Monopoly Regulation with Non-Paying Consumers

PRELIMINARY AND INCOMPLETE

Martín Besfamille* Nicolás Figueroa† León Guzmán‡

May 30, 2017

Abstract

This article is concerned with studying the effects of delegation in a monopoly regulation context. A regulator, with welfare maximization objectives, fixes the price and makes transfers to a monopolist which faces non-paying consumers. A costly technology, which influences the evasion level, exists. In a full information scenario, the regulator not only chooses the price and the transfers so as to maximize social welfare, but also decides the optimal level of investment. This document studies the moral hazard situation that arises when the regulator must delegate the investment decision to the firm and the monopolist undertakes the investment so as to maximize profit. This is a common situation since usually the firm's know-how is required in order to undertake an efficient investment. It is shown that under certain conditions, delegation implies a lower level of anti-evasion technology than socially optimal. The effect over price is however ambiguous and will ultimately depend in the nature of price and investment as complements or substitutes for the regulator. It is also shown that these effects depend on the nature of the firm's utility. Comparative statics exercises are done with respect to the cost of transfers for society.

*Instituto de Economía, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile; Email:mbesfamille@uc.cl.

†Instituto de Economía, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile; Email: nicolasf@uc.cl.

‡MA Student Instituto de Economía, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile; Email: leguzman@uc.cl.

1 Introduction

Public utility markets are important for the economy. Markets for provision of electricity, water, and public transportation, among others, are not only big, but also facilitate the functioning of other activities in the economy. Given their scope, and considering they usually operate as natural monopolies, the regulator finds a special interest in carefully regulating such markets. Although such regulatory exercise has as first order objective to pursue efficiency by means of regulating the price, as long as the achievement of this first goal depends on it, the regulator will also have to assure the participation of the firm. This is particularly true in a context in which the monopoly faces payment evasion since this harms the firm's profit.

Transfers from the regulator to the firm are a straightforward way of tackling the latter problem. Many concerns, however, could be raised on the use of transfers. First, the transfers must be financed. Achieving this by means of taxation in other markets implies pushing prices away from the first best marginal cost in such markets. Moreover, tax collection is usually inefficient which implies resources are lost in the process. Second, as NG (1974) points out, even political implementation of transfers is a major concern since usually the set of taxpayers financing the transfers and consumers who benefit from them are not the same which hinders its implementation. A second possible option is to directly attack the evasion, the root of the problem, by means of investing in an anti-evasion technology. The firm may invest in such technology and be able to change the probability of evasion detection and therefore the expected cost the evaders face.

In a full-information scenario, the regulator not only chooses the socially optimal price and transfers scheme, but also decides the optimal level of investment in the anti-evasion technology. This is, however, usually an utopia. The nature of the investment could require the firm to undertake it in order for it to be efficient. Therefore, the best the regulator can do is to decide the optimal price and transfers scheme, and to delegate the investment decision to the firm.

A moral hazard problem appears naturally. When considering the investment decision, both the regulator and the firm take into consideration the cost of investment and its effect over the firm's profits. The regulator, however, takes into consideration other factors. First, only the regulator cares about the welfare that stems from agents consumption; not only the formal consumer's utility but it also takes into consideration evader's utility. Although increasing the investment level, ceteris paribus, affects the social welfare because it imposes an extra cost on informal agents, it allows the regulator to reduce inefficiencies on price whose effects exceed the latter costs. Second, high anti-evasion technology investment reduces the need of financing the firm via transfers. Since transfers are potentially more costly than investment, the regulator finds interest in doing so. It must be noticed that the firm's and the regulator's interests are not completely orthogonal as is usual in the basic principal-agent moral hazard framework (principal likes effort and low salary while agent likes high salary and low effort). Such perpendicularity, however, is by no means a necessary condition for the moral hazard problem to arise. It suffices there to exist certain misalignment of interests between the agents as is the case in the world described here (both care about the profits but only the regulator cares about the consumption welfare and the cost of transfers for society). As is usual in the moral hazard literature, interests misalignment must be addressed by means of contingent transfers which produce the right incentives for the firm to invest. This article tries to inquire in the nature of the distortions that appear when delegation of the investment decision takes place. Instead of following the usual modeling in moral hazard literature and considering ex-ante voluntary participation, ex-post voluntary participation is required. Such modeling seems to be close to reality for this context since when regulating natural monopolies, the regulator cannot commit on not assuring a minimum utility level to the firm when the bad state of nature takes place.

We separate our analysis in two cases. In case I it is assumed that the firm's profit consists only in the transfers received from the monopolist and the cost of the investment. The regulator, therefore bears the cost of the production process and receives the revenues from sales. This modeling is in line with the standard moral hazard literature in a regulation context. In this scenario, the standard sub-optimal investment under delegation appears. We also describe how delegation implies distortions over the price placed by the regulator in this case. The direction of such distortions, however, is ambiguous and depends on the nature of the price and anti-evasion technology as complements or substitutes. In case II it is assumed that the firm, besides receiving transfers and paying the cost of the investment, claims the operational profit of the production process. When the price and the anti-evasion technology are complements, the before results hold. The same is not true when considering the price and the anti-evasion technology as substitutes. Interestingly, we are not able to rule out the possibility of an opposite effect over the investment level when the price and the anti-evasion technology are substitutes. Finally, comparative statics exercises are done in the main parameter of the model, the cost of transfers.

The rest of the article is structured as follows. Section 2 presents a brief review of the evasion related literature. After the description of the model in Section 3, Sections 4 and 5, the central sections of the article, present formal definitions for the first and second best problems and derive some comparisons between the two scenarios. Comparative statics exercises are also presented. Finally, Section 6 concludes. All omitted proofs are found in Appendix *A*.

2 Literature

Correct evasion modeling must consider how the regulatory actions affect both the extensive and intensive margins of demand. One early paper that carefully considers both dimensions is Oi (1971). This paper studies profit maximization by means of a two part tariff pricing in a context in which individuals abandon the market when utility of formal consumption fails to exceed the utility of the outside option. Oi (1974)'s modeling is not directly extendable to a context in which some individuals decide to become non-paying agents since it implicitly, by measuring consumer welfare through consumer surplus, assumes a quasi-linear utility world. This has two important consequences. First, the decision between formal and informal consumption becomes homogeneous between consumers. Since the aim of any evasion model should be to capture the heterogeneity in such decision, quasi-linear utilities are not useful. Second, since pricing in this context usually depends on the difference between the average demand and the marginal demand, and since no income effect implies both quantities are the same, trivial pricing decisions arise (e.g. $P = MC$ and participation is assured via transfers).

Our modeling technology follows NG (1974). First, it avoids the quasi-linear utility caveats mentioned before by expressing total social welfare in terms of consumer's indirect utility functions. Second, it considers socially optimal two part tariff in a general equilibrium context instead of profit maximizing pricing. NG (1974) is able to characterize optimal pricing in terms of average and marginal demand.

The latter papers, although relevant, do not directly consider a context with evasion. An early paper studying fare evasion is Boyd et al. (1989). This paper develops a simple fare evasion model considering differences between the common perceived probability of detection and the actual probability of detection. The difference is influenced by the observation of investment expenditure by the firm. Another early paper that studies evasion in the context of public transportation markets is Kooreman (1993). This paper extends Boyd et al. (1989)'s idea of perceived probability of detection and models a world in which heterogeneity in the decision of evading results from assuming differences in the perceived probability of detection among consumers. We believe it is hard to rationalize the reason of this differences in perception. After all, individuals have access to the same information when assessing such probability. In this paper we assume individuals share a probability of detection and it is instead, income differences that induce heterogeneity in the decision of being a formal versus an informal consumer. Moreover, Kooreman (1993) focuses in monetary fines as the way the firms dissuade evaders. Here, the focus is on reputational costs. This approach has the advantage of separating the cost of evasion from the individuals income which makes the problem more tractable. Also, since evasion monitoring is very costly, considering reputation costs, which can be interpreted as propaganda without the need of inspectors, seems to be close to reality in many markets.

More recent research has studied evasion in an exclusive empirical manner. Killias et al. (2009) is concerned with studying how certainty of detection affects the incentives to evade payment by studying a field experiment in Zurich in which ticket inspection was reintroduced to the public transportation system. They find evasion levels dropped even in hours of the day in which no inspection took place. Guarda et al (2016A) point out how bus system fare evasion is considerably lower in bus stops located at high income areas. This may be a consequence of high income individuals having a higher reputation cost. It is also consistent with low income individuals being more able to avoid monetary sanctions. They identify methods to address evasion. Guarda et al. (2016B) is close to our analysis in considering a situation in which inspectors are unable to fine the evaders. They suggest inspection strategies could be cost-effective even without considering fines. Barabino et al (2015) characterizes the optimal inspection level in a proof-of-payment transit system by considering passengers heterogeneity, variability of perceived inspection levels by passengers and human considerations in inspection.

