INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

576

2023

# Ramsey pricing revisited: Natural monopoly regulation with evaders?

**M. Besfamille, N. Figueroa y L. Guzmán.**

www.economia.uc.cl · ISSN (edición impresa) 0716-7334 · ISSN (edición electrónica) 0717-7593

# Ramsey pricing revisited: Natural monopoly regulation with evaders*

M. Besfamille† N. Figueroa‡ L. Guzmán§

October 21, 2023

**Abstract**

We consider a model featuring a single-product natural monopoly, which faces evaders, i.e., individuals that may not pay the price. By exerting a costly effort, the firm can deter evasion. To maximize the total surplus, a regulator sets the price, the level of deterrence effort, and socially costly transfers to ensure the monopoly's participation.

We obtain a modified Ramsey formula, which clearly shows that the mere existence of evaders dampens the use of the price as a mean to finance the firm's deficit. The regulated price is always below the monopoly price and, under sufficient conditions, also below marginal cost. Then, we generalize the model to incorporate moral hazard. Finally, we undertake an empirical application of our results, which shows quantitatively that the downward tendency of regulated prices in a context of high evasion is significant.

**Keywords:** Regulation, natural monopoly, evasion and marginal cost of public funds.

**JEL Codes:** D42, H2, L43 and L51

*We thank L. Basso, M. Batarce, S.-I. Bessho, E. Bond, G. Coloma, C. Conlon, W. Cont, M. Dugrov, K. Fuyiwara, A. Galichon, A. García, D. Habermacher, J. Kline, J. Massó, A. López García, A. McLennan, C. Mezzetti, J. P. Montero, C. Moskovits, C. Müller, F. Navajas, P. O'Callaghan, H. Ogawa, A. Pereira, D. Pérez-Castrillo, A. Piolatto, J. Poblete, D. Ray, P. Rey, Y. Sato, H. Silva, R. Troncoso, S. Urbiztonodo, J.P. Xandri, and seminar and conference participants at FIEL, CIRJE (University of Tokyo), Universidad Alberto Hurtado, Universidad Autónoma de Barcelona, Universidad de los Andes, University of Queensland, IIPF (2017, 2020) and Mini TOI (2016) for useful comments and suggestions. The authors thank financial support from the Complex Engineering Systems Institute (CONICYT-PIA-FB0816) and from CAF-Banco de Desarrollo de América Latina.

†Instituto de Economía, Pontificia Universidad Católica de Chile and CESifo. Email: mbesfamille@uc.cl. Corresponding author.

‡Instituto de Economía, Pontificia Universidad Católica de Chile and ISCI. Email: nicolasf@uc.cl.

§Department of Economics, New York University. Email: lgl266@nyu.edu.

# 1 Introduction

Public services, like electricity distribution and public transportation, are crucial for the economy, not only in terms of number of users, total expenditures or resources invested, but also as essential inputs for many different activities. Indeed, Auriol et al. (2021) report that, in 2017, the average share of total household expenditures in infrastructure services was approximately 20%, both in developed and developing economies. Given their scope, and the fact that they usually operate as natural monopolies, such public services are often regulated. Although their regulation entails a myriad of dimensions, how to price them and how much to subsidize them are two of the most important issues that regulators of public services have to deal with. For example, electricity tariffs are regulated in 55% of OECD countries[1] and in 74% of the 82-non OECD countries surveyed in the Global Electricity Regulatory Index (Rana et al. (2022)).

Clearly, regulated prices are set within the same institutional context that frames the market where public services are provided. In that sense, the existence of weak institutional settings where individuals consume or use public services without paying their price (or evade) is a real-world problem that affects their regulation. Although this misbehavior is difficult to measure, many investigations demonstrate that it is not an anomaly. Smith (2004) provides estimations for electricity theft in 102 countries. More recent studies show that the percentage of electricity that is produced and not paid for, but that is not wasted due to technical transmission and distribution losses, is 22.3% in India[2] and 16.3% in Brazil.[3] This phenomenon is not limited to developing countries. According to World Bank (2014), non-technical electricity loss rates are 8.7% in Canada, 9.6% in Spain, and 12.5% in Hong Kong. Also, many bus systems across the world face fare evasion, as Table 1 illustrates.

[1]Retrieved from https://www.oecd.org/economy/reform/OECD-PMR-Sector-Database-2018.xlsx.
[2]*Report on Performance of Power Utilities 2020-2021*, Power Finance Corporation Ltd, New Delhi, India.
[3]*Perdas de Energia Elétrica na Distribuição*, Agência Nacional de Energia Elétrica, 2021, Brasilia, Brazil.

| City | Fare evasion rate in % (year) | City | Fare evasion rate in % (year) |
|---|---|---|---|
| Melbourne | 8 (2018) | Lyon | 24.8 (2017) |
| San Francisco | 9.5 (2011) | Santiago | 31.7 (2022) |
| Washington DC | 14.2 (2019) | New York | 34.1 (2022) |
| Bogotá | 15.4 (2019) | Reggio Emilia | 43 (2011) |

Table 1: Fare evasion rate in public bus systems, selected cities.
The evasion rate is the percentage of bus riders that do not pay the fare. Sources: Public Transport Victoria, Lee (2011), Washington Metropolitan Area Transit Authority, Transmilenio, Egu and Bonnel (2020), Ministerio de Transporte de Chile, New York Metropolitan Transit Authority, and Bucciol et al. (2013).

Yurtseven (2015) and Troncoso and de Grange (2017) establish that, among different factors, evasion is positively correlated with the price. Hence, regulators should recognize that the relevance of prices as feasible instruments to fund public services' providers or to mitigate their financial deficits is attenuated by the existence and extension of evasion.[4] In fact, we conjecture that evasion does more than that: It really *dampens* the use of prices to fund public services. The goal of this paper is thus to incorporate evasion into the theory of natural monopoly regulation to precisely evaluate this hypothesis.

We consider a model featuring a unit mass of risk-neutral individuals that earn identical income and consume two goods. One of these goods is provided by a monopolist (hereinafter, we refer to it as 'good $x$'); while the other is the numeraire, produced in a perfectly competitive market. Individuals have different willingness to pay for good $x$. They choose whether to consume a unit of this good formally, informally (i.e., without paying its price), or to drop out of its market. Individuals are also heterogeneous concerning the subjective cost incurred when they are detected evading, and they decide to consume informally by trading-off the monetary payment and the expected subjective cost. In our model formal buyers, evaders, and individuals that drop out of the market of good $x$ coexist. As expected, the mass of formal consumers (or equivalently, the formal demand) decreases with the price of good $x$, and increases with the detection probability.

