily satisfied when the density is non-decreasing for $\gamma \in [\underline{\gamma}, \gamma_H]$. Part (ii) of Proposition 1 is also easily interpreted, as it captures the first order condition for the optimal placement of the critical type, γ_H, when attention is restricted to cap allocations that satisfy the IR constraint. In the absence of an IR constraint, the first order condition for an optimal simple cap allocation is defined in Lemma 1 and is equivalent to $G(\overline{\gamma}) = 0$. As part (ii) of Proposition 1 indicates, when the IR constraint is included, the first order condition becomes a weak inequality, $G(\overline{\gamma}) \geq 0$, which reflects the fact that a higher value for γ_H is feasible and thus cannot be attractive while a lower value for γ_H is infeasible even if it is potentially attractive.[20]

Proposition 1 is established for general regulator welfare functions, $w(\gamma, \pi)$. For further intuition, we provide stronger sufficient conditions that identify demand and distribution functions and welfare weights that satisfy the sufficient conditions defined in Proposition 1. We develop this approach in two ways. First, in the next section, we examine the case of a log demand function and show that the sufficient conditions in Proposition 1 are then satisfied under a simple restriction on the distribution function and welfare weight. Second, in the subsequent section, we show that the sufficient conditions in Proposition 1 are also sure to hold if the density is non-decreasing and an easy-to-check inequality holds, where the inequality describes a relationship between properties of the demand function and the welfare weight.

Finally, an interesting feature of Proposition 1 is that the optimal form of regulatory policy does not depend on the details of the distribution function, provided that the distribution function is such that the sufficient conditions in Proposition 1 hold. By contrast, when transfers are included as in the Baron and Myerson (1982) model, the specific form of the optimal regulatory policy requires detailed knowledge of the underlying distribution function. The results presented in the next two sections further clarify this point by imposing additional structure on demand functions and characterizing the distribution functions that satisfy the sufficient conditions in Proposition 1.

5 Log Demand Function

In this section, we consider the special case of a log demand function. This case is quite tractable and results in a rich set of distribution functions and welfare weights for which the sufficient conditions in Proposition 1 are satisfied.

In the *log demand example*, $P(\pi) = C_0 - C_1 \ln(\pi)$, where $C_1 > 0$, $\Pi = (0, C_0/C_1)$

[20] To see the point in more detail, let us define the expected welfare to the regulator under a cap allocation as

$$\mathbb{Z}(\gamma_c) = \int_{\underline{\gamma}}^{\gamma_c} w(\gamma, \pi_f(\gamma))dF(\gamma) + \int_{\gamma_c}^{\overline{\gamma}} w(\gamma, \pi_f(\gamma_c))dF(\gamma).$$

It is straightforward to confirm that

$$\mathbb{Z}'(\gamma_c)|_{\gamma_c=\gamma_H} = \pi_f'(\gamma_H)(\overline{\gamma} - \gamma_H)G(\overline{\gamma}),$$

where we recall that $\pi_f'(\gamma_c) < 0$. In the regulator's problem, since it is (not) feasible to raise γ_c above γ_H (lower γ_c below γ_H) slightly, it follows that $G(\overline{\gamma}) \geq 0$ is necessary for the optimal determination of γ_H among cap allocations that satisfy the IR constraint.

and $C_1 e^{g(\overline{\gamma})} > \sigma > [C_1 - (\overline{\gamma} - \underline{\gamma})]e^{g(\underline{\gamma})}$ for $g(\gamma) \equiv (C_0 - C_1 - \gamma)/C_1$.[21] We may verify that Assumptions 1 and 2 hold under these restrictions, with the flexible allocation taking the form $\pi_f(\gamma) = e^{g(\gamma)}$. In particular, for the log demand example, γ_H satisfies $e^{g(\gamma_H)}[C_1 - (\overline{\gamma} - \gamma_H)] = \sigma$, and so it follows from $\sigma \geq 0$ that $C_1 \geq \overline{\gamma} - \gamma_H$. A convenient feature of the log demand example is that $v(\pi)$ is linear, with the exact form being given as $v(\pi) = C_1 \pi$. As a consequence, we have that $\kappa = 1$ for the case of log demand.

Our next step is to express the function $G(\gamma)$ for the log demand example. Using $w_\pi(\widetilde{\gamma}, \pi_f(\gamma_H)) = \gamma_H - \widetilde{\gamma} + C_1/\alpha$, $\kappa = 1$ and (4), we find that

$$\begin{aligned} G(\gamma) &= \frac{1}{\gamma - \gamma_H} \int_{\gamma_H}^{\gamma} (\gamma_H - \widetilde{\gamma} + C_1/\alpha) f(\tilde{\gamma}) d\tilde{\gamma} - (1 - F(\gamma)) \qquad (9) \\ &= \frac{1}{\gamma - \gamma_H} \int_{\gamma_H}^{\gamma} (F(\widetilde{\gamma}) + (C_1/\alpha) f(\tilde{\gamma})) d\tilde{\gamma} - 1, \end{aligned}$$

where the second expression for $G(\gamma)$ follows after integration by parts.

We are now prepared to examine the sufficient conditions in Proposition 1. We begin by showing that part (ii) of Proposition 1 holds. To see that $G(\overline{\gamma}) \geq 0$, we observe from the first expression for $G(\gamma)$ in (9) that

$$\begin{aligned} G(\overline{\gamma}) &= \frac{1}{\overline{\gamma} - \gamma_H} \int_{\gamma_H}^{\overline{\gamma}} (\gamma_H - \widetilde{\gamma} + C_1/\alpha) f(\tilde{\gamma}) d\tilde{\gamma} \\ &\geq \frac{1}{\overline{\gamma} - \gamma_H} \int_{\gamma_H}^{\overline{\gamma}} (\gamma_H - \overline{\gamma} + C_1) f(\tilde{\gamma}) d\tilde{\gamma} \\ &\geq 0, \end{aligned}$$

where the first inequality uses $\alpha \in (0, 1]$ and $\widetilde{\gamma} \leq \overline{\gamma}$ and the second inequality uses $C_1 \geq \overline{\gamma} - \gamma_H$ as established above using $\sigma \geq 0$.

Our remaining task is to characterize conditions under which parts (i) and (iii) of Proposition 1 hold. For the log demand example, our approach is to impose an assumption under which the monotonicity condition in part (iii) of Proposition 1 holds *globally* (i.e., for all $\gamma \in \Gamma$). We then show that this assumption ensures that part (i) of Proposition 1 holds as well.

To this end, we observe that for the log demand example

$$\kappa F(\gamma) + w_\pi(\gamma, \pi_f(\gamma)) f(\gamma) = F(\gamma) + (C_1/\alpha) f(\gamma).$$

[21] To verify that $C_1 e^{g(\overline{\gamma})} > [C_1 - (\overline{\gamma} - \underline{\gamma})]e^{g(\underline{\gamma})}$, let $z(x) \equiv [C_1 - (\overline{\gamma} - x)]e^{g(x)}$ and observe that $z'(x) = [(\overline{\gamma} - x)/C_1]e^{g(x)} > 0$ for $x \in [\underline{\gamma}, \overline{\gamma})$.

Therefore, part (iii) of Proposition 1 holds globally if $F(\gamma) + (C_1/\alpha)f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$. Next, using the second expression for $G(\gamma)$ in (9), we may define

$$K(\gamma) \equiv G(\gamma) + 1 = \frac{1}{\gamma - \gamma_H} \int_{\gamma_H}^{\gamma} (F(\tilde{\gamma}) + (C_1/\alpha)f(\tilde{\gamma}))d\tilde{\gamma}.$$

We see that $K(\gamma_H) = G(\gamma_H) + 1 = F(\gamma_H) + (C_1/\alpha)f(\gamma_H)$. For $\gamma > \gamma_H$, we find that

$$K'(\gamma) = G'(\gamma) = \frac{F(\gamma) + (C_1/\alpha)f(\gamma)}{\gamma - \gamma_H} - \frac{1}{(\gamma - \gamma_H)^2} \int_{\gamma_H}^{\gamma} (F(\tilde{\gamma}) + (C_1/\alpha)f(\tilde{\gamma}))d\tilde{\gamma},$$

and so $K'(\gamma) = G'(\gamma) \geq 0$ if and only if

$$\int_{\gamma_H}^{\gamma} [(F(\gamma) + (C_1/\alpha)f(\gamma)) - (F(\tilde{\gamma}) + (C_1/\alpha)f(\tilde{\gamma}))]d\tilde{\gamma} \geq 0.$$

Therefore, if $F(\gamma) + (C_1/\alpha)f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$, then $G'(\gamma) \geq 0$ for $\gamma > \gamma_H$, and so part (i) of Proposition 1 then holds as well.