This article extends NG (1974) monopoly regulation model by incorporating the moral hazard problem. The basic moral hazard modeling, as in Bolton and Dewatripont (2005), is considered; two realizations of output and a continuum of effort possibilities. We also compare the results in terms of the utility of the firm. Case I considers firm's utility as in Laffont and Tirole (1993) framework in which the regulator receives revenues from sales, bears the production cost, and gives transfers to the firm. Then, Case II, considers the scenario in which the firm claims the operational profits of the production process besides receiving transfers and paying the cost of the investment. To the best of our knowledge no research has been done in trying to understand delegation distortions in a monopoly regulation context by using the moral hazard framework.

3 The Model

Consider an economy which produces two goods. One of these, z, is produced competitively, while the other one, x, is provided by a monopolist. Consumers must take decisions in two margins. A first decision answers the question of how to consume x. Each individual may decide to be a formal consumer, which naturally implies paying a unit price for consumption of the monopolist's good, or may become an informal or non-paying agent, consuming some positive amount of good x without paying the unit price for those units. A second decision that consumers face answers the question of how much of good x to consume. A formal agent's utility is given by the solution to problem (1).

$$v(P,y) := \max_{x,z} \quad u(x) + z \quad \text{s.t.} \quad Px + z \leq y \tag{1}$$

where P denotes the price of good x and the price of good z is normalized to unity. We will denote the solution to (1) by $x_F(P)$. Utility of non-paying consumers is given by the solution to problem (2).

$$\mathring{v}(y) := \max_{x,z,x_i} \quad u(x_F + x_I) + z - \gamma h(y) \quad \text{s.t.} \quad Px_F + z \leq y \tag{2}$$

where x_I is the amount of evasion consumption, x the amount of formal units of x demanded, γ is the probability of detection (assumed to be constant from the consumers point of view), and $h(y)$ is a reputation cost that results from evading. It is said that an individual is an informal consumer iff it chooses $x_I > 0$. This is, it considers consuming some units of x informally. Lemma 3.1 characterizes the decision of informal consumers. It states the the set of informal consumers and the set of formal consumers are disjoint.

Lemma 3.1 *If individual y chooses $x_I > 0$ then it must be that $x_F(P) = 0$. This is, no individual chooses to act as an evader and a formal consumer simultaneously.*

The validity of Lemma 3.1 results from the assumption that $h \cdot y$ is paid (conditional on detection) by any consumer that chooses $x_I > 0$ regardless of x_I's magnitude. This assumption seems natural in the context studied here in which the action of evasion, and not the amount evaded, is what is punished. Lemma 3.1, therefore, implies that non-paying consumers devote all their income to consumption of good z. Also, since γ cannot be influenced by the consumers, the solution to problem (2) is such that the evasion consumption is common for all evaders and equal to the satiation consumption level which, with some abuse of notation, is denoted by x_I. Such satiation consumption is assumed to be high enough such that all formal consumers demand a quantity $x_F(P)$ below x_I. Assumption 1 imposes standard restrictions on the individual's utility function.

Assumption 1 *The function $u(\cdot) : \mathbb{R}_+ \to \mathbb{R}$ satisfies the following conditions:*

(i) $(\exists x_I \in \mathbb{R})$ *such that $u(\cdot)$ is strictly increasing and strictly concave in $(0, x_I)$ and strictly decreasing for all $x \geq x_I$.*

(ii) *$u(\cdot)$ is a $\mathbb{C}^2$ function.*

Part (i) of Assumption 1 considers the existence of a satiation point. This seems to be a natural assumption in public utility markets; the focus of this article. For example, if x is public transportation, since demanding high levels of x implies less leisure time, consumers could bound their demands by x_I. Part (i) plays a technical role and is a fairly standard assumption. Since each consumer maximizes utility, agents will choose the consumption format that gives them the higher utility level. Notice that if the reputation cost is ignored, the informal consumption utility strictly dominates the formal consumption utility [1]. Therefore, consideration of the reputation cost is what generates heterogeneity in the decision of formal against informal consumption. The income level of the individual indifferent between both consumption formats is denoted by $\hat{y}$. Such income, $\hat{y}$, is characterized by equality of both of the indirect utilities: $u(x_F(P)) + \hat{y} - Px_F(P) = u(x_I) + \hat{y} - \gamma hy$ [2]. Naturally, $\hat{y}$ responds to changes in P, h, γ, and x_I. Lemma 3.2 characterizes the consumer's decisions at each side of $\hat{y}$. Its proof follows directly from comparison of the formal and informal utilities. Figure 1 depicts graphically the argument behind Lemma 3.2 [3].

Lemma 3.2 *Consider P and γ as given. For consumers with $y > \hat{y}(P, \gamma)$ formal consumption dominates informal consumption and for consumers with $y < \hat{y}(P, \gamma)$ informal consumption dominates formal consumption.*

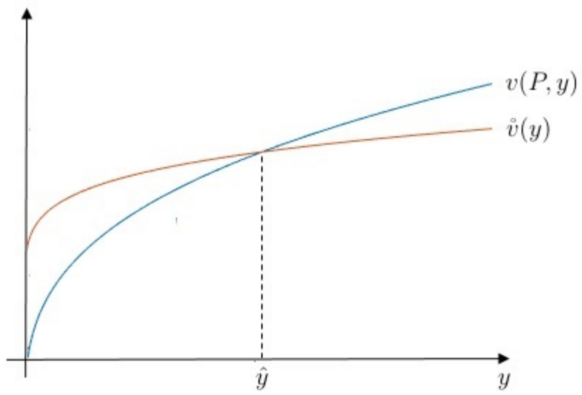

Figure 1: Formal and Informal Indirect Utility

Given a uniform population density function, $f(y)$, aggregate formal and informal consumption are defined as $\hat{X} := \int_{\hat{y}}^{\overline{y}} x_F(P) f(y) dy$ and $\mathring{X} := \int_{\underline{y}}^{\hat{y}} x_I f(y) dy$, respectively. It will also be useful to define $M := \int_{\hat{y}}^{\overline{y}} f(y) dy$; the number of formal consumers in the economy.

A world in which the probability of detection is restricted to take two values, $\gamma \in$

[1]Informal consumers enjoy a higher consumption of x and a higher consumption of z.

[2]It is possible that for all incomes in $[\underline{y}, \overline{y}]$ one consumption format dominates the other format. In this case we define $\hat{y} = \underline{y}$ if formal consumption dominates informal and $\hat{y} = \overline{y}$ if informal consumption always dominates. We ignore this trivial scenarios.

[3]Lemma 3.2's results are partially a consequence of the assumption that the cost function $h(\cdot)$ is increasing in income y. We want to emphasize that this modeling does not responds to an ideological vision. Instead it tries to capture, in our modeling, the positive correlation between evasion and income observed.

$\{\gamma_h, \gamma_l\}$ is considered. There exists a costly anti-evasion technology which allows the firm to influence the probability distribution over the realizations of γ as described by the function $Pr(\gamma = h|a) := \rho(a)$; where $a \in [0, \infty)$ is the investment level. Notice that since only two realizations of γ are possible, the whole distribution over γ can be characterized by describing the probability of high γ, γ_h, conditional on investment level. Assumption 2 imposes structure on the conditional probability distribution over the probability of detection.

Assumption 2 *The function $\rho(\cdot) : \mathbb{R}_+ \to [0, 1]$ satisfies the following conditions:*

(i) *$\rho(\cdot)$ is strictly increasing and strictly concave in all its domain.*

(ii) *$\rho(\cdot)$ is $\mathbb{C}^2$ and such that $\rho^{'}(0) > 1$.*

The assumption that $\rho(\cdot)$ is an increasing function of a captures the idea that higher investment levels make the event of γ_h to take place more likely. The assumption of $\rho(\cdot)$ being concave resembles the standard decreasing marginal value argument. Notice that since a quasi-linear utility world is considered, the demand of any formal individual is implicitly defined by: $u^{'}(x_F(P)) = P$. This implies that the marginal utility of consuming x_F, $u^{'}(\cdot)$, coincides with the inverse demand function x_F^{-1}. Therefore, the following identity holds:

$$u(x_F(P)) = \int_0^{x_F(P)} u^{'}(x)dx = \int_0^{x_F(P)} x_F^{-1}(x)dx$$

By a similar argument, $u(x_I)$, can be expressed as:

$$u(x_I) = \int_0^{x_I} u^{'}(x)dx = \int_0^{x_I} x_F^{-1}(x)dx$$

Integrating the indirect utilities of formal and informal consumers leads to the following expression for the consumer surplus $S(P, \gamma) - MPx_F(P)$:

$$\begin{aligned}
... &= \int_{\underline{y}}^{\hat{y}} \mathring{v}(y)f(y)dy + \int_{\hat{y}}^{\overline{y}} v(P, y)f(y)dy \\
&= \int_{\underline{y}}^{\hat{y}} (u(x_I) + y - \gamma hy)f(y)dy + \int_{\hat{y}}^{\overline{y}} (u(x_F(P)) + y - Px_F(P))f(y)dy \\
&= u(x_I) \int_{\underline{y}}^{\hat{y}} f(y)dy + [u(x_F(P)) - Px_F(P)] \int_{\hat{y}}^{\overline{y}} f(y)dy - \gamma \int_{\underline{y}}^{\hat{y}} hyf(y)dy + \int_{\underline{y}}^{\overline{y}} yf(y)dy \\
&= M \int_0^{x_F(P)} x_F^{-1}(x)dx + (1 - M) \int_0^{x_I} x_F^{-1}(x)dx - \int_{\underline{y}}^{\hat{y}(P,\gamma)} hyf(y)dy + \int_{\underline{y}}^{\overline{y}} yf(y)dy - MPx_F(P) \\
&= S(P, \gamma) - MPx_F(P)
\end{aligned}$$