The monopolist produces good $x$ with a commonly known decreasing average cost technology. Crucially, the variable cost depends upon total consumption, which includes that of evaders. To deal with this issue, the firm exerts a costly effort that increases the detection probability, deterring in this way informal consumption. The monopoly's earnings come from formal sales and public transfers that hinge on the realized level of

[4]Remarkably, this important topic has been absent in the economic literature of regulation. Informal consumption is not even mentioned in Sherman (1989), one of the main books about the conventional theory of monopoly's regulation, or in more recent contributions that present the advances of the incentive theory of regulation, like Laffont (1994), Armstrong and Sappington (2007), and Auriol et al. (2021).

evasion.

The regulator chooses the price and the level of effort to maximize the expected social welfare, net of transfers to the firm. In the spirit of Laffont and Tirole (1985, 1993), these transfers are costly because they are financed with distortionary taxation, and thus society bears a marginal cost of public funds (mcpf) greater than one.[5] In any case, the regulator must induce the monopoly to remain active, because there are very high losses if the latter withdraws from the market after initiating its activities.

We first characterize the optimal regulation when deterrence effort is observable and contractible. When choosing a higher price of good $x$, the regulator trades-off the extra revenues the firm receives (which are socially valuable because they save on transfers) against the rise in the mass of evaders (which aggravates the aggregate subjective costs and triggers more financial losses) and in the mass of individuals that drop out of the market of this good. In particular, we show that even when transfers are not distortionary good $x$ cannot be provided free of charge: the regulator sets a strictly positive price below the marginal cost,[6] coupled with a strictly positive level of deterrence effort. Moreover, under some conditions on the distribution of subjective costs in the population, the optimal price and effort are complements, and both increase with the mcpf.

We then rewrite the optimal pricing rule as a modified Ramsey formula, that relates the Lerner index to two terms. The first one reflects the well-established trade-off between financing the firm's deficit on formal sales with prices or transfers, while the second term is new, and it is related to the funding of the extra costs caused by evaders. These two terms have opposite signs, which implies not only that evasion pushes the regulated price downwards (in particular, it is always below the monopoly price), but also that there is a priori no clear prediction regarding the comparison between the optimal price and the marginal cost. Actually, we find a simple and reasonable sufficient condition ensuring that, under the threat of evasion, the regulated price is always lower than the marginal cost, **for any value of the mcpf**. To the very best of our knowledge, this is the first contribution to the public economics literature that presents such a result, which confirms our initial conjecture that evasion can present a big challenge when regulating prices in public services.

[5]The mcpf is the money measure of the welfare cost of raising an additional dollar of tax revenues. See Dahlby (2008) for theoretical considerations and applications of this concept.

[6]We are not the first to obtain such result. For example, in the context of a budget-constrained natural monopoly financed by a two-part tariff, Ng and Weisser (1974) show that if the demand of the marginal individual (i.e., the one that is indifferent between consuming or not the good produced by the firm) is above the average demand of inframarginal consumers, the regulator optimally sets the variable price below the marginal cost. In Srinagesh (1991) and Armstrong and Vickers (1993), pricing below marginal cost can emerge when regulation allows non-linear pricing with bundling and price discrimination, respectively.

Next, we assume that deterrence effort is neither observable nor contractible, and we modify the detection technology to incorporate the moral hazard dimension into the model. In this case, the firm has to be induced to exert any level of deterrence effort. Thus, the regulator has to take into account the social cost of the informational rent left to the firm when the extent of evasion is low. As expected, the optimal level of deterrence effort is lower than under full information, pushing down the likelihood of the low evasion case. To attenuate this impact, the regulator optimally decreases the price below its value under full information. Surprisingly, the pricing rule is not distorted by the presence of moral hazard, which is reminiscent to the 'dichotomy property' exposed by Laffont and Tirole (1990). By comparing our findings to theirs, we explain why in the context of our model this result is intuitive, although unexpected.

Finally, to assess quantitatively the downward tendency of regulated prices when evasion is high, we present an empirical application of the optimal pricing rule, based on a sufficient statics approach. Using elasticities of evasion and total demand of the bus component of Transantiago (the public transportation system in the capital of Chile) that were obtained from previous contributions, coupled with an estimation of the mcpf for this country, we compute the optimal fare for the period 2013-2015. Our result is 49.4% and 14.08% lower than the (average) observed fare and the marginal cost, respectively. These are significant differences that, to some extent, seem to validate our theoretical conclusions.

## 1.1 Related Literature

Our paper connects with different strands of the literature. Clearly, it stands on the analysis of pricing by the conventional theory of natural monopoly regulation, which has been undertaken under two different theoretical approaches. In the Ramsey-Boiteux tradition under full information (see Ramsey 1927, Boiteux 1956, Baumol and Bradford 1970), the monopoly must break-even because an exogenous constraint prevents the regulator to subsidize it. Following the concerns raised by Meade (1944) regarding marginal-cost pricing and using the formalism presented in Caillaud et al. (1988), Laffont and Tirole (1985, 1993) analyze optimal regulation under asymmetric information, when transfers to the firm are authorized but are socially costly because they are funded with distortionary taxation. In any case, the optimal regulation is a second-best one: by setting properly the price above the marginal cost, the firm's deficit is financed inducing a minimal loss in consumer surplus. Moreover, when the firm can be subsidized, the higher the mcpf, the higher the optimal price should be. In particular, when the mcpf tends to infinity, the regulated price should converge to the monopoly price. Undoubtedly, incorporating evasion into Laffont and Tirole's setting, as we do, qualifies some of their conclusions.

Recently, this theory has been extended to incorporate regulatory problems in developing countries. As Estache and Wren-Lewis (2009) claim, in those countries "*There is a clear concern that public institutions are unable to collect adequate revenue to allow direct subsidies when the ability of consumers to pay for services is limited.*" Our paper contributes to this body of work in at least two ways. We explicitly take into account that weak institutional settings allow consumers to choose whether to pay or not for the service. Then, by highlighting fare evasion as an important matter for regulators, we complement the analysis made by Laffont (2005) on the optimal way to solve the tension mentioned in the previous citation. Moreover, we bring this topic to the forefront of the regulatory discussion, since this tension is becoming relevant also in more advanced economies, as the data presented above suggests.