We may now summarize our findings in this section as follows:

Proposition 2. *Consider the log demand example and let $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ be defined as in Definition 2. If $F(\gamma)+(C_1/\alpha)f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$, then an optimal solution to the regulator's problem is achieved with the IR-cap allocation $\pi_{IR}(\gamma)$ such that $\pi_{IR}(\gamma) = \pi_f(\gamma)$ for all $\gamma \in [\underline{\gamma}, \gamma_H)$ and $\pi_{IR}(\gamma) = \pi_f(\gamma_H)$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$.*

The assumption that $F(\gamma)+(C_1/\alpha)f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$ clearly holds if the density $f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$. We emphasize, however, that the assumption can hold as well for densities that are non-monotonic or even for densities that decrease across the full support.[22] Furthermore, the details of the distribution otherwise matter only in so far as they determine γ_H. Different densities that satisfy the assumption and yield the same value for γ_H generate exactly the same optimal regulatory policy.

6 Density and Relative Concavity Conditions

In this section, we provide propositions that identify easy-to-check conditions that ensure the satisfaction of the sufficient conditions in Proposition 1. Building on the intuition featured at the start of Section 4, our approach here is to impose conditions on the density and the relative concavity of the regulator and monopolist's objectives that ensure the satisfaction of

[22]Consider, for example, $f(\gamma) = e^{-\gamma}$ where $\underline{\gamma}$ and $\overline{\gamma}$ satisfy $e^{-\underline{\gamma}} - e^{\overline{\gamma}} = 1$. This density is everywhere strictly decreasing, and $F(\gamma) + (C_1/\alpha)f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$ provided that $\alpha > C_1$.

the sufficient conditions in Proposition 1 and thereby guarantee the optimality of the IR-cap allocation. We also discuss the implications of our propositions for examples with linear, constant elasticity and exponential demand functions, respectively.

We begin by identifying a lower bound for our relative concavity parameter, κ, such that the IR-cap allocation is optimal when κ exceeds this bound and the density is non-decreasing:

Proposition 3. *Let $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ be defined as in Definition 2. If $\kappa \geq \frac{1}{2}$ and $f'(\gamma) \geq 0$ for all $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, then an optimal solution to the regulator's problem is achieved with the IR-cap allocation $\pi_{IR}(\gamma)$ such that $\pi_{IR}(\gamma) = \pi_f(\gamma)$ for all $\gamma \in [\underline{\gamma}, \gamma_H)$ and $\pi_{IR}(\gamma) = \pi_f(\gamma_H)$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$.*

Proof: Recall that we know that $w_\pi = -\gamma + P(\pi) + \pi P'(\pi) - \frac{1}{\alpha}\pi P'(\pi)$. Hence:

$$\begin{aligned}
G(\gamma) &\equiv \frac{1}{\gamma - \gamma_H}\int_{\gamma_H}^{\gamma} w_\pi(\tilde{\gamma}, \pi_f(\gamma_H))f(\tilde{\gamma})d\tilde{\gamma} - \kappa(1 - F(\gamma)) \\
&= \frac{1}{\gamma - \gamma_H}\int_{\gamma_H}^{\gamma} (-\tilde{\gamma} + P(\pi_f(\gamma_H)))f(\tilde{\gamma})d\tilde{\gamma} - \kappa(1 - F(\gamma)) \\
&\quad +\pi_f(\gamma_H)P'(\pi_f(\gamma_H))\left(1 - \frac{1}{\alpha}\right)\frac{F(\gamma) - F(\gamma_H)}{\gamma - \gamma_H} \\
&= \frac{1}{\gamma - \gamma_H}\left[\int_{\gamma_H}^{\gamma} (\overline{\gamma} - \tilde{\gamma})f(\tilde{\gamma})d\tilde{\gamma}\right] - \kappa(1 - F(\gamma)) \\
&\quad + \left(\frac{\sigma}{\pi_f(\gamma_H)} + \pi_f(\gamma_H)P'(\pi_f(\gamma_H))\left(1 - \frac{1}{\alpha}\right)\right)\frac{F(\gamma) - F(\gamma_H)}{\gamma - \gamma_H},
\end{aligned}$$

where we have used that $P(\pi_f(\gamma_H)) = \overline{\gamma} + \sigma/\pi_f(\gamma_H)$. We now consider the three parts of the sufficient conditions in order.

Consider part (i) of the sufficient conditions. We find that

$$\begin{aligned}
G'(\gamma) &= \frac{(\overline{\gamma} - \gamma)f(\gamma)}{\gamma - \gamma_H} - \frac{1}{(\gamma - \gamma_H)^2}\left[\int_{\gamma_H}^{\gamma} (\overline{\gamma} - \tilde{\gamma})f(\tilde{\gamma})d\tilde{\gamma}\right] + \kappa f(\gamma) \\
&\quad + \frac{\frac{\sigma}{\pi_f(\gamma_H)} + \pi_f(\gamma_H)P'(\pi_f(\gamma_H))(1 - \frac{1}{\alpha})}{(\gamma - \gamma_H)}\left(f(\gamma) - \frac{F(\gamma) - F(\gamma_H)}{\gamma - \gamma_H}\right)
\end{aligned}$$

Suppose $f'(\gamma) \geq 0$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$ and $\kappa \geq \frac{1}{2}$. Then, for all $\gamma \in (\gamma_H, \overline{\gamma}]$,

$$\begin{aligned}
G'(\gamma) &\geq \frac{(\overline{\gamma}-\gamma)f(\gamma)}{\gamma-\gamma_H} - \frac{1}{(\gamma-\gamma_H)^2}\left[\int_{\gamma_H}^{\gamma}(\overline{\gamma}-\tilde{\gamma})f(\gamma)d\tilde{\gamma}\right] + \kappa f(\gamma) \\
&\quad + \frac{\frac{\sigma}{\pi_f(\gamma_H)} + \pi_f(\gamma_H)P'(\pi_f(\gamma_H))(1-\frac{1}{\alpha})}{(\gamma-\gamma_H)}\Big(f(\gamma) - \frac{F(\gamma)-F(\gamma_H)}{\gamma-\gamma_H}\Big) \\
&\geq \frac{(\overline{\gamma}-\gamma)f(\gamma)}{\gamma-\gamma_H} - \frac{1}{(\gamma-\gamma_H)^2}\left[\int_{\gamma_H}^{\gamma}(\overline{\gamma}-\tilde{\gamma})f(\gamma)d\tilde{\gamma}\right] + \kappa f(\gamma) \\
&= f(\gamma)\left[\kappa - \frac{1}{2}\right] \\
&\geq 0,
\end{aligned}$$

where the first and second inequalities use $f'(\gamma) \geq 0$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$, the second inequality uses $\sigma \geq 0$ and $\alpha \in (0,1]$, and the third inequality uses $\kappa \geq \frac{1}{2}$. Further, under the same conditions, repeated applications of L'Hopital's rule yields $G'(\gamma_H) \geq 0$. Thus, if $f'(\gamma) \geq 0$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$ and $\kappa \geq \frac{1}{2}$, then $G'(\gamma) \geq 0$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$. Hence, if $f'(\gamma) \geq 0$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$ and $\kappa \geq \frac{1}{2}$, then $G(\overline{\gamma}) \geq G(\gamma)$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$.

Now consider part (ii) of the sufficient conditions. Using again that $\sigma \geq 0$ and $\alpha \in (0,1]$, we see that

$$\begin{aligned}
G(\overline{\gamma}) = \frac{1}{\overline{\gamma}-\gamma_H}&\left[\int_{\gamma_H}^{\overline{\gamma}}(\overline{\gamma}-\gamma)f(\gamma)d\gamma\right] \\
&+ \left(\frac{\sigma}{\pi_f(\gamma_H)} + \pi_f(\gamma_H)P'(\pi_f(\gamma_H))\left(1-\frac{1}{\alpha}\right)\right)\frac{1-F(\gamma_H)}{\overline{\gamma}-\gamma_H} > 0,
\end{aligned}$$

where the strict inequality uses $\gamma_H < \overline{\gamma}$. So part (ii) of the sufficient conditions is automatically satisfied.