Moreover considering the firm's total revenue is given by $R(P, \gamma) := P\hat{X}(P, \gamma)$, total social surplus, $V(P,\gamma)$, is then given by equation (3):

$$\begin{aligned} V(P,\gamma) &:= S(P,\gamma) - MPx_F(P) + (1+\mu)R(P,\gamma) \\ &:= S(P,\gamma) + \mu R(P,\gamma) \end{aligned} \tag{3}$$

Assuming a constant marginal cost, total production cost is given by $C(P,\gamma) := c\big[\hat{X}(P,\gamma) + \mathring{X}(P,\gamma)\big]$. Total social welfare considers social surplus, production costs, transfers, and the investment cost of the anti-evasion technology denoted by $\Psi(a)$. Equation (4) gives the exact mathematical expression for the total social welfare.

$$\begin{aligned} SW &:= V(P,\gamma) - (1+\mu)\big[C(P,\gamma) + T\big] + (T - \Psi(a)) \\ &:= V(P,\gamma) - (1+\mu)C(P,\gamma) - \mu T - \Psi(a) \end{aligned} \tag{4}$$

Parameter $\mu \in (0,1)$, captures the cost of public resources for society. How costly is the regulatory action is captured by the magnitude of μ. We distinguish between two cases. In Case I, the firm's utility is given by (5):

$$U := T - \Psi(a) \tag{5}$$

Assumption 3 imposes standard restrictions over investment cost in the anti-evasion technology; function $\Psi(\cdot)$.

Assumption 3 *The function $\Psi(\cdot\,) : \mathbb{R}_+ \to \mathbb{R}$ satisfies the following conditions:*

(i) *$\Psi(\cdot\,)$ is strictly increasing and convex in all its domain.*

(ii) *$\Psi(\cdot\,)$ is $\mathbb{C}^2$ and such that $\Psi'(0) = 0$*

Naturally, $\Psi(\cdot)$ must be an increasing function since the technology, after all, is not cost-less. It is also assumed that the cost of investment is a convex function as is standard. The before formulation is not always realistic since in many natural monopolies the firm bears the cost of production and also receives the operational profit. Therefore, we consider a second scenario, Case II, which takes into account this possibility. In the main sections the results between the two cases will be contrasted. Under Case II modeling, total social surplus is given by equation (6) [4]:

$$\begin{aligned} V(P,\gamma) &:= S(P,\gamma) - MPx_F(P) + R(P,\gamma) \\ &:= S(P,\gamma) \end{aligned} \tag{6}$$

[4]Since the regulator does not receives the revenues from sales, the value of public funds μ is not considered in the social surplus equation

The operational profit of the firm is denoted by $\Pi(P,\gamma) := (P-c)\hat{X}(P,\gamma) - c\mathring{X}(P,\gamma)$. Total Social welfare, and the utility of the firm in the second scenario are given by equations (7) and (8), respectively. .

$$\begin{aligned} SW &:= V(P,\gamma) - C(P,\gamma) - (1+\mu)T + (T - \Psi(a)) \\ &:= S(P,\gamma) - C(P,\gamma) - \mu T - \Psi(a) \end{aligned} \tag{7}$$

Assumption 4 imposes some restrictions over the nature of the firms profits which captures the idea that the firm studied must be financed in order to achieve the target $\overline{\pi}$.

Assumption 4 $(\forall a \in [0,\infty)) \quad \Pi(P,\gamma) - \Psi(a) \leq \overline{\pi}$.

In words, for any level of investment, the firms profits are below the target level when transfers are zero. Intuitively, assumption 4 states that the firms problems go beyond the evasion problems and that the regulator will not be able to finance the firm entirely with investment in anti-evasion technology. The problem is therefore a question of how much to finance with each instrument and not which instruments to use. Notice this is not a hard assumption since after all, the benefit of investing is bounded by the profits when the realization of γ is γ_h.

Finally, regardless of which case is considered, the timing of interaction is as follows: In $t = 1$ the regulator establishes the price and the contingent transfers to the firm. Two possibilities arise in $t = 2$. In a full information scenario the regulator chooses the socially optimal level of investment a. In a second best context, the firm chooses the investment level without the regulator being able to observe the decision. In $t = 3$ consumption takes place, both the regulator and the firm learn about the realization of γ, and the regulator gives contingent transfers to the firm.

3.1 Assumptions Discussion

Reputation Cost

Before continuing, it seems important to discuss some of the assumptions in the modeling. First, it is necessary to motivate the way in which evasion costs have been modeled. It is possible to model the evasion cost as a monetary fine. This article, instead, focuses in the case in which the cost is of a reputation nature. This may be interpreted in two different ways. The most straightforward interpretation corresponds to the firm hiring inspectors which pursue evaders in the market. In the case of being caught, an individual pays no monetary sanction, but must face a reputation cost when society becomes aware of his actions. Notice it is possible to state the latter argument without the need of the firm contracting inspectors. The important aspect is visibility; other individuals being able to observe evaders actions. For example, a second interpretation argues that the evaders cost could result from some individuals in society looking down evasion with contempt. In this context, the anti-evasion technology investment can be interpreted as the firm investing in propaganda that increases society's dislike for evasion. Such anti-evasion technology has some attractive features. Actual persecution of evaders in large markets could be extremely costly. For example, in public transportation systems it would require hiring inspectors to patrol the routes. In practice, although many firms formally impose monetary sanctions to evaders, given the very low probability of being caught, it seems plausible that it is reputation cost (in some of its many forms), and not monetary fines, that discourages evasion.

Also notice that even when the focus is on reputation and not on fines, one could argue that $h \cdot y$ could be modeled as a reputation reward formal consumers receive. This alternative formulation implies changes with regard to why the regulator values a change in the probability of detection γ. Intuitively, in a reputation cost world, increases in γ, ceteris paribus, reduce the total consumer surplus since continuity assures that the only relevant effect over the is the reduction in utility of informal individuals that remain as informal after the increase in γ. The gains from increasing γ have to do with an increase in the firms revenues and with a production cost reduction. When modeling a reputation reward world the gains from increasing γ have to do with rewarding honest consumers. Since we are interested in capturing in the model the firm's operational benefits from increasing γ we stick to the former formulation. Results, when considering moral rewards are of a similar nature.

Constant Evasion Consumption

A second point worth to discuss is the modeling of evasion consumption as constant for all individuals. The important assumption behind this is that changes in consumption do not affect the probability of detection of each individual. In this sense, a good example of a market in which this is true is the electricity provision market. Changing the amount of evasion in such markets has no effect on the probability of detection. The probability of being detected is the same even when evader's consumption is zero. What is important is whether or not the consumer has the technical set up required for evasion [5]. Perhaps, public transportation is an example of a market in which such assumption is not satisfied since the probability of detection is increasing in the amount of trips.

The modeling could be changed so as to allow the probability of detection to depend on the evasion level. The problem of evaders would then be given by:

[5] Such technical set up cannot be easily removed when individuals are not consuming

$$
\begin{aligned}
\mathring{v}(y) \coloneqq \max_{x,z,x_i} \quad & u(x+x_I)+z-\gamma(x_I)h(y) \\
\text{s.t.} \quad & Px+z \leq y \\
& z \leq y
\end{aligned}
$$

Now, since $\gamma(\cdot)$ is a function of y, it is possible that the optimal level of evasion changes with income. Moreover, Lemma 3.1 does not holds. However, the results that follow in this article are qualitatively similar under this new specification. Now increases in γ have a differentiated effect over informal's utility but in any case evasion is discouraged (the qualitative nature of the indifferent consumer would not change). Therefore, we stick with the constant probability of detection modeling. One can also argue that price effects should be considered in the informal's problem. However, there seems to be no good reason for why it should be the case. Although prices should affect the decision of being informal versus being formal, conditional on the decision of evading, prices should play no role in the problem of deciding how much to evade.