Silva and Kahn (1993) analyze a firm that provides an excludable public good, funded through fees from formal consumers. As these authors assume self-financing as an institutional constraint, the fee is endogenously chosen by dividing costs among paying individuals. Free-riders increase costs, and therefore it is optimal to design the system so that they are kept away. Buehler et al. (2017) present a model where a profit-maximizing firm can obtain revenues from prices and fines on detected evaders. The firm engages in a form of price discrimination, charging lower 'expected prices' (through expected fines) to consumers with a low willingness to pay, who choose to evade. When these authors consider a welfare maximizing firm, price discrimination vanishes, and thus marginal cost pricing emerges at the optimum. Although in this last case financial losses occur, Buehler et al. (2017) do not formalize the funding of such deficits. None of these two contributions examine the tradeoff between revenue financing (through prices or fines) and direct subsidies which are costly for a regulator, which explains why we find contrasting results.

Recently, Ramos and Silva (2023) study theoretically and numerically optimal public transportation policies when evasion is pervasive. Despite the fact that these authors also obtain a downward tendency of fares under these circumstances, some of their results differ from ours. Unlike them, we do provide a modified Ramsey formula which clearly exhibits the role of the mcpf and the evasion concern. Moreover, the simplicity of this formula enables us to show that, under some plausible conditions, the optimal fare is always lower than the marginal cost, even for large values of the mcpf.

Last, our model shares many features with contributions of the tax compliance literature (see Slemrod (2019) for a recent survey), in particular the formalization of evaders as optimizing individuals and the use of a deterrence technology to modify their behavior. As in Chander and Wilde (1998), we also show how the existence of evaders distort the optimal value of some policy instruments. Notwithstanding these similarities, the focus of our paper is very different: the financing of the firm's deficit through transfers, when the use of the price is severely constrained by evaders.

The layout of the remainder of the paper is as follows. Section 2 describes the model, and discusses some assumptions. Section 3 analyzes optimal regulation when deterrence effort is observable. Section 4 incorporates moral hazard. Section 5 presents the empirical application. Section 6 concludes. All proofs are relegated to the Appendix.

# 2 The Model

We consider a two-goods economy with a unit mass of individuals. The good $x$ is provided by a monopolist, while the numeraire $m$ is produced in a perfectly competitive market.

**Individuals** Individuals earn income $y$, and have unit demand for good $x$. They have quasilinear utilities on good $m$, and have different willingness to pay for good $x$, represented by $\theta$.

Good $x$ can be consumed in two different ways: formal and informal. Formal consumption is done at price $p$,[7] while informal consumption *does not*. Informal consumption leads to an expected loss $ez$, where $e$ is the probability of being detected evading the payment of $p$, and $z$ is the money value of the subjective cost faced in this circumstance. This subjective cost, which is heterogeneous across individuals, may be caused by a loss of reputation, waste of time or social stigma.

Formally, $z$ and $\theta$ are distributed according to strictly positive, continuous, independent and bounded density functions $f$ on $[0, \overline{z}]$ and $g$ on $[0, \overline{\theta}]$, with cumulative distributions $F$ and $G$, respectively. For technical reasons that will become clear later on (see Lemma 3), we adopt the following assumption.

**Assumption 1** The density $f$ is log-concave, with $\overline{z}\frac{f'(\overline{z})}{f(\overline{z})} > -1$.

[7] We assume that income $y$ is sufficiently high, i.e. $y > \overline{z}$, so that all individuals can afford to buy good $x$.

An individual might also decide not to consume good $x$, and instead spend his entire income on $m$: We say that such an individual 'drops out of the market of good $x$'. We will hereinafter characterize individuals based on whether and how they decide to consume good $x$.

Expected utilities of formal and informal consumers are denoted by $U^F$ and $U^I$, respectively. Similarly, let $U^O$ be the expected utility of an individual that drops out of the market of good $x$. Assuming that everybody observes the detection probability before consuming, these expected utilities are given by

$$U^F \equiv \theta + y - p, \quad U^I \equiv \theta + y - ez, \quad U^O \equiv y.^{8} \tag{1}$$

We impose that, when they are indifferent between consuming or not good $x$, individuals choose to consume it. Similarly, when they are indifferent between consuming good $x$ formally or informally, individuals choose to purchase it. We denote the individuals' indirect utility function by $\mathcal{U}(p, e, z, \theta)$.

Comparing expressions (1), three relevant thresholds appear. The value of the willingness to pay that makes an individual being indifferent between consuming good $x$ formally and dropping out of its market is $\widehat{\theta}(p) \equiv p$. The value of the reputation cost that makes an individual being indifferent between consuming good $x$ formally and informally is $\widehat{z}(p,e) \equiv \frac{p}{e}$. Finally, $\varphi(e,z) \equiv ez$ is the value of the willingness to pay that makes an individual being indifferent between consuming informally good $x$ and dropping out of its market.

Figure 1 depicts these thresholds, and illustrates, in the $[0,\overline{z}] \times [0,\overline{\theta}]$ plane, the configuration of decisions adopted by individuals, regarding the way they consume good $x$.[9] The figure shows the most general case, that is a configuration in which formal consumers, evaders, and individuals that drop out of the market of good $x$ coexist.[10] Consumers with a relatively high willingness to pay, $\theta > \widehat{\theta}(p)$, become informal if their subjective cost $z$ is small enough, $z \leq \widehat{z}(p,e)$. On the other hand, consumers with relatively low willingness to pay, $\theta \leq \widehat{\theta}(p)$, would never pay for the good and become informal only if their subjective cost is small enough. In fact, the higher the willingness to pay, the higher the reputation cost they are willing to bear when evading.

[8]This assumption regarding $U^O$ is a shortcut to a more sophisticated model, where individuals that drop out of the market of good $x$ would have different outside options because they can consume other goods. For the sake of simplicity, we normalize to 0 the utility obtained, under this circumstance, by any of those individuals.

[9]Throughout the paper, we assume that $\overline{\theta} > \overline{z}$. This assumption simplifies considerably the analysis.

[10]This case emerges provided $p < e\overline{z}$. Otherwise, there would be no formal consumers. Without loss of optimality, we do not consider this other scenario. See Section 3 and Appendix B.

*Ceteris paribus*, when $p$ increases, the kink $(\widehat{z}(p,e),\widehat{\theta}(p))$ moves along the frontier $\varphi(e,z)$ upwards, implying that less individuals consume formally. Hence, despite the fact that each individual may consume only one unit of good $x$, formal demand and aggregate consumption are not constant. On the other hand, when the detection probability $e$ increases *ceteris paribus*, the slope of the frontier $\varphi(e,z)$ increases, and thus $\widehat{z}(p,e)$ moves to the left. This entails that less individuals evade.