Finally, we can rewrite part (iii) of our sufficient conditions as

$$\kappa F(\gamma) + \frac{1}{\alpha}v'(\pi_f(\gamma))f(\gamma) \text{ nondecreasing, for all } \gamma \in [\underline{\gamma}, \gamma_H].$$

Differentiating, we may represent this condition as follows:

$$f(\gamma)\left[\kappa + \frac{\frac{1}{\alpha}v''(\pi_f(\gamma))}{b''(\pi_f(\gamma))}\right] + \frac{1}{\alpha}v'(\pi_f(\gamma))f'(\gamma) \geq 0 \text{ for all } \gamma \in [\underline{\gamma}, \gamma_H], \tag{10}$$

where we use $\pi_f'(\gamma) = 1/b''(\pi_f(\gamma))$. Now, by the definition of κ, we know that

$$\kappa = \min_{\gamma,\pi \in \Gamma\times\Pi} \left\{ \frac{w_{\pi\pi}(\gamma,\pi)}{b''(\pi)} \right\} = \min_{\pi\in\Pi} \left\{ 1 + \frac{\frac{1}{\alpha}v''(\pi)}{b''(\pi)} \right\} \le 1 + \frac{\frac{1}{\alpha}v''(\pi_f(\gamma))}{b''(\pi_f(\gamma))}.$$

Thus, $\frac{\frac{1}{\alpha}v''(\pi_f(\gamma))}{b''(\pi_f(\gamma))} \ge \kappa - 1$ and so condition (10) is sure to hold if

$$2f(\gamma)\left[\kappa - \frac{1}{2}\right] + \frac{1}{\alpha}v'(\pi_f(\gamma))f'(\gamma) \ge 0 \text{ for all } \gamma \in [\underline{\gamma}, \gamma_H].$$

Since $v'(\pi_f(\gamma)) > 0$, we thus conclude that part (iii) of our sufficient conditions holds if $\kappa \ge \frac{1}{2}$ and $f'(\gamma) \ge 0$ for all $\gamma \in [\underline{\gamma}, \gamma_H]$. □

Proposition 3 formally captures the intuition presented at the start of Section 4. As anticipated, Proposition 3 utilizes a non-decreasing density. To understand the role of κ, observe that

$$\kappa = \min_{(\gamma,\pi)\in\Gamma\times\Pi} \left\{ \frac{w_{\pi\pi}(\gamma,\pi)}{b''(\pi)} \right\} = \min_{\pi\in\Pi} \left\{ 1 + \frac{\frac{1}{\alpha}v''(\pi)}{b''(\pi)} \right\},$$

where the latter equality follows since $w(\gamma,\pi) = -\gamma\pi + b(\pi) - \sigma + \frac{1}{\alpha}v(\pi)$. Based on Proposition 3, a key question is whether $\kappa \ge \frac{1}{2}$. This will clearly be the case if $v(\pi)$ is concave (or linear). If $\frac{1}{\alpha}v(\pi)$ is too convex relative to the concavity of $b(\pi)$, however, then our condition that $\kappa \ge \frac{1}{2}$ will fail. Clearly, if $v(\pi)$ is convex and α is sufficiently small, then κ will fall below $\frac{1}{2}$.[23]

To gain further insight, we explore general properties of demand functions under which $\kappa \ge \frac{1}{2}$. Let us define

$$\rho(\pi) = 1 + \frac{\frac{1}{\alpha}v''(\pi)}{b''(\pi)}.$$

Notice that $\kappa = \min_{\pi\in\Pi}\{\rho(\pi)\}$. Substituting, we find that

$$\rho(\pi) = \frac{(P''(\pi)\pi + P'(\pi))(1 - \frac{1}{\alpha}) + P'(\pi)}{P''(\pi)\pi + 2P'(\pi)} = \frac{(\frac{P''(\pi)}{P'(\pi)}\pi + 1)(1 - \frac{1}{\alpha}) + 1}{\frac{P''(\pi)\pi}{P'(\pi)} + 2}.$$

Thus, for $\pi \in \Pi$, $\rho(\pi) \ge \frac{1}{2}$ if and only if the following key inequality holds:

$$\frac{2(\alpha - 1)}{(2 - \alpha)} \ge \frac{P''(\pi)}{P'(\pi)}\pi \tag{11}$$

[23]Note, though, that α cannot not fall so far as to make $w(\gamma,\pi)$ convex in π and thereby violate Assumption 1.

We thus may now report the following proposition:

Proposition 4. *Let $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ be defined as in Definition 2. If inequality (11) holds for all $\pi \in \Pi$ and $f'(\gamma) \geq 0$ for all $\gamma \in \gamma \in [\underline{\gamma}, \overline{\gamma}]$, then an optimal solution to the regulator's problem is achieved with the IR-cap allocation $\pi_{IR}(\gamma)$ such that $\pi_{IR}(\gamma) = \pi_f(\gamma)$ for all $\gamma \in [\underline{\gamma}, \gamma_H)$ and $\pi_{IR}(\gamma) = \pi_f(\gamma_H)$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$.*

It is useful to understand when the key inequality (11) holds. Recall that $\alpha \in (0, 1]$. Using (11), if $P''(\pi) = 0$, then $\rho(\pi) \geq \frac{1}{2}$ if and only if $\alpha = 1$. If $P''(\pi) < 0$, then $\kappa < \frac{1}{2}$ follows. Our stronger sufficient conditions do not apply, therefore, when $P''(\pi) < 0$. Finally, if $P''(\pi) > 0$ and $\pi > 0$, then $\rho(\pi) > \frac{1}{2}$ if $\alpha = 1$. Notice in this last case that $\rho(\pi) \geq \frac{1}{2}$ can hold for $\alpha < 1$.[24]

We can also interpret the inequality in (11) in terms of our preceding discussion about the curvature properties of $v(\pi)$. When $\kappa < 1$, $v(\pi)$ brings a convex ingredient to $w(\gamma, \pi)$. As noted, we can allow for such convexity, as long as it is not so great as to push κ below 1/2. The left hand side of (11) rises with α and takes the values of -1 and 0 when α equals 0 and 1, respectively. Since $v(\pi)$ is concave if and only if the right hand side of (11) is less than or equal to -1, we conclude that (11) is sure to hold when $v(\pi)$ is concave. If $v(\pi)$ is convex with $P''(\pi) > 0$, then the right hand side of (11) is negative for $\pi > 0$, and so (11) holds for α sufficiently close to 1 when $\pi > 0$. In the case where $v(\pi)$ is convex with $P''(\pi) = 0$, the right hand side of (11) is zero, and hence (11) holds if and only if $\alpha = 1$. Finally, if $v(\pi)$ is convex with $P''(\pi) < 0$, then the right hand side of (11) is positive, and hence (11) cannot be satisfied.

We conclude our discussion by considering three examples. In the *linear demand example*, $P(\pi) = \mu - \beta\pi$, where $\mu > \overline{\gamma}$, $\beta > 0$, $\alpha \geq 1/2$ and $\Pi = [0, \mu/\beta)$. In this case, $v''(\pi) = \beta > 0$, and so $v(\pi)$ is convex. The *constant elasticity demand example* specifies that $P(\pi) = (\pi)^{-1/\epsilon}$ where $\epsilon > 1$, $\underline{\gamma} > 0$ and $\Pi = (0, \infty)$. For this example, $P'(\pi) = -(1/\epsilon)(\pi)^{-(1+1/\epsilon)} < 0$ and $P''(\pi) = (1/\epsilon)(1 + 1/\epsilon)(\pi)^{-(2+1/\epsilon)} > 0$, and so $v''(\pi) = -\pi^{-(1+1/\epsilon)}(1/\epsilon)^2 < 0$, which indicates that $v(\pi)$ is concave. Finally, in the *exponential demand example*, $P(\pi) = \beta e^{-\pi}$, where $\beta > \max\{\overline{\gamma}, \underline{\gamma}e^2\}$, $\alpha \geq 1/2$ and $\Pi = [0, 2)$. This example yields $P'(\pi) = -\beta e^{-\pi} < 0$ and $P''(\pi) = \beta e^{-\pi} > 0$, and so $v''(\pi) = \beta e^{-\pi}(1 - \pi)$, which indicates that $v(\pi)$ is convex for $\pi \in [0, 1)$ and concave for $\pi \in (1, 2]$.