Stochastic Anti-Evasion Technology

Finally, an important question is why the anti-evasion technology has been modeled as affecting the probability distribution of γ. It is also possible to model a world in which investment deterministically affects the realization of the probability of detection. Two reasons explain why the article chooses the former over the latter.

First, it must be noticed that observability of the firms actions is a crucial aspect of the moral hazard problem. Suppose that investment in anti-evasion technology, a, deterministically changes the probability of detection as described by a technology function λ; $a \to \lambda(a)$. If the regulator knows the function λ, which seems as the natural assumption, then after actions take place and the level of evasion is known, the regulator could exactly calculate the level of investment of the firm by considering the function λ^{-1}. This implies that the motivation in the second best scenario vanishes. The regulator could always observe the firm's effort and therefore the question is whether or not it is capable of punishing the firm harshly so as to implement the first best results. Since this is not the focus of this article such modeling is ignored.

4 Full Information and Second Best - Case I

In this section we focus in Case I modeling. Therefore, social welfare and the firm's utility are given by functions (4) and (5), respectively. Since this article is concerned with a regulator that cares about investing in order to avoid evasion, it must be the case that SW is an increasing function in γ. Moreover, we are interested in whether the price and the anti-evasion technology are complements or substitutes as regulatory instruments; are increases in the investment level more desirable when price level is low? Let's consider the first issue. First, differentiation of (4) with respect to γ leads to the following expression:

$$\frac{\partial SW}{\partial \gamma} = \frac{\partial S(P,\gamma)}{\partial \gamma} + \mu\frac{\partial R(P,\gamma)}{\partial \gamma} - (1+\mu)\frac{\partial C(P,\gamma)}{\partial \gamma} \tag{8}$$

Direct examination of the three derivatives shows that the total effect of an increase in γ is ambiguous. In one hand $\frac{\partial S(P,\gamma)}{\partial \gamma} < 0$ changes in γ generate two effects: (i) the utility of informal consumers decreases, (ii) some consumers move from informal consumption to formal consumption. Since we have assumed continuity in utility, this last effect sums to zero at an aggregate level leaving only the negative effect alive. However $\mu\frac{\partial R(P,\gamma)}{\partial \gamma} - (1+\mu)\frac{\partial C(P,\gamma)}{\partial \gamma} > 0$ since increasing γ means more formal consumption, which increases sales revenues, and lowers production costs because of the reduction in evasion. Lemma 4.1 establishes that if the evasion consumption is high enough, the gains from increasing γ offset the negative effects of doing so.

Lemma 4.1 *(SW as an increasing function of γ): If the evasion consumption level is high enough, $x_I > \frac{h}{cf(\cdot)}$, and the marginal formal consumer's price elasticity is below one, $-\frac{\partial X_F(P)}{\partial P}\frac{P}{X_F(P)} < 1$, then $\frac{\partial SW}{\partial \gamma} > 0$.*

Intuitively, the benefits of increasing γ increase with x_I since high evasion levels make the production process more costly and therefore the gains from avoiding it higher. Since we are interested in studying a regulator which finds attractive investing in theanti-evasion technology, we assume the premises of Lemma 4.1 hold on for what remains of section 4.

We now turn our attention to the relationship between γ and P as complements or substitutes from the regulator's point of view. Direct differentiation leads to (7).

$$\frac{\partial SW}{\partial P \partial \gamma} = \frac{\partial S(P,\gamma)}{\partial P \partial \gamma} + \mu\frac{\partial R(P,\gamma)}{\partial P \partial \gamma} - (1+\mu)\frac{\partial C(P,\gamma)}{\partial P \partial \gamma} \tag{9}$$

Again, the sign of $\frac{\partial SW}{\partial P \partial \gamma}$ is ambiguous. Intuitively, $\frac{\partial S(P,\gamma)}{\partial P \partial \gamma} < 0$ since an increase in P makes increases in γ less attractive since a higher γ increases the amount of formal consumers which now pay a higher P. This effect, ceteris paribus, implies that the anti-evasion technology and the price act as substitutes. However, as a second effect, increasing P makes it more attractive to increase γ since this translates into more formal consumers paying P and therefore higher sales revenues. In this last effect anti-evasion technology and price act as complements. As will be seen, how this ambiguity is solved, has important consequences over the regulator's price decision when going from a full information scenario to a second best scenario. Instead of considering sufficient conditions for the sign of (10) to be clear, we will characterize results for both cases.

4.1 Full Information and Second Best Comparison

In a first best context the regulator, besides choosing P and T, decides the optimal level of investment in the anti-evasion technology. This implies that besides voluntary participation (VP), the regulator faces no other constraint. Also notice that, ex-post VP constraints are considered instead of ex-ante VP as is common in the moral hazard literature. The reason has to do with the context in which we are applying the moral hazard framework. Natural monopolies are usually sensible markets. This implies that the regulator lacks the commitment to assure the firm it will not give transfers to it if ex-post the firm faces losses. Considering ex-post VP constraints is then the natural way to model such lack of commitment. A formal definition of the solution to the first best problem is given by (11).

Definition A Full Information (FI) contract is defined as a solution to the following maximization problem:

$$\max_{a,P,T_h,T_l} \quad \mathbb{E}_\gamma \Big\{ V(P,\gamma) - (1+\mu)C(P,\gamma) - \mu T \Big\} - \Psi(a) \tag{10}$$

$$\text{s.t.} \quad T_h - \Psi(a) \geq \overline{\pi}$$

$$T_l - \Psi(a) \geq \overline{\pi}$$

Notice this is a complex problem in four control variables. We follow Grossman and Hart (1983)'s approach by dividing the problem into two sub-problems. We first consider the minimum cost at which a can be implemented while fixing a given level of P. Implementation, in this context, means satisfying the ex-post VP constraints. The existence of a solution to such problem is guaranteed since it is a convex optimization problem with linear constraints. Formally, the sub-problem is characterized by (12).

Definition The minimum cost of implementing a for a given level of P, $\Omega^{FI}(a,P)$, in the Full Information context, is characterized by the following minimization problem:

$$\Omega^{FI}(a,P) := \min_{T_h,T_l} \quad \Big\{ \rho(a)T_h + (1-\rho(a))T_l \Big\} \tag{11}$$

$$\text{s.t.} \quad T_h - \Psi(a) \geq \overline{\pi}$$

$$T_l - \Psi(a) \geq \overline{\pi}$$

Notice that since the regulator does not need to induce high investment, and given the transfers are costly, it will design a contract such that both ex-post VP constraints are binding. If this were not the case, the regulator could reduce transfers such that both constraints still hold while the implementation cost is reduced. Lemma 4.2 captures this intuition.

Lemma 4.2 *(Binding ex-post VP constraints): For $a > 0$, the solution of problem* (12) *is such that $T_h - \Psi(a) = \overline{\pi}$ and $T_l - \Psi(a) = \overline{\pi}$.*

The second stage problem then considers choosing price and the level of anti-evasion technology investment that maximizes expected social welfare considering the function $\Omega^{FI}(a, P)$. Formally, the problem is given by (13).

Definition The solution to the Full Information second sub-problem, a^{FI}, P^{FI} is characterized by the following maximization problem:

$$\max_{a,P} \; SW^{FI}(a,P) := \mathbb{E}_\gamma \Big\{ V(P,\gamma) - (1+\mu)C(P,\gamma) \Big\} - \Psi(a) - \mu\Omega^{FI}(a,P) \tag{12}$$

As stated before, usually the regulator is not able to directly choose the level of investment, a. The nature of the investment could require delegation of the decision in order for it to be efficient. The regulator's second best problem is defined by (14).

Definition A Second Best (SB) contract is defined as a solution to the following maximization problem:

$$\max_{a,P,T_h,T_l} \; \mathbb{E}_\gamma \Big\{ V(P,\gamma) - (1+\mu)C(P,\gamma) - \mu T \Big\} - \Psi(a) \tag{13}$$

$$\text{s.t.} \;\; T_h - \Psi(a) \; \geq \overline{\pi}$$

$$T_l - \Psi(a) \; \geq \overline{\pi}$$

$$a \in \underset{\tilde{a}}{\operatorname{argmax}} \, \mathbb{E}_\gamma \left\{ T - \Psi(\tilde{a}) \right\}$$

The crucial difference between FI and SB problems is that since now the regulator delegates the decision of investment to the firm, it must then be the case that given price and contingent transfers, the firm chooses a so as to maximize its utility. We capture this intuition modeling the problem as if the regulator still chooses a but faces an incentive compatibility constraint which dictates that it must choose the profit maximizing a after the election of P and the transfers T_i; $i \in \{H, L\}$. In an analogous fashion to the first best problem, we break the SB problem into two sub-problems by finding the minimum cost of implementation of P and a before choosing them.