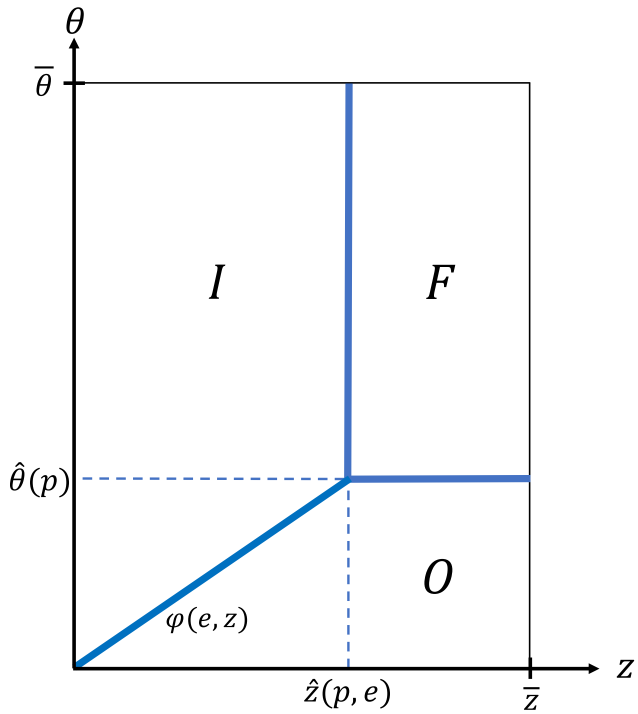



Figure 1: Individuals' decisions

$I$ and $F$ represent informal and formal consumption of good $x$, respectively. $O$ stands for dropping out of its market.

**The firm** The monopolist produces good $x$ using a decreasing average cost technology. Specifically, the production cost is given by

$$C(X) = K + c \cdot \left(X^F + X^I\right), \tag{2}$$

where $K$ is a fixed cost, $c$ is the (constant) marginal production cost, and $X^F$, $X^I$ denote formal and informal aggregate consumption of good $x$, respectively. Therefore, the production cost depends upon total demand $X = X^F + X^I$, which includes consumption by evaders. We assume that fixed and marginal costs are common knowledge.

For institutional reasons, the firm is unable to levy fines on evaders. However, it can deter informal consumption through a detection technology. By exerting a costly effort, the firm affects the probability to catch evaders. Until Section 4, we assume that such probability $e$ is equal to the effort level.

From (1) and Figure 1, we can see that the levels of $p$ and $e$ have an impact on the individuals' decisions regarding consumption of good $x$, and thus on the mass of formal and informal consumers. We then write $X^F(p,e)$, $X^I(p,e)$, and $X(p,e)$.

The firm's earnings come from formal consumers and transfers from the regulator. In fact, its ex-post utility is given by

$$V(p,e,T) \equiv pX^F(p,e) + T - c \cdot \left[X^F(p,e) + X^I(p,e)\right] - \psi(e), \tag{3}$$

where $T$ stands for transfers received from the regulator, and $\psi(e)$ is the effort cost. The function $\psi$ is strictly increasing and convex, and satisfies $\psi(0) = 0, \lim_{e\to 0} \psi'(e) = 0$ and $\lim_{e\to 1} \psi(e) = \infty$. We assume that the firm can leave the market at any time, and we normalize its outside option to zero.

**The regulator** The regulator chooses the price $p$ and non-negative transfers $T$. As we assume that effort is observable and contractible, the regulator can also choose the level of effort $e$ to be exerted by the firm. We follow Laffont and Tirole (1985, 1993) and assume that the regulator maximizes the social welfare

$$W(p,e,T) \equiv CS(p,e) + V(p,e,T) - (1+\lambda)T, \tag{4}$$

where

$$CS(p,e) \equiv \int_0^{\overline{z}} \int_0^{\overline{\theta}} \mathcal{U}(p,e,z,\theta)dG(\theta)dF(z),$$

stands for the net consumer surplus. In order to raise $T$, the government taxes other (non-modelled) sectors of the economy in a distortionary way. In this sense, $\lambda > 0$ represents the deadweight loss of taxation, and $1+\lambda$, the marginal cost of public funds.

Instead of expressing the social welfare in terms of transfers, it is more convenient to do it considering the firm's utility. In this way, the trade-off faced by the regulator becomes evident. Indeed, solving for $T$ in (3), and replacing it in (4), we can rewrite $W(p,e,T)$ as a function of $V$ instead of $T$, and therefore we have

$$W(p,e,V) = S(p,e) + \lambda pX^F(p,e) - (1+\lambda)[C(X(p,e)) + \psi(e)] - \lambda V, \tag{5}$$

where $S(p,e)$ is the gross consumer surplus. As in Laffont and Tirole (1985, 1993), the regulator values positively income from formal sales (because they decrease the need to make transfers), but dislikes leaving rents to the firm (because they are socially costly). But here, due to the fact that evaders' social weight is the same than for the other individuals, informal consumption imposes two different costs to society. On the one hand, by not paying the price, the social benefit of the firm's income is lower than in the conventional model of natural monopoly regulation. On the top of that, informal consumers suffer the subjective cost $z$ when caught evading.

We assume that society bears very high costs if the firm withdraws after being allowed to operate.[11] Therefore, in order to induce it to stay active **under any circumstance**, the regulator must satisfy the firm's ex-post voluntary participation constraints $V \geq 0$.

**The timing** At $t = 1$, the regulator sets the price $p$, contingent transfers $T$ and requires a level of effort $e$. At $t = 2$, individuals learn the probability $e$ and decide either to consume or not good $x$, and if they consume, whether to do it formally or informally.[12] Finally, production takes places as to meet total consumption, and all payoffs are realized. Figure 2 summarizes the timing.

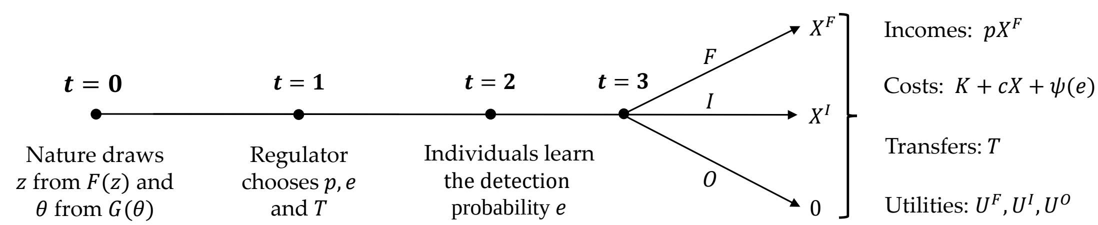



Figure 2: Timing of the model

## 2.1 Discussion

Some features of the model deserve some comments. First, we assume that the monopoly cannot impose fines to individuals that are caught evading. Indeed, in some

[11]For example, if good $x$ represents electricity distribution, many long term decisions, like urban design or the localization of firms, depend upon its normal operation. If this public service collapses, the regulator may face social discontent, bankruptcies, etc.