We now compute κ for our three examples. For the linear demand example, $v''(\pi) = \beta$

[24] For π at or sufficiently close to zero, $\rho(\pi) \geq \frac{1}{2}$ holds for $\alpha \in (0, 1]$ if $\lim_{\pi \to 0} \frac{P''(\pi)}{P'(\pi)}\pi \leq -1$. The constant elasticity demand example that we consider below satisfies this limit condition. We also recall that, since our participation constraint requires positive output when $\sigma > 0$, an assumption that $\pi > 0$ is not restrictive when $\sigma > 0$.

and $b''(\pi) = -2\beta$, and so it follows that

$$\kappa = 1 - \frac{1}{2\alpha}.$$

Given $\alpha \leq 1$, our condition that $\kappa \geq \frac{1}{2}$ holds for this example if and only if $\alpha = 1$, so that the planner maximizes aggregate social surplus.[25] Our stronger sufficient conditions thus do not hold for the linear demand example when the regulator gives greater weight to consumer welfare.[26] We note, though, that the case in which the regulator maximizes aggregate social surplus is of some special interest from a normative standpoint.

For the constant elasticity demand example, $v''(\pi) = -(\pi)^{-(1+1/\epsilon)}(1/\epsilon)^2$ and $b''(\pi) = -(\pi)^{-(1+1/\epsilon)}(\epsilon - 1)(1/\epsilon)^2$. It thus follows for the constant elasticity example that

$$\kappa = 1 + \frac{1}{\alpha(\epsilon - 1)} > 1.$$

Thus, our condition that $\kappa \geq \frac{1}{2}$ clearly holds for the constant elasticity example, regardless of the value of $\alpha \in (0, 1]$.[27] Notice that a lower value for α helps in this case, since v brings additional concavity in the constant elasticity demand example.

For the exponential demand example, $v''(\pi) = \beta e^{-\pi}(1 - \pi)$ and $b''(\pi) = \beta e^{-\pi}(\pi - 2)$. We find for this example that $\rho(\pi)$ is minimized over $\pi \in [0, 2)$ at $\pi = 0$. It thus follows that

$$\kappa = 1 - \frac{1}{2\alpha}.$$

As in the linear demand example, given $\alpha \leq 1$, our condition that $\kappa \geq \frac{1}{2}$ holds for this example if and only if $\alpha = 1$, so that the planner maximizes aggregate social surplus.[28] A

[25] For the linear demand example where $\mu > \overline{\gamma}$, $\beta > 0$, $\alpha \geq 1/2$ and $\Pi = [0, \mu/\beta)$, Assumption 1 holds. The flexible allocation take the form $\pi_f(\gamma) = (\mu - \gamma)/(2\beta)$ and is interior. Assumption 2 holds if and only if $\frac{(\mu+\underline{\gamma})}{2} - \frac{\sigma 2\beta}{\mu - \underline{\gamma}} < \overline{\gamma} < \frac{(\mu+\overline{\gamma})}{2} - \frac{\sigma 2\beta}{\mu - \overline{\gamma}}$. When these inequalities hold, there exists $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ such that $P(\pi_f(\gamma_H)) = \overline{\gamma} + \frac{\mu}{\pi_f(\gamma_H)}$.

[26] As mentioned in the Introduction, Alonso and Matouschek (2008) also analyze the linear demand example. They posit throughout that $\alpha = 1$ and do not include a participation constraint; however, they do consider a more general family of (uni-modal) distributions. We note that our assumption of a non-decreasing density ensures that we can apply Proposition 1 but is certainly not necessary for the application of this proposition.

[27] For the constant elasticity demand example where $\epsilon > 1$, $\underline{\gamma} > 0$ and $\Pi = (0, \infty)$, Assumption 1 holds. The flexible allocation takes the form $\pi_f(\gamma) = (\frac{\gamma\epsilon}{\epsilon - 1})^{-\epsilon}$ and is interior. Assumption 2 holds if and only if $\frac{\gamma\epsilon}{\epsilon-1} - \sigma(\frac{\gamma\epsilon}{\epsilon-1})^{\epsilon} < \overline{\gamma} < \frac{\overline{\gamma}\epsilon}{\epsilon-1} - \sigma(\frac{\overline{\gamma}\epsilon}{\epsilon-1})^{\epsilon}$. When these inequalities hold, there exists $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ such that $P(\pi_f(\gamma_H)) = \overline{\gamma} + \frac{\sigma}{\pi_f(\gamma_H)}$.

[28] For the exponential demand example where $\beta > \max\{\overline{\gamma}, \underline{\gamma}e^2\}$, $\alpha \geq 1/2$ and $\Pi = [0, 2)$, Assumption 1 holds. The flexible output is interior under these restrictions. Assumption 2 holds if and only if σ falls in an intermediate range defined by $-\overline{\gamma}\pi_f(\underline{\gamma}) + b(\pi_f(\underline{\gamma})) < \sigma < -\overline{\gamma}\pi_f(\overline{\gamma}) + b(\pi_f(\overline{\gamma})))$, where $\pi_f(\gamma)$ is implicitly defined by the first order condition $b'(\overline{\pi}) - \gamma = 0$. When σ falls in this intermediate range, there exists

novel feature of the exponential demand example, however, is that $\rho(\pi)$ varies with π. If $\sigma > 0$, then outputs near zero violate the participation constraint, even for the lowest cost monopolist; hence, if $p(\pi)$ is minimized over the subset of Π in which π is individual rational for the lowest cost monopolist, then the resulting κ exceeds $\frac{1}{2}$ when $\alpha = 1$. Our condition would then not require $\alpha = 1$ but would instead hold if α were close to unity.

Finally, we also recall the log demand example considered in detail in the previous section. For this example, v is linear and so $\kappa = 1$; thus, if the density is non-decreasing, then Proposition 3 holds and indicates that the IR-cap allocation is optimal. Proposition 2 is consistent with Proposition 3, since the distributional assumption in Proposition 2 also holds if the density is non-decreasing. As remarked at the close of the previous section, however, Propositon 2 holds for a broader family of densities and allows even for densities that are everywhere decreasing.

7 Optimal Regulation with Exclusion

We now return to our discussion in Section 2 and consider the setting in which the regulator is allowed to select a menu of permissible outputs such that some types of the monopolist are excluded in that they choose to produce zero output. A key and perhaps surprising finding is that our characterizations of the solution to the regulator's problem for the setting without exclusion, as provided in our propositions above, lead directly to analogous characterizations of the solution to the regulator's problem for the setting with exclusion.

To begin, we define the *regulator's problem when exclusion is allowed*:

$$\max_{\pi:\Gamma\to\Pi_0} \int_\Gamma \mathbf{1}(\gamma)\cdot \left(-\gamma\pi(\gamma) + P(\pi(\gamma))\pi(\gamma) - \sigma + \frac{1}{\alpha}\left(\int_0^{\pi(\gamma)} P(z)dz - P(\pi(\gamma))\pi(\gamma)\right)\right) dF(\gamma)$$

subject to:

$$\gamma \in \arg\max_{\tilde{\gamma}\in\Gamma} \mathbf{1}(\tilde{\gamma})\cdot(-\gamma\pi(\tilde{\gamma}) + P(\pi(\tilde{\gamma}))\pi(\tilde{\gamma}) - \sigma), \text{ for all } \gamma \in \Gamma$$

$$0 \le \mathbf{1}(\gamma)\cdot(-\gamma\pi(\gamma) + P(\pi(\gamma))\pi(\gamma) - \sigma), \text{ for all } \gamma \in \Gamma$$

where $\Pi_0 \equiv \Pi$ when $0 \in \Pi$, $\Pi_0 \equiv \Pi \cup \{0\}$ when $0 \notin \Pi$, and $\mathbf{1}(\gamma)$ is an indicator function such that $\mathbf{1}(\gamma) = 1$ if $\pi(\gamma) > 0$ and $\mathbf{1}(\gamma) = 0$ if $\pi(\gamma) = 0$.[29] Notice that, even when $\sigma > 0$, the constraints now allow for the possibility of types for which $\pi(\gamma) = 0$, since the IR constraint

$\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ such that $P(\pi_f(\gamma_H)) = \overline{\gamma} + \frac{\sigma}{\pi_f(\gamma_H)}$.