Definition The minimum cost of implementing a for a given level of P in the second best context, $\Omega^{SB}(a,P)$, is characterized by the following minimization problem:

$$\Omega^{SB}(a,P) := \min_{T_h, T_l} \left\{ \rho(a)T_h + (1-\rho(a))T_l \right\} \tag{14}$$

$$\text{s.t.} \quad T_h - \Psi(a) \geq \overline{\pi}$$

$$T_l - \Psi(a) \geq \overline{\pi}$$

$$a \in \underset{\tilde{a}}{\operatorname{argmax}}\, \mathbb{E}_\gamma \{T - \Psi(\tilde{a})\}$$

Lemma 4.3 is analogous to Lemma 4.2. It establishes that conditional on the investment level being positive, the solution to problem (15) involves giving some informational rent too the firm. This result is standard in moral hazard literature. Since now the regulator delegates the decision of a, any investment must be induced by giving higher transfers in case of the realization of γ_h.

Lemma 4.3 *(Non-Binding ex-post VP Constraint): For $a > 0$, the solution of problem* (10) *is such that $T_h - \Psi(a) > \overline{\pi}$ and $T_l - \Psi(a) = \overline{\pi}$.*

Also notice that the extra restriction, naturally, leads to a function $\Omega^{SB}(a,P)$ that responds differently to a and P than $\Omega^{FI}(a,P)$. Lemma 4.4 captures this by qualitatively characterizing the derivatives of $\Omega^i(a,P)$; $i \in \{SB, FI\}$ and comparing their magnitudes. The before lemmas will be useful for establishing further results.

Lemma 4.4 *(Properties of $\Omega^i(a,P)$; $i \in \{SB, FI\}$): If $a > 0$, then the cost functions that result from the minimization sub-problem satisfy:*

(i) $\frac{\partial \Omega^{SB}(a,P)}{\partial a} \geq \frac{\partial \Omega^{FI}(a,P)}{\partial a} > 0$

(ii) $\frac{\partial \Omega^{SB}(a,P)}{\partial P} = \frac{\partial \Omega^{FI}(a,P)}{\partial P} = 0$

Intuitively, Lemma $4,4$ establishes that achieving high anti-evasion technology investment levels involves a cost beyond that in FI in the second best scenario. Since the firm dislikes investing, the regulator faces the additional cost of having to create the right incentives for the firm to invest. Notice that both implementation cost functions (FI and SB) do not depend on P. Clearly this depends on the fact that Case I considers the utility of the firm as only depending on the transfers and in the investment cost. Changes in P play no role in giving incentives to the firm. In Case II, we change the modeling so as to incorporate the effects of the price.

Once the cost function is obtained from the cost minimization sub-problem (15), we consider the sub-problem of finding the optimal level of investment, a, and price, P, considering the reduced form of the minimum implementation cost $\Omega^{SB}(a,P)$. Formally this sub-problem is given by (16).

Definition The solution to the Second Best second sub-problem, a^{SB}, P^{SB} is characterized by the following maximization problem:

$$\max_{a,P} \; SW^{SB}(a,P) := \mathbb{E}_\gamma \Big\{ V(P,\gamma) - (1+\mu)C(P,\gamma) \Big\} - \Psi(a) - \mu\Omega^{SB}(a,P) \tag{15}$$

Breaking the problem into two separates steps allows us to compare the optimal investment and price levels between the full information and the second best scenarios. Proposition 4.5, formally states this result.

Proposition 4.5 *(Full Information and Second Best Comparison): The vectors* (a^{FI}, P^{FI}) *and* (a^{SB}, P^{SB}) *satisfy the following properties:*

(i) a^{SB}*, is below (weakly),* a^{FI}*.*

(ii) *If* $\frac{\partial SW}{\partial P \partial \gamma} > 0$*, then* P^{SB} *is below (weakly)* P^{FI}*.*

(iii) *If* $\frac{\partial SW}{\partial P \partial \gamma} < 0$*, then* P^{SB} *is above (weakly)* P^{FI}*.*

Proposition 4.5 is interesting. In one hand, it establishes the standard result of sub-optimal investment under delegation. However it also states that the effect over the price level is ambiguous and will ultimately depend on the price and the anti-evasion technology being complements or substitutes. Intuitively, delegating the investment decision makes implementation of a more costly. This leads the regulator to reduce the investment level it induces. If $\frac{\partial SW}{\partial P \partial \gamma} > 0$ price and anti-evasion technology are complements from the regulator's point of view, since a is lower, the marginal benefit from increasing the price is now lower which leads to a reduction in P. The opposite effect holds if $\frac{\partial SW}{\partial P \partial \gamma} < 0$. Although we leave a detailed proof for Appendix A, the highlights of the proof are as follows. Consider the following maximization problem:

$$\max_{a,P} \; \tilde{SW}(a,P,t) := \; (1-t) \cdot SW^{SB}(a,P) + t \cdot SW^{FI}(a,P) \tag{16}$$

where $t \in (0,1)$ is an artificial parameter. In Appendix A it is shown that $\tilde{SW}$ is a supermodular function in (a,P,t) or $(a,-P,t)$ depending on the nature as complements or substitutes of P and a. In either case, the comparative statics exercise of a and P with respect to parameter t are clear since problem (17) is an euclidean unconstrained optimization problem. We are particularly concerned with what happens to the argmax set when t changes from $t = 0$ to $t = 1$ (why?). The reason follows from noticing that $\tilde{SW}(a,P,t=0) = SW^{SB}(a,P)$ and $\tilde{SW}(a,P,t=1) = SW^{FI}(a,P)$. Therefore, the before comparative statics exercise is actually a comparison between the second best and the full information problems solutions.

It is also interesting to think of the consumer welfare implications of this result. Complementarity between price and the anti-evasion technology clearly leaves consumers better since a lower P implies higher utility of formal consumers. It is true that some individuals may change their formal-informal status because of the change in a. However, continuity assures that this changes have no effect over total consumer surplus at an aggregate level. The result when P and a are substitutes is not clear since both generate opposite effects.

4.2 First Order Condition Approach

The incentive compatibility constraint in the SB problem has been framed in terms of a maximization problem of the firm. As shown in appendix A, we have characterized the incentive compatibility constraint by means of the first order condition of such problem. This approach, in general, implies looking at a relaxed problem. Following Rogerson (1985), certain conditions must hold in order to assure the approach is valid in this case. The two requirements are:

- **(i)** (MLRP): $(\forall a \in [0, \infty))$ $\frac{\rho^{'}(a)}{\rho(a)} > \frac{-\rho^{'}(a)}{(1-\rho(a))}$.
- **(i)** (CDF): $\big(\forall a, a^{'} \in [0, \infty)\big)$ $(\forall \alpha \in [0, 1])$ $F(\gamma_i, \alpha a + (1 - \alpha)a^{'}) \leq \alpha F(\gamma_i, a) + (1 - \alpha)F(\gamma_i, a^{'})$

where $F(\cdot)$ is the cumulative distribution function of the random variable γ. The first condition is the monotone likely-hood ratio property which is satisfied in our model given that only two outcome realizations are considered (γ_h, γ_l). Condition (ii) is a convexity property which is assumed to hold throughout the article. Throughout the article it is assumed that such conditions hold so that the incentive compatibility constraints are well captured by the firm's FOC.

4.3 Comparative Statics

One interesting question is what happens to the optimal (first best or second best) anti-evasion technology investment and the price when certain parameters of the model change. In particular we are concerned with understanding the response to changes in μ. Proposition 4.6 starts by tackling the before question for Case I. As will be seen in Section 5 the analysis is different when considering Case II scenario.

Proposition 4.6 *(Comparative Statics Exercise in μ): The solution vector (a^i, P^i); $i \in \{FI, SB\}$ satisfies the following properties:*

(i) *a^i is (weakly) decreasing in μ.*

(ii) *If $\frac{\partial SW}{\partial P \partial \gamma} < 0$, then P^i is (weakly) increasing in μ.*

(iii) *If $\frac{\partial SW}{\partial P \partial \gamma} > 0$, then P^i is (weakly) decreasing in μ.*

Notice, again, that the effect over the optimal investment level is unambiguous. It is interesting to see, as in proposition 4.5, that the effect over the optimal price depends on the nature as complements or substitutes of the price and the investment for the regulator. Intuitively, an increase in μ makes transfers, under any realization of γ more costly. Since $T_h \geq T_l$ (strictly in the second best context), the regulator wants to avoid both transfers; in particular the high one, T_h. This means that the regulator reduces its investment level so as to make less probable the realization of γ_h and therefore less probable having to pay T_h. Changing μ has no direct effect over the marginal value of changing P since Lemma 4.4 assures that Ω^i does not depends on P. However, it does has an indirect effect through a.

5 Full Information and Second Best - Case II

Up to this point, we have consider a situation in which the regulator bears the cost of production and receives the revenues from sales. In such a world the utility of the firm only depends of transfers and the cost of investment; in particular it is constant with respect to changes in the price. Although, this modeling is congruent with the standard moral hazard framework, it has two limitations.