[12]The fact that we assume that individuals observe the detection probability is consistent with Bucciol et al. (2013), who found that passengers' beliefs on ticket inspection frequency is very close to the actual figure.

countries, firms that provide public services are either not allowed to fine evaders directly or they have to rely upon a costly and uncertain judicial procedure to do so.[13] More importantly, fines are seldom an important source of income, either for providers of public services or for the government, and even less if judicial costs are considered.[14] Therefore, as our focus is the use of the price as a source of funding for the firm in a context of evasion, not incorporating fines does not seem to be too restrictive.

Individuals face an idiosyncratic cost $z$ if they are caught evading. As we have already mentioned, this cost represents the social stigma and reputation loss due to being detected,[15] and the waste of time and disutility from subsequent prosecution. Like fines, all these inconveniences act as deterrents for evasion. It has been documented that moral costs or social sanctions do matter for individuals that evade the payment for public services (see, among others, Sterner and Sheng 2013, Dai et al. 2019 and Ayal et al. 2021). Moreover, with the goal to affect the reputation of a particular individual, enforcement authorities have sometimes exposed evaders in the social networks.[16]

Our setting differs with respect to the canonical model that deals with the regulation of a multiproduct natural monopoly. With heterogeneous individuals, the regulator chooses both $p$ and $e$, introducing, for evaders, a 'generalized cost' $p + ez$. Individuals differ in their subjective cost $z$, and therefore they self-select among the different options (formal and informal consumption, dropping out of the market). However, this is not equivalent to the standard second-degree price discrimination problem of a regulated natural monopoly (as in Laffont and Tirole (1990)), since the regulator is not allowed to offer 'menus' of $(p, e)$, because effort cannot be tailored to different consumers.

In our model, the regulator fully credits the evaders' utility to welfare. Although this is a conventional assumption in the deterrence literature, one might be concerned

[13]For example, in Chile, to fine evaders, inspectors from private providers of public transport need to be accompanied by the police. And even if caught, evaders might not be penalized after all. Calvo (2015) shows that, in all *comunas* of Santiago, the percentage of detected evaders that were finally fined was only less than 9 percent. This incapacity to fine fare evaders is not a distinctive feature of developing countries, as the study of Bijleveld (2007) in the Netherlands revealed.

[14]From 2019 to 2021, Hydro-Québec raised additional CA$ 16.5 million from theft electricity, 0.04% of its revenues. In 2022, the New York Metropolitan Transportation Authority collected US$ 2951 millions in bus and subway fares, and only US$ 11.38 millions in fines. Hence, fines represent merely 0.39% of its operating revenues.

[15]In some cases, individuals even feel harassed by inspectors. See https://www.theage.com.au/national/victoria/fare-evasion-on-melbourne-public-transport-at-lowest-recorded-level-20150102-12gty3.html

[16]In Argentina, the electricity company *Edenor* published in January 2019 on *Twitter* that one of their inspections found an illegal electricity installation in a franchise of Maru Botana, a well known baker and TV presenter. After a couple of months, Maru Botana decides to settle out of court and pay the corresponding fine. See https://www.clarin.com/sociedad/edenor-escracho-maru-botana-colgarse-luz-locales_0_thDI9lXJ_.html.

by the fact that, for different reasons, real life regulators may distinguish between formal consumers and evaders. We explore this possibility in the Conclusions.

## 3 Optimal regulation

In this section, we characterize the optimal regulatory scheme, assuming that effort $e$ is observable and contractible, and thus the regulator can impose its level to the firm. The regulator chooses the price $p$, the effort $e$, and the firm's rents $V$ to maximize the social welfare, while inducing the firm's ex-post voluntary participation. Formally, the problem faced by the regulator is

$$\max_{p,e,V} W(p,e,V) \;\text{ s.t. }\; V \geq 0. \quad (VP) \tag{6}$$

Let's denote by $p^E$, $e^E$ and $V^E$ the solutions to (6). Since transfers, and therefore rents to the firm, are costly, the regulator designs a scheme which grants no utility to the monopoly.

**Lemma 1** *Optimal regulation requires $V^E = 0$.*

This lemma enables us to rewrite (6) as[17]

$$\max_{p,e} W(p,e) = \underbrace{y + \int_0^{\widehat{z}} \int_{\varphi}^{\overline{\theta}} \theta \, dG(\theta) dF(z) + \int_{\widehat{z}}^{\overline{z}} \int_{\widehat{\theta}}^{\overline{\theta}} \theta \, dG(\theta) dF(z)}_{\text{(a)}}$$
$$\underbrace{- \int_0^{\widehat{z}} \int_{\varphi}^{\overline{\theta}} ez \, dG(\theta) dF(z)}_{\text{(b)}} + \underbrace{\lambda p X^F(p,e)}_{\text{(c)}} - \underbrace{(1+\lambda)\big(C(X(p,e)) + \psi(e)\big)}_{\text{(d)}}. \tag{7}$$

Block (a) captures aggregate individual benefits from consumption of the numeraire $m$ and the good $x$. Block (b) corresponds to the expected aggregate welfare loss from subjective costs borne by evaders. Block (c) represents the expected social value of revenues from formal sales, stemming from a reduction of costly transfers, and thus rents to the firm. Finally, block (d) expresses the social value of production and effort costs. The next lemma characterizes the solution to (7).

[17]From now on, we omit the dependence of the thresholds $\widehat{z}$, $\widehat{\theta}$, and $\varphi$ on the other variables or parameters of the model to simplify notation.