[29]For γ such that $\mathbf{1}(\gamma) = 0$ and thus $\pi(\gamma) = 0$, it is understood that the integrand takes the value of zero even if the term that multiplies $\mathbf{1}(\gamma)$ is infinite. Similar remarks apply to the constraints and to the program as re-written later in this section.

as represented here holds when $\mathbf{1}(\gamma) = 0$.

Our first step is to show that, if exclusion occurs, then the excluded types are defined by a threshold type, $\gamma_t < \overline{\gamma}$ such that $\pi(\gamma) = 0$ if $\gamma > \gamma_t$ and $\pi(\gamma) > 0$ if $\gamma \leq \gamma_t$, where type γ_t is indifferent between producing and not.[30] Suppose to the contrary that the constraints for the regulator's problem when exclusion is allowed are satisfied by some function $\pi(\gamma)$ for which, for some γ_1 and γ_2 with $\underline{\gamma} \leq \gamma_1 < \gamma_2 \leq \overline{\gamma}$, we have that $\pi(\gamma_1) = 0 < \pi(\gamma_2)$. A monopolist with type γ_1 would then gain by violating the incentive compatibility constraint and selecting instead the output intended for type γ_2 :

$$\begin{aligned} & \mathbf{1}(\gamma_2) \cdot (-\gamma_1 \pi(\gamma_2) + P(\pi(\gamma_2))\pi(\gamma_2) - \sigma) \\ > \; & \mathbf{1}(\gamma_2) \cdot (-\gamma_2 \pi(\gamma_2) + P(\pi(\gamma_2))\pi(\gamma_2) - \sigma) \\ \geq \; & 0 \\ = \; & \mathbf{1}(\gamma_1) \cdot (-\gamma_1 \pi(\gamma_1) + P(\pi(\gamma_1))\pi(\gamma_1) - \sigma), \end{aligned}$$

where the first inequality follows since $\pi(\gamma_2) > 0$ and $\gamma_1 < \gamma_2$, the second inequality follows from the participation constraint for a monopolist with type γ_2, and the equality follows since $\pi(\gamma_1) = 0$.

Our second step is to consider the necessary features of the solution to the regulator's problem when exclusion is allowed. In particular, we assume for now that a solution to this problem entails exclusion. As just argued, the set of excluded types then must be defined by an interval $(\gamma_t, \overline{\gamma}]$ with $\gamma_t < \overline{\gamma}$, where the participation constraint for threshold type γ_t necessarily binds. Conditional on the solution excluding exactly the set of types in the interval $(\gamma_t, \overline{\gamma}]$, we seek to characterize the solution's associated allocation function.

To this end, we define a "truncated" distribution given by the support $\Gamma_t \equiv [\underline{\gamma}, \gamma_t]$ and the distribution function $F_t(\gamma) \equiv F(\gamma)/F(\gamma_t)$. The *regulator's truncated problem* may now be defined as the regulator's problem once we replace Γ with Γ_t and likewise $F(\gamma)$ with $F_t(\gamma)$. With these definitions at hand, let $\pi(\gamma)$ be any function that satisfies the constraints of the regulator's problem when exclusion is allowed and that entails exclusion for exactly the types in the interval $(\gamma_t, \overline{\gamma}]$. When $\pi(\gamma)$ is restricted to the domain $\Gamma_t \equiv [\underline{\gamma}, \gamma_t]$, it is straightforward to verify that it satisfies the constraints of the regulator's truncated problem. The regulator's truncated problem is thus a relaxed problem, in that it includes any (domain-restricted) function that satisfies the constraints of the regulator's problem when exclusion is allowed while excluding exactly the types in the interval $(\gamma_t, \overline{\gamma}]$. Going the other direction, consider any function $\pi_t(\gamma)$ that satisfies both the constraints of the regulator's truncated problem *and* the property that the associated participation constraint binds for type γ_t so

[30]For simplicity, we assume that the indifferent type γ_t produces positive output, $\pi(\gamma_t) > 0$.

that $\sigma = P(\pi_t(\gamma_t))\pi_t(\gamma_t) - \gamma_t\pi_t(\gamma_t)$. If we now define the function $\widehat{\pi}(\gamma)$ by $\widehat{\pi}(\gamma) = \pi_t(\gamma)$ for $\gamma \in \Gamma_t$ and $\widehat{\pi}(\gamma) = 0$ for $\gamma \in (\gamma_t, \overline{\gamma}]$, then we may easily confirm that $\widehat{\pi}(\gamma)$ satisfies the constraints of the regulator's problem when exclusion is allowed.[31] Of course, the function $\widehat{\pi}(\gamma)$ also excludes exactly the types in the interval $(\gamma_t, \overline{\gamma}]$.

Based on these relationships, we conclude that, if the regulator's problem when exclusion is allowed has a solution for which the set of excluded types is defined by an interval $(\gamma_t, \overline{\gamma}]$ with $\gamma_t < \overline{\gamma}$, and if the associated regulator's truncated problem is solved by a function $\pi_t(\gamma)$ for which the associated participation constraint binds for type γ_t so that $\sigma = P(\pi_t(\gamma_t))\pi_t(\gamma_t) - \gamma_t\pi_t(\gamma_t)$, then the regulator's problem when exclusion is allowed is solved by the function $\widehat{\pi}(\gamma)$ defined by $\widehat{\pi}(\gamma) = \pi_t(\gamma)$ for $\gamma \in \Gamma_t$ and $\widehat{\pi}(\gamma) = 0$ for $\gamma \in (\gamma_t, \overline{\gamma}]$.

We now record our observations to this point as follows:

Proposition 5. *(i) Suppose that $\pi(\gamma)$ satisfies the constraints of the regulator's problem when exclusion is allowed and that some types are excluded (i.e., $\pi(\gamma) = 0$ for some types). Then the set of excluded types is characterized by an interval, $(\gamma_t, \overline{\gamma}]$ with $\gamma_t < \overline{\gamma}$. (ii) Assume that the solution to the regulator's problem when exclusion is allowed entails exclusion over an interval $(\gamma_t, \overline{\gamma}]$, with $\gamma_t < \overline{\gamma}$. Let $\pi_t(\gamma)$ be a solution to the regulator's truncated problem. If $\sigma = P(\pi_t(\gamma_t))\pi_t(\gamma_t) - \gamma_t\pi_t(\gamma_t)$, then $\widehat{\pi}(\gamma)$ is a solution to the regulator's problem when exclusion is allowed, where $\widehat{\pi}(\gamma)$ is defined by $\widehat{\pi}(\gamma) = \pi_t(\gamma)$ for $\gamma \in \Gamma_t$ and $\widehat{\pi}(\gamma) = 0$ for $\gamma \in (\gamma_t, \overline{\gamma}]$*

Our third step is to use our characterization of the solution to the regulator's problem and thereby characterize the solution to the regulator's truncated problem. Our approach is to utilize our propositions in preceding sections for the regulator's problem once they are adjusted to apply to the truncated setting.

[31] The participation constraint clearly holds for $\gamma \in \Gamma_t$, since $\widehat{\pi}(\gamma) = \pi_t(\gamma)$ for $\gamma \in \Gamma_t$ satisfies the participation constraint for the regulator's truncated problem. The participation constraint also holds for $\gamma \in (\gamma_t, \overline{\gamma}]$, since $\widehat{\pi}(\gamma) = 0$ for $\gamma \in (\gamma_t, \overline{\gamma}]$. The incentive compatibility constraint likewise holds for $\gamma \in \Gamma_t$ and $\widetilde{\gamma} \in \Gamma_t$. For $\gamma \in \Gamma_t$ and $\widetilde{\gamma} \in (\gamma_t, \overline{\gamma}]$, the incentive compatibility constraint holds since $\widehat{\pi}(\gamma) = 0$ for $\gamma \in (\gamma_t, \overline{\gamma}]$ and the participation constraint holds for $\gamma \in \Gamma_t$. The incentive compatibility constraint trivially holds for $\gamma \in (\gamma_t, \overline{\gamma}]$ and $\widetilde{\gamma} \in (\gamma_t, \overline{\gamma}]$. Finally, the incentive compatibility constraint holds for $\gamma \in (\gamma_t, \overline{\gamma}]$ and $\widetilde{\gamma} \in \Gamma_t$, since

$$\begin{aligned}
& \mathbf{1}(\widetilde{\gamma}) \cdot (- \gamma\pi(\widetilde{\gamma}) + P(\pi(\widetilde{\gamma}))\pi(\widetilde{\gamma}) - \sigma) \\
< \; & \mathbf{1}(\widetilde{\gamma}) \cdot (-\gamma_t\pi(\widetilde{\gamma}) + P(\pi(\widetilde{\gamma}))\pi(\widetilde{\gamma}) - \sigma) \\
\leq \; & \mathbf{1}(\gamma_t) \cdot (-\gamma_t\pi(\gamma_t) + P(\pi(\gamma_t))\pi(\gamma_t) - \sigma) \\
= \; & 0,
\end{aligned}$$

where the first inequality follows since $\gamma > \gamma_t$ and $\pi(\widetilde{\gamma}) > 0$, the second inequality follows from the incentive compatibility constraint for type γ_t with $\gamma \in \Gamma_t$ and $\widetilde{\gamma} \in \Gamma_t$, and the equality follows since the participation constraint binds for type γ_t.