(i) In practical grounds, although in some public utility monopolies such assumptions hold, in general this is not the case.

(ii) In theoretical terms, the before modeling ignores the role of price as an instrument to induce high investment.

In this section we depart from this formulation and contrast the results with those found in section 4. Since the arguments of this section are of a similar nature of those in section 4 the exposition is quick and less detailed. Also, in order to keep the exposition simple, the notation in this section is identical to the one in section 4. It should be clear from the context which case we are talking about. We start by establishing sufficient conditions for the positiveness of $\frac{\partial SW}{\partial \gamma}$ as in Section 4. Notice that now SW is given by equation (7).

Lemma 5.1 *(SW as an increasing function of γ): If the evasion consumption level is high enough, $x_I > \frac{h}{cf(\cdot)}$, then $\frac{\partial SW}{\partial \gamma} > 0$.*

In this new scenario the Full Information and the Second best problems are analogous to the ones defined in Section 4 except for considering equation (8) as the firm's utility. The following programs give exact definitions of each problem:

Definition A FI contract is defined as a solution to the following maximization problem:

$$\max_{a,P,T_h,T_l} \quad \mathbb{E}_\gamma \left\{ V(P,\gamma) - C(P,\gamma) - \mu T \right\} - \Psi(a) \tag{17}$$

$$\text{s.t.} \quad \Pi(P,\gamma_h) + T_h - \Psi(a) \geq \overline{\pi}$$

$$\Pi(P,\gamma_l) + T_l - \Psi(a) \geq \overline{\pi}$$

Definition A SB contract is defined as a solution to the following maximization problem:

$$\max_{a,P,T_h,T_l} \quad \mathbb{E}_\gamma \left\{ V(P,\gamma) - C(P,\gamma) - \mu T \right\} - \Psi(a) \tag{18}$$

$$\text{s.t.} \quad \Pi(P,\gamma_h) + T_h - \Psi(a) \geq \overline{\pi}$$

$$\Pi(P,\gamma_l) + T_l - \Psi(a) \geq \overline{\pi}$$

$$a \in \underset{\tilde{a}}{\operatorname{argmax}}\, \mathbb{E}_\gamma \left\{ \Pi(P,\gamma) + T - \Psi(\tilde{a}) \right\}$$

We can reason in a similar way as in Section 4 breaking the Full Information problem into two sub-problems obtaining $\Omega^i(a,P)$; $i \in \{SB, FI\}$. Again, since no investment must be induced, in the FI problem both ex-post VP constraints will be binding. Some information rent will be given to the firm under the SB since regardless of the form the firm's utility take, the regulator must differentiate the outcome after γ_h and γ_l so as to induce $a > 0$. We state the before intuition in Lemma 5.2 and Lemma 5.3. Also, as in case I, we can study the properties of the derivatives of Ω^i. Note that now the implementation cost functions respond to changes in P. Lemma 5.2 formally states the results of such exercise. Since the mechanics of Lemmas 5.2 and 5.3 proofs are identical to the argument in their analogues in section 4, we omit the proof of them.

Lemma 5.2 *(Ex-post VP constraints FI/SB): For $a > 0$, the solutions of problem* (18) *and* (19) *are such that:*

(i) For problem (18), $\Pi(P,\gamma_h) + T_h - \Psi(a) = \overline{\pi}$ *and* $\Pi(P,\gamma_l) + T_l - \Psi(a) = \overline{\pi}$.

(ii) For problem (19), $\Pi(P,\gamma_h) + T_h - \Psi(a) > \overline{\pi}$ *and* $\Pi(P,\gamma_l) + T_l - \Psi(a) = \overline{\pi}$.

Lemma 5.3 *(Properties of $\Omega^i(a,P)$; $i \in \{SB, FI\}$): If $a > 0$, then the cost functions that result from the minimization sub-problem satisfy:*

(i) $\frac{\partial \Omega^{SB}(a,P)}{\partial a} \geq \frac{\partial \Omega^{FI}(a,P)}{\partial a} > 0$

(ii) $0 > \frac{\partial \Omega^{SB}(a,P)}{\partial P} \geq \frac{\partial \Omega^{FI}(a,P)}{\partial P}$

We can now state an analogous result to Proposition 4.5. As is shown in the proof of the proposition, the crucial difference between this result and Proposition 4.5 is that the argument for (ii) doesn't holds anymore.

Proposition 5.4 *(Full Information and Second Best Comparison): If* $\frac{\partial SW}{\partial P \partial \gamma} > 0$*, then* $a^{SB} \leq a^{FI}$ *and* $P^{SB} \geq P^{FI}$*.*

Proposition 5.3 suggests the possibility of the investment level in the Full information scenario being smaller than the investment level in the Second Best. This is interesting since it contrasts with the standard result of sub-optimal effort in the traditional moral hazard literature. Intuitively, when going from the Second Best to the Full information problem the regulator finds more attractive an increase in P and a since both reduce more the implementation cost. However, when P and a are substitutes from the regulators perspective, increases in P make it less attractive to increase a. If this last effect is large enough so as to off-set the first effect over a, it could be the case that the regulator reduces the total investment level in the Full Information scenario compared to the second Best scenario.

5.1 Comparative Statics

This section mimics the comparative static exercise of section 4.3. Proposition 5.4 is analogous to proposition 4.6. Again, both results are qualitatively identical for the FI and the SB.

Proposition 5.5 *(Comparative Statics Exercise in μ): The solution vector* (a^i, P^i); $i \in \{FI, SB\}$ *is such that if* $\frac{\partial SW}{\partial P \partial \gamma} < 0$*, then* $(-a^i, P^i)$ *is weakly increasing in* μ*.*

Notice, again, that we are not able to establish a result when the price and the anti-evasion technology are substitutes. The intuition is similar to the argument in proposition 5.1. Since in case II, the price plays a role in changing the implementation cost, the net effect of changing μ is ambiguous. Again, the possibility of counter-intuitive results of the level of investment increasing with μ cannot be ruled out and will ultimately depend on the magnitude of the crossed effects.

6 Conclusion

Throughout this article we consider the delegation problem that arises when considering regulating a monopoly facing evasion. In a full information context, the regulator can decide the optimal level of invest in a costly anti-evasion technology which reduces evasion. In a second best context, it delegates the decision to the firm which chooses investment so as to maximize it's profit. We break the analysis into two cases. Case I considers the regulator receives revenues from sales and bears the production cost. In Case II the firm besides receiving transfers and paying for the investment, receives the operational profit of the production process. In Case I the standard sub-investment result under delegation appears. Under delegation, the regulator must induce high effort through transfers. Since such transfers are costly the regulator considers a lower level of investment. The effects of delegation over the price level are ambiguous and will ultimately depend on the nature of price and the investment as complements or substitutes. In Case II, the same FI vs SB results hold under complementarity. Interestingly, however, the comparison between the full information and the delegation scenario is ambiguous when considering price and the investment as substitutes. Moreover, we are not able to rule out a situation in which the investment level is higher under delegation. Definite comparative exercises would require considering the magnitude, and not only the qualitative nature of the cross effects between the control variables. This result shows how even the most basic standard moral hazard results are sensible to changes in the context studied. Finally we have considered comparative statics exercises in the cost of public resources for both Case I and Case II.

We conclude by emphasizing some limitations of our analysis and some possible extensions. First, notice that we have ignored other possible decision variables, besides price, transfers, and the investment in the anti-evasion technology. One could think of other ways in which the regulator and firm could deal with evasion. For example, it may be argued that the quality of the service could be changed as to compensate for evasion losses. This scenario is interesting since changing the evasion level would then have an impact over formal consumers when the firm is unable to discriminate in quality. Indeed, this is a way in which many public utilities monopolists operate.

Second, it is possible to think of different interpretations for γ and h. In many scenarios evasion requires investing in a costly setup. We could then interpret h as the cost of such set-up and γ as a parameter which affects such cost. Also considering reputation rewards and monetary fines instead of reputation costs is an option. Although the mathematical workout would be similar, the interpretation of the results would be different. As regard to γ it would be interesting to see how the results change when considering a continuum of values instead of the anti-evasion technology we consider.

Finally, we have focused the comparative statics exercises in μ. The comparative exercise is technically easier in μ given the linear form in which it enters the social welfare function. Obviously, understanding how the optimal price and investment levels respond to changes in other parameters is also interesting.