**Lemma 2** *The optimal price $p^E$ and effort $e^E$ are implicitly given by the first-order conditions*

$$(1+\lambda)g(\widehat{\theta})\frac{\partial\widehat{\theta}}{\partial p}\big[1-F(\widehat{z})\big](p^E-c)+(1+\lambda)f(\widehat{z})\frac{\partial\widehat{z}}{\partial p}\big[1-G(\widehat{\theta})\big]p^E=\lambda X^F, \tag{8}$$

$$-(1+\lambda)f(\widehat{z})\frac{\partial\widehat{z}}{\partial e}\big[1-G(\widehat{\theta})\big]p^E+(1+\lambda)c\int_0^{\widehat{z}}g(\varphi)\frac{\partial\varphi}{\partial e}f(z)dz \\ -\int_0^{\widehat{z}}z\big[1-G(\varphi)\big]f(z)dz=(1+\lambda)\psi'(e^E). \tag{9}$$

Equation (8) highlights the marginal benefits and costs of an increase in $p$. The right-hand side corresponds to the marginal benefit, which is the extra revenue, valued at $\lambda$, coming from inframarginal formal consumers. The two terms in the left hand side correspond to the marginal costs. The first one is standard, as it measures the social loss from consumers that, after an increase in the price, leave the market of good $x$. The second term is new. There is a mass $f(\widehat{z})\frac{\partial\widehat{z}}{\partial p}\big[1-G(\widehat{\theta})\big]$ of new evaders. Each one saves the payment of the price $p^E$, while also bears an expected subjective cost of $e^E z$. But, as these individuals are *marginal evaders* (who are indifferent between paying to consume formally or not), these benefits and costs cancel out. However, each new evader causes a direct financial loss to the firm, loss that amounts to the forgone payment and has a social cost of $(1+\lambda)p^E$ (because it must be covered by transfers). It is important to note that the social loss generated by a formal consumer that leaves the market of good $x$ is lower than the corresponding one that arises with an evader, since the latter stops paying but keeps consuming the good.

To determine the optimal level of deterrent effort, in (9) the regulator balances the marginal increase in additional revenues plus cost savings, both accruing from less evasion, net of the change in subjective costs faced by evaders, and its socially marginal cost.[18]

In Appendix B, we show that the regulator always sets a price-effort scheme that induces the configuration with formal consumers. Otherwise, she would face only evaders and individuals that do not consume good $x$, with the consequent total loss of sales income.

Next, we characterize the optimal price-effort scheme when transfers are not socially

[18]Throughout the paper, we impose parameter conditions to ensure that the second-order conditions hold at the optimum, and thus that interior solutions are unique.

costly.

**Proposition 1** *When $\lambda = 0$, the optimal price-effort scheme satisfies $0 < p^E < c$ and $e^E > 0$.*

In a setting with no evaders, marginal cost pricing emerges as the optimal solution when $\lambda = 0$. But when individuals can consume good $x$ informally, the regulator optimally decreases the price below marginal cost. Although some individuals with a willingness to pay lower than the marginal cost end-up buying, which is inefficient, this solution attenuates the incentives to misbehave of some potential evaders. However, even when $\lambda = 0$, good $x$ cannot be provided free of charge ($p^E > 0$), and thus deterrence effort is also needed ($e^E > 0$). To grasp the intuition of this last result, consider an initial situation with $p = e = 0$. Although in this case the equilibrium is indeterminate (i.e., all individuals are indifferent between consuming formally and informally), expected welfare is uniquely defined. Then, we simultaneously increase $p$ and $e$. These changes marginally decrease the mass of evaders and formal consumers, which triggers a reduction in production costs, but also increase the aggregate subjective costs borne by inframarginal evaders. In fact, we can always find a way to make these increases in $p$ and $e$ such that the efficiency gains are much larger than the rise in subjective costs, because the mass of inframarginal evaders can be made arbitrarily small.

The following lemma shows the most important comparative statics of the optimal price-effort scheme.

**Lemma 3** *The optimal price $p^E$ and effort $e^E$ increase with the deadweight loss $\lambda$.*

In this model, raising deterrence effort has two opposing effects on the expected aggregate subjective costs borne by evaders: It increases their expected level for each evader (the inframarginal effect) but also reduces the number of informal consumers (the marginal effect). Assumption 1 ensures that, at the optimum, the marginal effect dominates,[19] and thus that price and effort are *complements*. When $e^E$ goes up, less individuals evade, and therefore the marginal benefit of increasing the price as a mean to finance the firm increases.

The comparative static with respect to $\lambda$ results from this complementarity. As $\lambda$ goes up, the social cost of transfers increases, making more attractive the use of higher prices to finance the firm. But this triggers evasion, which calls for higher effort, to deter

[19] In other words, even if the regulator is utilitarian and puts equal social weights on formal and informal consumers, she prefers to face less evaders.

it. The aforementioned complementarity between price and effort reinforces these effects, pushing both regulatory instruments further upwards. This strengthens our last comment to Proposition 1, in the sense that both the price and the deterrent effort are necessary regulatory tools when some individuals consume informally.

### 3.1 A modified Ramsey pricing rule

How would a regulator address practically the problem analyzed in this paper, using the tools provided by the received theory of natural monopoly regulation? It would be a mistake to simply incorporate informal consumption in the definition of the price elasticity of demand, and then to set the price according to the conventional Ramsey formula (see Laffont and Tirole 1993)

$$\frac{p^E - c}{p^E} = -\frac{\lambda}{1+\lambda}\frac{1}{\eta^D}, \tag{10}$$

with $\eta^D \equiv \frac{\partial(X^F+X^I)}{\partial p}\frac{p}{X^F+X^I}$. In fact, individuals who start evading after a price increase are very different than those who drop out of the market, since the former still consume and thus generate costs for the firm. The following proposition shows that (10) is not the proper way to deal with evasion, by deriving the correct pricing rule.

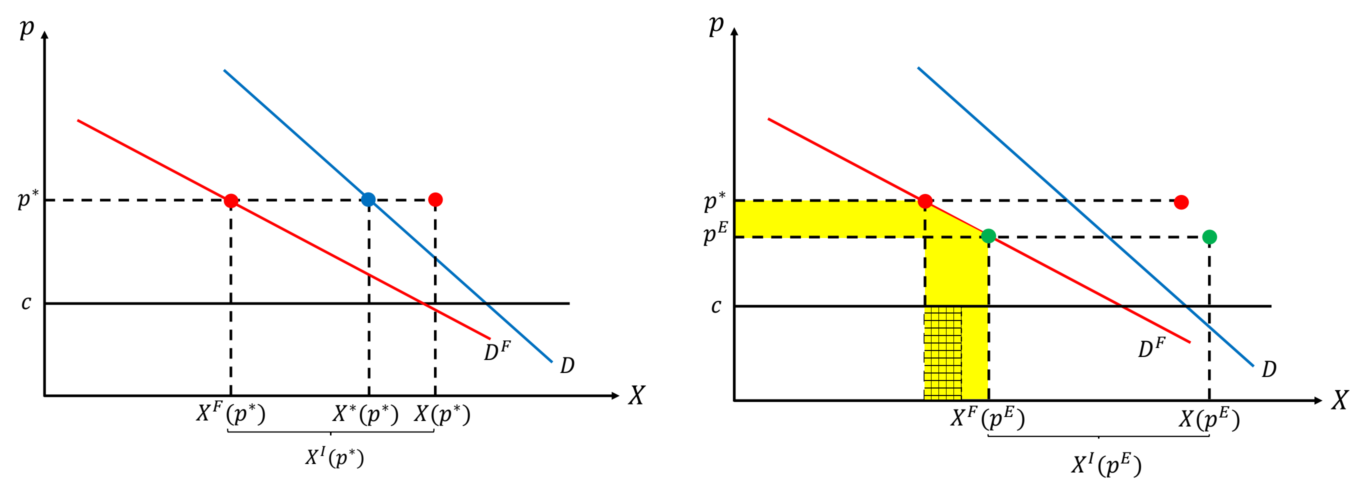
**Proposition 2** *Let $\eta^F \equiv \frac{p}{X^F}\frac{\partial X^F}{\partial p} < 0$ and $\eta^I \equiv \frac{p}{X^I}\frac{\partial X^I}{\partial p} > 0$ be the price elasticity of formal demand and informal consumption, respectively. The optimal pricing rule is*