To this end, we now assume that Assumption 2 holds when $\overline{\gamma}$ is replaced with γ_t. We refer to this new assumption as *Assumption 2,t*. Drawing on Definition 2, we can then define $\gamma_{H,t}$ to be the counterpart to γ_H, where $\gamma_{H,t} \in (\underline{\gamma}, \gamma_t)$ satisfies

$$-\gamma_t \pi_f(\gamma_{H,t}) + b(\pi_f(\gamma_{H,t})) = \sigma.$$

The *IR-cap,t allocation* may now be defined just as was the IR-cap allocation, once $\overline{\gamma}$ is replaced with γ_t and γ_H is replaced by $\gamma_{H,t}$; thus, the IR-cap,t allocation is defined by

$$\pi_{IR,t}(\gamma) = \begin{cases} \pi_f(\gamma) & ; \gamma \in [\underline{\gamma}, \gamma_{H,t}] \\ \pi_f(\gamma_{H,t}) & ; \gamma \in (\gamma_{H,t}, \gamma_t] \end{cases}.$$

We note that Assumption 1 continues to hold for the truncated support Γ_t, since $\overline{\gamma} > \gamma_t$.

Finally, to state our propositions below, it is convenient to further define the *extended IR-cap,t allocation* $\widehat{\pi}_{IR,t}(\gamma)$ as the IR-cap,t allocation when the domain is extended to Γ and shutdown is imposed for types above γ_t. Formally, the extended IR-cap,t allocation is defined by $\widehat{\pi}_{IR,t}(\gamma) = \pi_{IR,t}(\gamma)$ for $\gamma \in [\underline{\gamma}, \gamma_t]$ and $\widehat{\pi}_{IR,t}(\gamma) = 0$ for $\gamma \in (\gamma_t, \overline{\gamma}]$.

With the basic setup now appropriately extended, we turn to our specific propositions. We begin with the log-demand case. For this case, and with $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ defined as in Definition 2, we know from Proposition 2 that, if $F(\gamma) + (C_1/\alpha)f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$, then optimal solution to the regulator's problem is achieved with the IR-cap allocation $\pi_{IR}(\gamma)$ such that $\pi_{IR}(\gamma) = \pi_f(\gamma)$ for all $\gamma \in [\underline{\gamma}, \gamma_H)$ and $\pi_{IR}(\gamma) = \pi_f(\gamma_H)$ for all $\gamma \in [\gamma_H, \overline{\gamma}]$. Observe, however, that the assumption that $F(\gamma) + (C_1/\alpha)f(\gamma)$ is non-decreasing for all $\gamma \in \Gamma$ in turn implies that the truncated distribution $F_t(\gamma)$ is such that $F_t(\gamma) + (C_1/\alpha)f_t(\gamma)$ is non-decreasing for all $\gamma \in \Gamma_t$, where $f_t(\gamma) \equiv f(\gamma)/f(\gamma_t)$. We now have the following result:

Proposition 6. *Assume that the solution to the regulator's problem when exclusion is allowed entails exclusion, and let the set of excluded types be defined by an interval* $(\gamma_t, \overline{\gamma}]$ *with* $\gamma_t < \overline{\gamma}$. *Let Assumption 2,t hold for this value of* γ_t. *Consider the log demand example. If* $F(\gamma)+(C_1/\alpha)f(\gamma)$ *is non-decreasing for all* $\gamma \in \Gamma$, *then an optimal solution to the regulator's problem when exclusion is allowed is achieved with the extended IR-cap,t allocation* $\widehat{\pi}_{IR,t}(\gamma)$ *such that* $\widehat{\pi}_{IR,t}(\gamma) = \pi_f(\gamma)$ *for all* $\gamma \in [\underline{\gamma}, \gamma_{H,t})$, $\widehat{\pi}_{IR,t}(\gamma) = \pi_f(\gamma_{H,t})$ *for all* $\gamma \in [\gamma_{H,t}, \gamma_t]$, *and* $\widehat{\pi}_{IR,t}(\gamma) = 0$ *for all* $\gamma \in (\gamma_t, \overline{\gamma}]$.

Under the assumptions stated in the proposition, we note that $\pi_t(\gamma) \equiv \pi_{IR,t}(\gamma)$ solves the regulator's truncated problem and also ensures that the IR constraint binds for type γ_t, so that $\sigma = P(\pi_t(\gamma_t))\pi_t(\gamma_t) - \gamma\pi_t(\gamma_t)$. This observation enables us to use Proposition 5 and conclude that $\widehat{\pi}_{IR,t}(\gamma)$ is a solution to the regulator's problem when exclusion is allowed.

We next extend Proposition 3 to the truncated environment. For this case, and with $\gamma_H \in (\underline{\gamma}, \overline{\gamma})$ defined as in Definition 2, we know from Proposition 3 that, if $\kappa \geq \frac{1}{2}$ and $f'(\gamma) \geq 0$ for all $\gamma \in \Gamma$, then optimal solution to the regulator's problem is achieved with the IR-cap allocation, $\pi_{IR}(\gamma)$. Recall now that $\kappa = \min_{\pi \in \Pi} \left\{ 1 + \frac{\frac{1}{\alpha} v''(\pi)}{b''(\pi)} \right\}$ is independent of the support of γ for the regulation model considered here. Furthermore, it is immediate that the assumption that $f'(\gamma) \geq 0$ for all $\gamma \in \Gamma$ in turn implies that the truncated density $f_t(\gamma)$ satisfies $f_t'(\gamma) \geq 0$ for all $\gamma \in \Gamma_t$, where $f_t(\gamma) \equiv f(\gamma)/F(\gamma_t)$. We now have the following result:

Proposition 7. *Assume that the solution to the regulator's problem when exclusion is allowed entails exclusion, and let the set of excluded types be defined by an interval $(\gamma_t, \overline{\gamma}]$ with $\gamma_t < \overline{\gamma}$. Let Assumption 2,t hold for this value of γ_t. If $\kappa \geq \frac{1}{2}$ and $f'(\gamma) \geq 0$ for all $\gamma \in \Gamma$, then an optimal solution to the regulator's problem when exclusion is allowed is achieved with the extended IR-cap,t allocation $\widehat{\pi}_{IR,t}(\gamma)$ such that $\widehat{\pi}_{IR,t}(\gamma) = \pi_f(\gamma)$ for all $\gamma \in [\underline{\gamma}, \gamma_{H,t})$, $\widehat{\pi}_{IR,t}(\gamma) = \pi_f(\gamma_{H,t})$ for all $\gamma \in [\gamma_{H,t}, \gamma_t]$, and $\widehat{\pi}_{IR,t}(\gamma) = 0$ for all $\gamma \in (\gamma_t, \overline{\gamma}]$.*

Once again, under the assumptions stated in the proposition, $\pi_t(\gamma) \equiv \pi_{IR,t}(\gamma)$ solves the regulator's truncated problem and also ensures that the IR constraint binds for type γ_t, so that $\sigma = P(\pi_t(\gamma_t))\pi_t(\gamma_t) - \gamma\pi_t(\gamma_t)$. We may thus again use Proposition 5 and conclude that $\widehat{\pi}_{IR,t}(\gamma)$ is a solution to the regulator's problem when exclusion is allowed.