Appendix A: Omitted proofs

Proof (Lemma 3.1):
Consider an individual of income y such that the solution of problem (2) leads to an evasion level of $x_i > 0$. The proof follows by contradiction. Consider that $x_F > 0$. The utility consumer y attains is:

$$u(x + x_I) + y - Px_F - \gamma hy$$

If instead, consumer y places x_I, the satiation level, he attains a utility level of:

$$u(x_I) + y - \gamma hy$$

Since x_I is the satiation point it must be that $u(x_i + x_F(P)) \leq u(x_I)$. Moreover, clearly $y - Px < y$. Therefore, since x_I and $x_F(P) = 0$ is feasible, it must be that the consumer is not maximizing its utility. Since this is a contradiction we conclude. ∎

Proof (Lemma 3.2):
Consider P and γ as given. Also consider the existence the existence of $\hat{y}(P, \gamma) \in (\underline{y}, \overline{y})$. From Lemma 3.1, we can write the informal utility of individual $\hat{y}$ as $u(x_I) + y - \gamma hy$. The result follows by proving that:

$$\frac{\partial}{\partial y}[u(x_F(P)) + y - Px_F(P)] > \frac{\partial}{\partial y}[u(x_I) + y - \gamma hy]$$

This is indeed the case since $\gamma h > 0$. Therefore, for any $y > \hat{y}(P, \gamma)$, formal consumption dominates informal consumption and for any $y < \hat{y}(P, \gamma)$, informal consumption dominates formal consumption. ∎

Proof (Lemma 4.1):
By differentiating the equality that defines $\hat{y}$:

$$\frac{\partial \hat{y}}{\partial \gamma} = -\frac{h(\hat{y})}{h'(\hat{y})\gamma}$$

$$\frac{\partial \hat{y}}{\partial P} = \frac{x_F(P)}{h'(\hat{y})\gamma}$$

Differentiation of $SW(P, \gamma)$ therefore is given by:

$$
\begin{aligned}
\frac{\partial SW}{\partial \gamma} &= \frac{\partial S(P,\gamma)}{\partial \gamma} + \mu \frac{\partial R(P,\gamma)}{\partial \gamma} - (1+\mu)\frac{\partial C(P,\gamma)}{\partial \gamma} \\
&= -\int_{\underline{y}}^{\hat{y}} h(y)\, f(y)dy - \gamma \frac{\partial \hat{y}}{\partial \gamma} h(\hat{y}) f(\hat{y}) - \mu P x_F(P) \frac{\partial \hat{y}}{\partial \gamma} f(\hat{y}) - c(x_I - x_F(P)) \frac{\partial \hat{y}}{\partial \gamma} f(\hat{y})(1+\mu) \\
&> -h(\hat{y}) \int_{\underline{y}}^{\hat{y}} f(y)dy + \frac{h^2(\hat{y})}{h'(\hat{y})} f(\hat{y}) + \mu P x_F(P) \frac{h(\hat{y})}{h'(\hat{y})\gamma} f(\hat{y}) + c(x_I - x_F(P)) \frac{h(\hat{y})}{h'(\hat{y})\gamma} f(\hat{y})(1+\mu) \\
&= \frac{h(\hat{y})}{\gamma h'(\hat{y})} f(\hat{y}) \Big[-(1 - M(\gamma)) \frac{\gamma}{f(\hat{y})} h'(\hat{y}) + \mu P x_F(P) + c(x_I - x_F(P))(1+\mu) \Big] \\
&> \frac{h(\hat{y})}{\gamma h'(\hat{y})} f(\hat{y}) \Big[-\frac{\gamma h'(\hat{y})}{f(\hat{y})} + \mu P x_F(P) + c(x_I - x_F(P))(1+\mu) \Big]
\end{aligned} \tag{19}
$$

where the sequence of equalities and inequalities follows by substituting for the derivatives of $\hat{y}$, noticing that $\int_{\underline{y}}^{\hat{y}} h(y)\, f(y)dy$ is bounded above by $h(\hat{y})(1-M)$ (since $h(y)$ is strictly increasing), and using the fact that $M \geq 0$. Since the term in $\Big[...\Big]$ is strictly increasing in P the last term of block (20) is bounded below by:

$$
\begin{aligned}
... &> \frac{-\gamma h}{f(\hat{y})} + c x_I (1+\mu) \\
&> \frac{-h}{f(\hat{y})} + c x_I \\
&> 0
\end{aligned} \tag{20}
$$

where the second inequality follows from the fact that $\gamma \in [0,1]$ and the last inequality is a consequence of the premises. ∎

Proof (Lemma 4.2):
Consider the first best sub-problem defined in equation (12). Assume that one of the participation constraints is not binding at the optimum; without loss of generality assume $T_h - \Psi(a) > 0$. The proof follows by contradiction. Notice that $(\exists \delta > 0)$ such that $(T_h - \delta) - \Psi(a) \geq 0$. Moreover, notice that the objective function is decreasing in T_h. Therefore, $T_h - \delta$ is feasible and increases the objective function which contradicts the assumption that T_h, T_l was optimal. ∎

Proof (Lemma 4.3):
Lemma 4.1 establishes that both ex-post VP constraints are binding in the full information problem. We now prove that for the second best context, the γ_h ex-post VP constraint is binding while the other one is not. The argument consists in showing any alternative leads to a contradiction. First notice that section 4.3 allows us to consider the incentive compatibility constraints in terms of the firm's FOC: $\rho'(a)\big[T_h - T_l\big] = \Psi'(a)$.

a. Consider $T_h - \Psi(a) > \overline{\pi}$ and $T_l - \Psi(a) > \overline{\pi}$ a similar argument to Lemma 5.1 leads to a contradiction.

b. Consider $T_h - \Psi(a) = \overline{\pi}$ and $T_l - \Psi(a) = \overline{\pi}$. Substituting into the FOC condition of the firm gives the following equality: $\rho^{'}(a)(0) = \Psi^{'}(a)$. Since, the premise considers $a > 0$ and $\Psi^{'}(a) > 0$ $\forall a > 0$ this leads to a contradiction.

c. Consider $T_h - \Psi(a) = \overline{\pi}$ and $T_l - \Psi(a) > \overline{\pi}$. Substituting into the FOC condition of the firm gives: $\rho^{'}(a)\big[\overline{\pi} + \Psi(a) - T_l\big] = \Psi^{'}(a)$. Because of the premise the term inside $\big[...\big]$ is negative. This contradicts the fact that $\forall a > 0$, $\rho^{'}(a), \Psi^{'}(a) > 0$.

Since all other options have been considered we conclude. ∎

Proof (Lemma 4.4):
Section 4.3 allows us to consider the incentive compatibility constraints in terms of the firm's FOC: $\rho^{'}(a)\big[T_h - T_l\big] = \Psi^{'}(a)$. The proof considers 3 steps $(i) - (iii)$.

(i) Consider problem (15) and an increase in a. Since the ex-post VP constraint for γ_l must be binding, it must be the case that T_l increases. Moreover since $\frac{\Psi^{'}(a)}{\rho^{'}(a)}$ increases in a, in order to satisfy the incentive compatibility restriction it must also be the case that T_h increases. That $\frac{\partial \Omega^{SB}}{\partial a} > 0$ is evident from writing the objective function of problem (15) as $\rho(a)(T_h - T_l)$ and considering the increases in a, T_h, T_l.

(ii) Lemma 4.2 shows that the both ex-post VP constraints are binding in the sub-problem of the full information context. Therefore increases in a lead to increases in T_h and T_l. Again, that $\frac{\partial \Omega^{SB}}{\partial a} > 0$ becomes evident from writing the objective function of sub-problem as $\rho(a)(T_h - T_l)$ and considering the fact that a, T_h, T_l all increased.

(iii) Comparing the sub-problem of minimizing the implementation cost between the full information and the second best context, the only difference is the incentive compatibility constraint in the second best problem. When a increases, incentive compatibility requires that T_h increases more than the increase in T_l for the difference $(T_h - T_l)$ to increase. Therefore it must be the case that $\frac{\partial \Omega^{SB}(a,P)}{\partial a} > \frac{\partial \Omega^{FI}(a,P)}{\partial a}$. ∎

Proof (Proposition 4.5):
Consider the following problem of maximization:

$$\max_{a,P} \ \tilde{SW}(a,P,t) := (1-t)\cdot SW^{SB}(a,P) + t\cdot SW^{FI}(a,P) \tag{21}$$

Notice that conditional on $t = 0$, problem (17) is equivalent to the second sub-problem of the second best program. Therefore the solution to both problems are analogous.

Analogously, conditional con $t = 1$, the solution to problem (17) is equivalent to the solution to the second sub-problem in the full information scenario. We can therefore do a comparative statics exercise in parameter t to obtain the comparison between the second best and the second best problems' solutions. Monotone comparative statics theory requires three conditions two hold:

(i) The constraint set of (17), conditional on $t = 1$, must dominate in strong order the constraint set of (17), conditional on $t = 0$.

(ii) The domain of $\tilde{SW}$ must be a lattice.

(iii) Function $\tilde{SW}$ must be super-modular in (a, P, t).