$$\frac{p^E - c}{p^E} = \underbrace{-\frac{\lambda}{1+\lambda}\frac{1}{\eta^F}}_{R\ concern} + \underbrace{\frac{c}{p^E}\frac{X^I(p^E,e^E)}{X^F(p^E,e^E)}\frac{\eta^I}{\eta^F}}_{Evasion\ concern}. \tag{11}$$

The proposition presents a modified version of the Ramsey pricing rule, as stated in Laffont and Tirole (1993). On the left-hand side of (11), the Lerner index evaluates the markup on *sold quantities*. The right-hand side is the sum of two terms. The first one captures the Laffont and Tirole's version of the well-known Ramsey concern for financing a natural monopoly's deficit on *formal* consumption with prices and costly transfers, where the marginal cost of public funds and the price-elasticity of formal demand play a crucial role. It is intuitive to understand why the latter does not consider evaders. As marginal evaders do not stop consuming good $x$ after its price increases, their utility remains unchanged, which implies that their shift has no impact on the expected social welfare. But the presence of informal consumers clearly pushes the (absolute) value of the price elasticity of formal demand upwards, which *ceteris paribus* implies a lower mark-up, and thus a regulated price lower than the optimal Ramsey price with transfers $p^*$, obtained in Laffont and Tirole (1993).

The second term captures another effect faced by the regulator, which is absent in models without evasion: the mass of informal consumers increases, pushing the variable cost of serving them upwards. This second effect is measured with respect to the loss of income generated by the price increase, and thus is negative, reinforcing the above mentioned decrease in the optimal markup. Surprisingly, this financial concern for the extra deficit caused by evaders does not depend explicitly upon the marginal cost of public funds. The reason for that is the following. By definition, the extra deficit on informal consumption can only be financed with transfers. Thus, the regulator faces no trade-off between instruments with different social cost, as was the case for the previous concern.

Figure 3 illustrates these considerations, for a given level of effort $e$. In Panel (a), we




(a) Ramsey pricing with evasion

(b) Optimal pricing with evasion

Figure 3: Optimal pricing

draw the demand for good $x$ when there is no evasion, $D$ (in blue), the marginal cost $c$, the optimal Ramsey price $p^*$, and total demand $X^*(p^*)$. Then, we let evasion to emerge; but we keep the price constant. The red curve depicts the demand of formal buyers, $D^F$, and thus $X^F(p^*)$ is the total quantity purchased. Finally, we also show total informal consumption $X^I(p^*)$, and total consumption $X(p^*)$. In Panel (b), we let the regulator to react optimally to the presence of evaders, and we depict the optimal price and quantities (in green). When deciding whether and how much to modify the price, the regulator has to take into account i) the concern about financing the firm's deficit on formal sales with the price and costly transfers, and ii) the financial consideration of the extra deficit caused by evaders' consumption. As we have already mentioned, $p^E \leq p^*$. So, when it decreases the price optimally, the regulator balances the social values of the net change in the firm's income and the increasing consumer surplus (in yellow) and the decrease in the cost of serving evaders (striped area).

Now, we use the modified Ramsey formula to derive some unexpected results and relevant policy implications.

The following corollary nests our model with the Ramsey-Laffont-Tirole setting.

**Corollary 1** *If all individuals bear an infinite subjective cost if they are caught evading,* $p^E \to p^*$ *and* $e^E \to 0$.

If all subjective costs are sufficiently high, our model behaves as if there were no evaders, and the regulator would only face formal consumers and those that drop out of the market of good $x$. As expected, the Ramsey pricing rule with transfers is optimal, and there is no need to exert deterrence effort.

The next corollary compares the optimal price and the price set by an unregulated monopoly in a formal market.

**Corollary 2** *Let* $p_M$ *be the price set by an unregulated monopoly that only serves formal buyers. For all values of the deadweight loss* $\lambda$, $p^E < p_M$.

When the marginal cost of public funds is very large, and thus welfare considerations vanish, the regulated price $p^E$ equals $p^E_M$, which is the price set by an unregulated monopoly that face evaders. In Appendix G, we show that $p^E_M$ is strictly lower than $p_M$, the price set by an unregulated monopoly in a formal market. Intuitively, if the regulator decreases slightly the price from its monopoly level, by an envelope argument the firm's forgone revenues from inframarginal formal buyers are equal to the new income from individuals that were out of the market. But there is a first-order gain: the income from former evaders. This corollary confirms that the downward tendency of prices when evasion is a concern, result conveyed by the modified Ramsey formula (11), remains valid when $\lambda \to \infty$. Then, as $p^E$ increases with $\lambda$, the inequality always holds.

## 3.2 Relationship between the optimal price and the marginal cost

The previous result relates the optimal price with the monopoly price, showing that the former is always lower. However, how far can the downward tendency of the regulated price go? To answer this question, we focus on another issue that previous contributions have extensively analyzed: the comparison between the optimal price $p^E$ and the marginal cost $c$.

Observe that at the right-hand side of the modified Ramsey formula (11), the first term is positive, while the second is negative. Therefore, contrary to the conventional theory of natural monopoly regulation, *a priori* there is no clear-cut prediction concerning the comparison between $p^E$ and $c$. The following proposition presents some conditions that clarify the comparison mentioned above.