Having now characterized the solution to the regulator's problem when exclusion is allowed for a given value of the threshold type, γ_t, we come to our fourth and final step and consider the determination of the optimal value for the threshold type. For this purpose, we find it convenient to select the output level $\pi_f(\gamma_{H,t})$ with the values for γ_t and $\gamma_{H,t}$ then implied. To simplify notation, we define

$$\pi_{H,t} \equiv \pi_f(\gamma_{H,t})$$

with γ_t then defined by $b(\pi_{H,t}) - \gamma_t \pi_{H,t} = \sigma$ so that

$$\gamma_t(\pi_{H,t}) = \frac{b(\pi_{H,t}) - \sigma}{\pi_{H,t}} \tag{12}$$

and with $\gamma_{H,t}$ then defined by

$$\gamma_{H,t}(\pi_{H,t}) = b'(\pi_{H,t}). \tag{13}$$

As Figure 5 illustrates, by selecting $\pi_{H,t} \geq \pi_f(\gamma_H)$, we thus define an IR-cap,t allocation $\pi_{IR,t}(\gamma)$ with $\gamma_t(\pi_{H,t}) \leq \overline{\gamma}$ and $\gamma_{H,t}(\pi_{H,t}) \leq \gamma_H$, where $\gamma_{H,t}(\pi_{H,t}) < \gamma_t(\pi_{H,t})$.[32] Note that, if

[32] Using (12), we confirm below that $\gamma_t(\pi_{H,t})$ is a decreasing function.

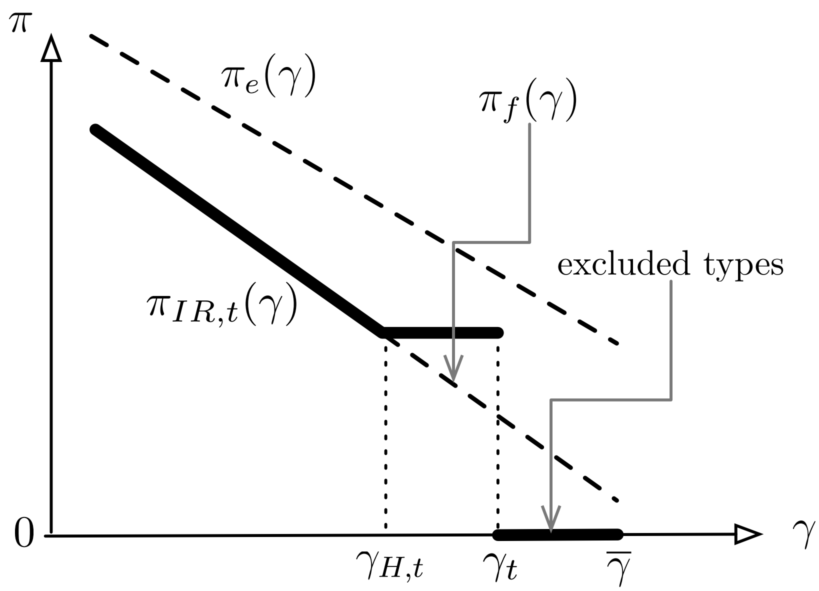

Figure 5: The proposed allocation with exclusion

we select $\pi_{H,t} = \pi_f(\gamma_H)$, then we have that $\gamma_t(\pi_{H,t}) = \overline{\gamma}$ and $\gamma_{H,t}(\pi_{H,t}) = \gamma_H$. We thus pick up the IR-cap allocation $\pi_{IR}(\gamma)$ as the special case where $\pi_{H,t} = \pi_f(\gamma_H)$.

Proceeding as in Section 2, we now re-write the regulator's problem when exclusion is allowed as follows:

$$\max_{\pi:\Gamma\to\Pi_0} \int_\Gamma \mathbf{1}(\tilde{\gamma})\cdot(w(\gamma,\pi(\gamma)))dF(\gamma) \qquad \text{subject to:}$$
$$\gamma \in \arg\max_{\tilde{\gamma}\in\Gamma} \mathbf{1}(\tilde{\gamma}) \cdot (-\gamma\pi(\tilde{\gamma}) + b(\pi(\tilde{\gamma})) - \sigma), \text{ for all } \gamma \in \Gamma$$
$$0 \leq \mathbf{1}(\gamma) \cdot (\, -\gamma\pi(\gamma) + b(\pi(\gamma)) - \sigma), \text{ for all } \gamma \in \Gamma$$

In this fourth step of our analysis, our focus is on the IR-cap,t allocation $\pi_{IR,t}(\gamma)$, which by construction satisfies the constraints of this problem. We thus consider the maximization of the objective of this problem with respect to the selection of $\pi_{H,t}$, where the objective to be maximized can be expressed accordingly in the following way:

$$W(\pi_{H,t}) \equiv \int_{\underline{\gamma}}^{\gamma_{H,t}(\pi_{H,t})} w(\gamma, \pi_f(\gamma))dF(\gamma) + \int_{\gamma_{H,t}(\pi_{H,t})}^{\gamma_t(\pi_{H,t})} w(\gamma, \pi_{H,t})dF(\gamma),$$

where we utilize that realized welfare is zero for $\gamma > \gamma_t(\pi_{H,t})$ since $\mathbf{1}(\gamma) = 0$ then holds.

We now consider the impact on regulator welfare of a small change in $\pi_{H,t}$. Using (12) and (13), we may easily establish that

$$\gamma_t'(\pi_{H,t}) = \frac{\gamma_{H,t}(\pi_{H,t}) - \gamma_t(\pi_{H,t})}{\pi_{H,t}} < 0.$$

Using this expression, we then find that

$$W'(\pi_{H,t}) = (\frac{\gamma_{H,t}(\pi_{H,t}) - \gamma_t(\pi_{H,t})}{\pi_{H,t}})w(\gamma_t(\pi_{H,t}), \pi_{H,t})f(\gamma_t(\pi_{H,t})) + \int_{\gamma_{H,t}(\pi_{H,t})}^{\gamma_t(\pi_{H,t})} w_\pi(\gamma, \pi_{H,t})dF(\gamma). \tag{14}$$

Intuitively, and as (14) captures, a higher value for $\pi_{H,t}$ lowers welfare by reducing $\gamma_t(\pi_{H,t})$ and thus expanding the set of excluded types but also impacts welfare by raising the output floor $\pi_{H,t}$ at which the highest active types pool.

As our preceding discussion indicates, the first term in (14) reflects the expansion in the set of excluded types and is negative. The sign of the second term is not immediately obvious, but we can show that it is positive when evaluated at $\pi_{H,t} = \pi_f(\gamma_H)$. This point of evaluation is of particular interest, since $W'(\pi_{H,t}) > 0$ at $\pi_{H,t} = \pi_f(\gamma_H)$ establishes the optimality of at least a small amount of exclusion.

To establish that the second term in (14) is positive when evaluated at $\pi_{H,t} = \pi_f(\gamma_H)$, we recall the characterization in Lemma 1 of the optimal simple cap allocation. We show there that the optimal simple cap allocation is described by a critical value $\gamma_c < \overline{\gamma}$ such that

$$\int_{\gamma_1}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_1))dF(\gamma) = 0 \tag{15}$$

when $\gamma_1 = \gamma_c$. Recall also our assumption that there is a unique $\gamma_c < \overline{\gamma}$ that solves (15). Given that (15) holds as well at the boundary (i.e., when $\gamma_1 = \overline{\gamma}$), it is then straightforward to verify that the assumption of a unique interior solution $\gamma_c < \overline{\gamma}$ implies that

$$\int_{\gamma_2}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_2))dF(\gamma) > 0$$

for all $\gamma_2 \in (\gamma_c, \overline{\gamma})$.[33] We also know that optimal simple cap allocation fails the IR constraint

[33] The optimal simple cap allocation maximizes the objective

$$\mathbb{Z}(\gamma_1) = \int_{\underline{\gamma}}^{\gamma_1} w(\gamma, \pi_f(\gamma))dF(\gamma) + \int_{\gamma_1}^{\overline{\gamma}} w(\gamma, \pi_f(\gamma_1))dF(\gamma).$$

The first derivative takes the form

$$\mathbb{Z}'(\gamma_1) = \int_{\gamma_1}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_1))\pi_f'(\gamma_1)dF(\gamma).$$

in the regulator's problem with $\gamma_H > \gamma_c$; hence, if we set $\pi_{H,t} = \pi_f(\gamma_H)$ and thus induce $\gamma_t(\pi_{H,t}) = \overline{\gamma}$ and $\gamma_{H,t}(\pi_{H,t}) = \gamma_H$, then we obtain

$$\int_{\gamma_{H,t}(\pi_{H,t})}^{\gamma_t(\pi_{H,t})} w_\pi(\gamma, \pi_{H,t}) dF(\gamma) = \int_{\gamma_H}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_H)) dF(\gamma) > 0.$$

We conclude that the second term in (14) is positive when we evaluate at the IR-cap allocation (i.e., when $\pi_{H,t} = \pi_f(\gamma_H)$).