Conditions (i) and (ii) are clearly satisfied since problem (17) is an unconstrained optimization problem and since the domain of $\tilde{SW}$ is a rectangular subset of $\mathbb{R}^2$, which is always a lattice. Moreover, since this is an euclidean maximization problem, pairwise super-modularity is a sufficient condition for (iii) to hold. We check this by studying the cross derivatives of $\tilde{SW}$. Direct differentiation leads to:

$$\begin{aligned}
\frac{\partial^2 \tilde{SW}}{\partial a \partial t} &= -\mu \left[\frac{\partial \Omega^{FI}(a,P)}{\partial a} - \frac{\partial \Omega^{SB}(a,P)}{\partial a} \right] > 0 \\
\frac{\partial^2 \tilde{SW}}{\partial P \partial t} &= -\mu \left[\frac{\partial \Omega^{FI}(a,P)}{\partial P} - \frac{\partial \Omega^{SB}(a,P)}{\partial P} \right] = 0 \\
\frac{\partial^2 \tilde{SW}}{\partial P \partial a} &= (1-t) \frac{\partial SW^{SB}(a,P)}{\partial P \partial a} + t \frac{\partial SW^{FI}(a,P)}{\partial P \partial a}
\end{aligned} \tag{22}$$

where the signs of the two first cross derivatives follow from Lemma 4.4. If $\frac{\partial^2 \tilde{SW}}{\partial P \partial t} > 0$ then $\tilde{SW}$ is super-modular in (a, P, t). Topkis' monotonicity theorem implies that $a^{FI} > a^{SB}$ and $P^{FI} > P^{SB}$. If $\frac{\partial^2 \tilde{SW}}{\partial P \partial t} < 0$ then $\tilde{SW}$ is super-modular in $(a, -P, t)$. Again, Topkis monotonicity theorem implies that $a^{FI} > a^{SB}$ and $P^{FI} < P^{SB}$. Notice that the effect over a is unambiguous as stated in the proposition. The effect over P depends on the nature of price and the anti-evasion technology as complements or substitutes. ∎

Proof (proposition 4.6):
We consider the results for the second best problem. The results are analogous for the FI problem. Consider a change from μ to μ'; $\mu' > \mu$. Consider problem (16):

$$\max_{a,P} \; SW^{SB}(a,P) := \mathbb{E}_{\gamma} \Big\{ V(P,\gamma) - (1+\mu) C(P,\gamma) \Big\} - \Psi(a) - \mu \Omega^{SB}(a,P) \tag{23}$$

Again, in order to apply the Topkis' Monotonicity result, the three conditions stated in Proposition 4.5 are required. Since this is an unconstrained problem in a rectangular euclidean domain, it suffices to see that condition (iii) is satisfied.

$$\begin{aligned}\frac{\partial^2 S\tilde{W}}{\partial a \partial \mu} &= -\frac{\partial \Omega^{SB}(a,P)}{\partial a} < 0 \\ \frac{\partial^2 S\tilde{W}}{\partial P \partial \mu} &= -\frac{\partial \Omega^{SB}(a,P)}{\partial P} = 0 \\ \frac{\partial^2 S\tilde{W}}{\partial P \partial a} &= \frac{\partial SW^{SB}(a,P)}{\partial P \partial a}\end{aligned} \tag{24}$$

where the signs of the two first cross derivatives follow from Lemma 4.4. If $\frac{\partial^2 SW^{SB}}{\partial P \partial t} > 0$ then SW^{SB} is super-modular in $(-a, P, \mu)$. Topkis monotonicity theorem implies that $a^{SB}(\mu^{'}) < a^{SB}(\mu)$ and $P^{SB}(\mu^{'}) > P^{SB}(\mu)$. If $\frac{\partial^2 SW}{\partial P \partial t} < 0$ then SW^{SB} is super-modular in $(-a, -P, \mu)$. Now, Topkis' monotonicity theorem implies that $a^{SB}(\mu^{'}) < a^{SB}(\mu)$ and $P^{SB}(\mu^{'}) < P^{SB}(\mu)$. ∎

Proof (Lemma 5.1):
By direct differentiation of equation (7) and considering the expressions for $\frac{\partial \hat{y}}{\partial \gamma}$ and $\frac{\partial \hat{y}}{\partial P}$ found in the proof of Lemma 4.1 we have:

$$\begin{aligned}\frac{\partial SW}{\partial \gamma} &= \frac{\partial S(P,\gamma)}{\partial \gamma} - \frac{\partial C(P,\gamma)}{\partial \gamma} \\ &= -\int_{\underline{y}}^{\hat{y}} h(y)\, f(y) dy - \mu\gamma \frac{\partial \hat{y}}{\partial \gamma} h(\hat{y}) f(\hat{y}) - c(x_I - x_F(P)) \frac{\partial \hat{y}}{\partial \gamma} f(\hat{y}) \\ &> -h(\hat{y}) \int_{\underline{y}}^{\hat{y}} f(y) dy + c(x_I - x_F(P)) \frac{h(\hat{y})}{h(\hat{y})\gamma} f(\hat{y}) \\ &= \frac{h(\hat{y})}{\gamma h^{'}(\hat{y})} f(\hat{y}) \Big[-(1 - M(\gamma)) \frac{\gamma}{f(\hat{y})} h^{'}(\hat{y}) + c(x_I - x_F(P)) \Big] \\ &> \frac{h(\hat{y})}{\gamma h^{'}(\hat{y})} f(\hat{y}) \Big[-\frac{\gamma h^{'}(\hat{y})}{f(\hat{y})} + c(x_I - x_F(P)) \Big]\end{aligned} \tag{25}$$

The inequalities follow the same reason as the argument in Lemma 4.2. Since the term in $\Big[...\Big]$ is strictly increasing in P the last term of block (26) is bounded below by (remember $h(y) = hy$):

$$\begin{aligned}... \quad &> \frac{-\gamma h}{f(\hat{y})} + c x_I \\ &> \frac{-h}{f(\hat{y})} + c x_I \\ &> 0\end{aligned} \tag{26}$$

Therefore we conclude that $\frac{\partial SW}{\partial \gamma} > 0$. ■

Proof (Proposition 5.4):
Consider the following problem of maximization:

$$\max_{a,P} \ \tilde{SW}(a,P,t) := \ (1-t)\cdot SW^{SB}(a,P) + t\cdot SW^{FI}(a,P) \tag{27}$$

This is the an analogous function as the one defined in the proof of Proposition 4.5. Following a similar argument, comparative statics in t require checking pairwise super-modularity. Direct differentiation leads to:

$$\begin{aligned}
\frac{\partial^2 \tilde{SW}}{\partial a \partial t} &= -\mu\left[\frac{\partial \Omega^{FI}(a,P)}{\partial a} - \frac{\partial \Omega^{SB}(a,P)}{\partial a}\right] > 0 \\
\frac{\partial^2 \tilde{SW}}{\partial P \partial t} &= -\mu\left[\frac{\partial \Omega^{FI}(a,P)}{\partial P} - \frac{\partial \Omega^{SB}(a,P)}{\partial P}\right] > 0 \\
\frac{\partial^2 \tilde{SW}}{\partial P \partial a} &= (1-t)\frac{\partial SW^{SB}(a,P)}{\partial P \partial a} + t\frac{\partial SW^{FI}(a,P)}{\partial P \partial a}
\end{aligned} \tag{28}$$

If $\frac{\partial SW}{\partial P \partial \gamma} > 0$ then $\tilde{SW}$ is supermodular in (a,P,t) which implies that $a^{SB} < a^{FI}$ and $P^{SB} < P^{FI}$. ■

Proof (proposition 5.5):
We consider the results for the second best problem. The results are analogous for the FI problem. Consider a change from μ to $\mu^{'}$; $\mu^{'} > \mu$. Consider the second-subproblem for Case II analogous to problem (16):

$$\max_{a,P} \ SW^{SB}(a,P) := \mathbb{E}_{\gamma}\Big\{V(P,\gamma) - (1+\mu)C(P,\gamma)\Big\} - \Psi(a) - \mu\Omega^{SB}(a,P) \tag{29}$$

Again, in order to apply the Topkis' Monotonicity result, the three conditions stated in Proposition 4.5 are required. Since this is an unconstrained problem in a rectangular euclidean domain, it suffices to see that condition (iii) is satisfied.

$$\begin{aligned}
\frac{\partial^2 \tilde{SW}}{\partial a \partial \mu} &= -\frac{\partial \Omega^{SB}(a,P)}{\partial a} < 0 \\
\frac{\partial^2 \tilde{SW}}{\partial P \partial \mu} &= -\frac{\partial \Omega^{SB}(a,P)}{\partial P} > 0 \\
\frac{\partial^2 \tilde{SW}}{\partial P \partial a} &= \frac{\partial SW^{SB}(a,P)}{\partial P \partial a} < 0
\end{aligned} \tag{30}$$

If $\frac{\partial^2 SW}{\partial P \partial t} < 0$ then SW^{SB} is super-modular in $(-a,P,\mu)$. Topkis' monotonicity theorem implies that $a^{SB}(\mu^{'}) < a^{SB}(\mu)$ and $P^{SB}(\mu^{'}) > P^{SB}(\mu)$. ■