As noted above, we are particularly interested in conditions that determine whether the regulator would gain by introducing even a small amount of exclusion. We thus evaluate (14) when $\pi_{H,t} = \pi_f(\gamma_H)$, $\gamma_t(\pi_{H,t}) = \overline{\gamma}$ and $\gamma_{H,t}(\pi_{H,t}) = \gamma_H$:

$$W'(\pi_f(\gamma_H)) = (\frac{\gamma_H - \overline{\gamma}}{\pi_f(\gamma_H)}) w(\overline{\gamma}, \pi_f(\gamma_H)) f(\overline{\gamma}) + \int_{\gamma_H}^{\overline{\gamma}} w_\pi(\gamma, \pi_f(\gamma_H)) dF(\gamma), \tag{16}$$

where as discussed the first (second) term is negative (positive). We may now make the general point that a small amount of exclusion is beneficial to the regulator if and only if $W'(\pi_f(\gamma_H)) > 0$:

Proposition 8. *Consider the regulator's problem when exclusion is allowed. With $W'(\pi_f(\gamma_H))$ given by (16), suppose that $W'(\pi_f(\gamma_H)) > 0$. Then the regulator can improve upon the IR-cap allocation by introducing a small amount of exclusion in the form of the extended IR-cap,t allocation for γ_t slightly below $\overline{\gamma}$.*

Guided by the general result obtained in Proposition 8, the optimality of some exclusion may be readily examined for specific demand functions.

Our results also provide a foundation for characterizing the globally optimal value for γ_t. Under log-demand, or when $\kappa \geq 1/2$, we may use the distributional assumptions given previously to confirm that the extended IR-cap,t allocation is optimal as lower values for γ_t are induced. Thus, the first-order condition derived above for $\pi_{H,t}$ continues to apply for such settings, since we know that the extended IR-cap,t allocation is then an optimal allocation conditional on the associated value for γ_t. For γ_t values that are sufficiently close to $\underline{\gamma}$, however, Assumption 2,*t* will fail, as a monopolist with cost type γ_t would then earn strictly positive profit when selecting the monopoly output for cost type $\underline{\gamma}$, $\pi_f(\underline{\gamma})$.

Finally, we note that the expression derived in (16) also plays a valuable role by isolating the key intuitive forces that determine the optimality of a small amount of exclusion. In

We also may confirm that $\mathbb{Z}'(\overline{\gamma}) = 0 < \mathbb{Z}''(\overline{\gamma})$. Given our assumption that the first-order condition $Z'(\gamma_1) = 0$ has a unique interior solution, denoted as γ_c, it now follows that $Z'(\gamma_c) = 0 > Z'(\gamma_2)$ for all $\gamma_2 \in (\gamma_c, \overline{\gamma})$. The expression given in the text now follows directly upon recalling that $\pi_f'(\gamma_2) < 0$.

particular, when a regulator considers the introduction of a small amount of exclusion, the regulator weighs the benefit of a higher output level (and thus a lower price) for pooled types as captured in the second term in (16) against the expected cost of a small interval of the highest-cost types that become excluded as captured in the first term in (16). One immediate implication of the tradeoff represented in (16) is that exclusion is less attractive when the expected social welfare associated with the highest-cost type, $w(\overline{\gamma}, \pi_f(\gamma_H))f(\overline{\gamma})$, is larger. Similarly, exclusion becomes attractive when the expected social welfare associated with the highest-cost type is sufficiently low, whether because this type has a sufficiently low density or because this type adds little to social welfare (as would be the case, for example, if $\alpha = 1$, $\sigma = 0$ and $P(0)$ is only slightly above $\overline{\gamma}$).

8 Conclusion

In this paper, we analyze the Baron and Myerson (1982) model of regulation under the restriction that transfers are infeasible. Extending the Lagrangian approach to delegation problems of Amador and Bagwell (2013b) to include an ex post participation constraint, we report sufficient conditions under which optimal regulation takes the simple and common form of price-cap regulation. We also identify families of demand and distribution functions and welfare weights that are sure to satisfy our sufficient conditions. We illustrate our sufficient conditions using examples with log demand, linear demand, constant elasticity demand and exponential demand, respectively. We also establish the optimality of price-cap regulation under different representations of the ex post participation constraint that differ as regards to whether the exclusion of some types is feasible. Finally, while we focus here on a model of regulation, we expect that our methods will be useful for other studies of applied mechanism design when transfers are infeasible and participation constraints play an important role.

References

Alonso, Ricardo and Niko Matouschek, "Optimal Delegation," *The Review of Economic Studies*, 2008, *75* (1), 259–293.

Amador, Manuel and Kyle Bagwell, "Money Burning in the Theory of Delegation," Working paper, Stanford University 2013.

_ **and** _ , "The theory of optimal delegation with an application to tariff caps," *Econometrica*, 2013, *81* (4), 1541–1599.

— , **Iván Werning, and George-Marios Angeletos**, "Commitment vs. Flexibility," *Econometrica*, 2006, *74* (2), 365–96.

Ambrus, Attila and Georgy Egorov, "Delegation and Nonmonetary Incentives," Working paper, Harvard University May 2009.

Armstrong, Mark and David E. M. Sappington, "Recent Developments in the Theory of Regulation," in Mark Armstrong and Robert Porter, eds., *Handbook of Industrial Organization*, Vol. 3, North-Holland, Amsterdam, 2007, pp. 1557–1700.

— **and John Vickers**, "A Model of Delegated Project Choice," *Econometrica*, 1 2010, *78* (1), 213–244.

Athey, Susan, Andrew Atkeson, and Patrick J. Kehoe, "The Optimal Degree of Discretion in Monetary Policy," *Econometrica*, 2005, *73* (5), 1431–1475.

— **, Kyle Bagwell, and Chris Sanchirico**, "Collusion and Price Rigidity," *The Review of Economic Studies*, 2004, *71* (2), 317–349.

Baron, David P., "Design of Regulatory Mechanisms and Institutions," in Richard Schmalensee and Robert D. Willig, eds., *Handbook of Industrial Organization*, Vol. 2, North-Holland, Amsterdam, 1989, pp. 1347–1447.

— **and Roger B. Myerson**, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 1982, *50* (4), 911–930.

Church, Jeffrey and Roger Ware, *Industrial Organization: A Strategic Approach*, McGraw-Hill, 2000.

Frankel, Alexander, "Aligned Delegation," Working paper, Stanford GSB December 2010.

Holmstrom, Bengt, "On Incentives and Control in Organizations." PhD dissertation, Stanford University 1977.

Joskow, Paul L. and Richard Schmalensee, "Incentive Regulation for Electric Utilities," *Yale Journal of Regulation*, 1986, *4*, 1–49.

Laffont, Jean-Jacques and Jean Tirole, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 1986, *94* (3), 614–41.

— **and** — , *A Theory of Incentives in Procurement and Regulation*, MIT Press: Cambridge, MA, 1993.

Loeb, Martin and Wesley A. Magat, "A Decentralized Method for Utility Regulation," *Journal of Law and Economics*, 1979, *22* (2), 399–404.

Martimort, David and Aggey Semenov, "Continuity in mechanism design without transfers," *Economics Letters*, 2006, *93* (2), 182–189.

Melumad, Nahum D. and Toshiyuki Shibano, "Communication in Settings with No Transfers," *The RAND Journal of Economics*, 1991, *22* (2), 173–198.

Milgrom, Paul and Ilya Segal, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, March 2002, *70* (2), 583–601.

Mylovanov, Tymofiy, "Veto-based delegation," *Journal of Economic Theory*, January 2008, *138* (1), 297–307.

Schmalensee, Richard, "Good Regulatory Regimes," *The RAND Journal of Economics*, 1989, *20* (3), 417–36.