**Proposition 3** *Let $\widetilde{c}$ be implicitly defined by $1 - F(\widetilde{c}) = \widetilde{c} f(\widetilde{c})$. If $c > \widetilde{c}$, for all values of the deadweight loss $\lambda$, the optimal price $p^E$ is lower than the marginal cost $c$. Otherwise, there exists a threshold $\widetilde{\lambda}$ above which $p^E > c$.*

It is quite natural that, for low $\lambda$, the price $p^E$ should be lower than the marginal cost $c$. In fact, when $\lambda = 0$, $p^E = c$ if there are no evaders, but their presence gives a rationale to keep prices below that level (see Proposition 1), while $e^E$ is also strictly positive. By a continuity argument, pricing below marginal cost and a positive effort hold for low values of $\lambda$.

But Proposition 3 goes further: under a mild sufficient condition, the optimal regulatory response to the threat of evasion is to price below marginal cost, for **any value of the deadweight loss** $\lambda$. Therefore, the inefficiency reported in Proposition 1 may persist, even when the size of the fiscal system's distortion is large.

To explain intuitively where does this sufficient condition come from, assume that $\lambda$ is very high, and thus the regulator is mainly concerned with avoiding transfers to the firm. Starting from marginal-cost pricing, a price decrease of $\delta$ has three effects (as in (8)). First, some individuals that were out of the market become formal, but as $p = c$, they do not contribute to the firm's profit. The second effect corresponds to revenue losses from charging a lower price to inframarginal formal consumers, which is equal to

$$\left[1 - F(\widehat{z}(c,e))\right]\left[1 - G(\widehat{\theta}(c))\right] \times \delta.$$

The third effect derives from extra revenue caused by a drop in the number of evaders. The magnitude of this effect depends on the firm's deterrence effort. In the spirit of finding a sufficient condition, we consider the most restrictive case, i.e., $e^E = 1$. in that case, these revenue gains are given by

$$\left[F(\widehat{z}(c,1)) - F(\widehat{z}(c-\delta,1))\right]\left[1 - G(\widehat{\theta}(c))\right] \times c.$$

At this point, the sufficient condition becomes transparent. The regulator chooses to decrease the price whenever

$$\left[F(\widehat{z}(c,1)) - F(\widehat{z}(c-\delta,1))\right]\left[1 - G(\widehat{\theta}(c))\right] \times c \geq \left[1 - F(\widehat{z}(c,1))\right]\left[1 - G(\widehat{\theta}(c))\right] \times \delta,$$

or $\left[F(c) - F(c-\delta)\right] \times c \geq \left[1 - F(c)\right] \times \delta$.[20] As $\delta \to 0$, the previous inequality boils down to $cf(c) \geq \left[1 - F(c)\right]$.

The proposition merely finds conditions over the marginal cost $c$ and the distribution of reputation costs $z$ as to guarantee that pricing below marginal cost holds even for very large values of $\lambda$. Observe that these conditions do not involve the distribution of the willingness to pay $\theta$. This reaffirms the assertion that, in this model, pricing below marginal cost is only a concern related with evasion.

One could suspect there must be something wrong with this result, in the following sense. If $\lambda \to \infty$, the regulator should give up using transfers. But, if she proceeds in this way, where does the money to fund the firm's deficit come from? As an answer, observe first that even if $\lambda \to \infty$ transfers do not tend to 0. Second, it is precisely in the context of this discussion where the implicit assumption that the firm cannot be shutdown becomes relevant, and relaxing it only enhances our result. Indeed, provided $c > \widetilde{c}$, $p^E < c$ until, for very large values of the mcpf, the firm is shut down because the market of good $x$ is non viable.

The following proposition shows how a particular comparative statics on the cumulative distribution function $F(z)$ affects the optimality of pricing below marginal cost.

**Proposition 4** *A shift in the cumulative distribution function of subjective costs, from $F_1(z)$ to $F_2(z)$, with $F_1(z)$ dominating $F_2(z)$ in the hazard rate order, increases the range of values of the marginal cost such that $p^E < c$ for all values of $\lambda$.*

If the cumulative distribution function $F_1(z)$ dominates $F_2(z)$ in the hazard rate order, it does also in the first stochastic order. So, a shift from $F_1(z)$ to $F_2(z)$ lowers subjective costs. Hence, *ceteris paribus*, evasion is more important at the margin, and thus the concern

[20] As we $e = 1$, $\widehat{z}(c,1) = c$ and $\widehat{z}(c+\delta,1) = c+\delta$.

for evasion in the modified Ramsey pricing rule (11) increases. This implies that there are more instances where the regulator optimally sets the price below marginal cost.

Finally, by explicitly focusing on the funding of the firm's deficit without fines, our results contrast those found by Buehler et al. (2017). In particular, expression (11) yields marginal cost pricing provided very particular conditions hold, whereas these authors always obtain this clear-cut result when the monopoly maximizes the social surplus. Moreover, in their setting, prices are always above marginal cost, provided the optimal fine is below the maximum legal limit. Proposition 3 also goes beyond the findings of Ramos and Silva (2023), where pricing below marginal cost is never possible when the mcpf is sufficiently high.

# 4 Optimal regulation under moral hazard

In this section we incorporate a moral hazard dimension into the model. To proceed, we modify the detection technology, as follows. The probability of catching an evader, now denoted by $\gamma$, can take two values: $\gamma \in \{\gamma_h, \gamma_\ell\}$, with $0 < \gamma_\ell, \gamma_h < 1$, and $\Delta\gamma \equiv \gamma_h - \gamma_\ell > 0$. By exerting the costly effort $e$, the firm affects the probability that $\gamma = \gamma_h$, $\rho(e)$.[21] The strictly increasing and concave function $\rho : [0, \infty) \to [0, 1)$ satisfies $\rho(0) = 0$. As in the previous section, we assume that, before deciding whether and how to consume, individuals know if the detection probability is $\gamma_h$ or $\gamma_\ell$.

Besides $\widehat{\theta}(p)$, new thresholds define the consumers' behavior. Let $\widehat{z}_\gamma(p) \equiv \frac{p}{\gamma}$ be the value of the reputation cost that makes an individual being indifferent between consuming good $x$ formally and informally. Also $\varphi_\gamma(z) \equiv \gamma z$ is the value of the willingness to pay that makes an individual being indifferent between consuming informally good $x$ and dropping out of its market. Figure 4 depicts those new thresholds (identified in different colours), and illustrates, in the $[0, \overline{z}] \times [0, \overline{\theta}]$ plane, the two different configurations of decisions adopted by individuals. Each configuration qualitatively replicates that of Figure 1, and emerges for each value of the detection probability $\gamma$.

[21] For example, effort $e$ can represent the number of inspectors the monopolist hires to catch evaders. The probability $\gamma$ illustrates the fact that the effective rate of detection of non-payers depends not only on the number of inspectors, but also upon other factors, out of the control of the firm.

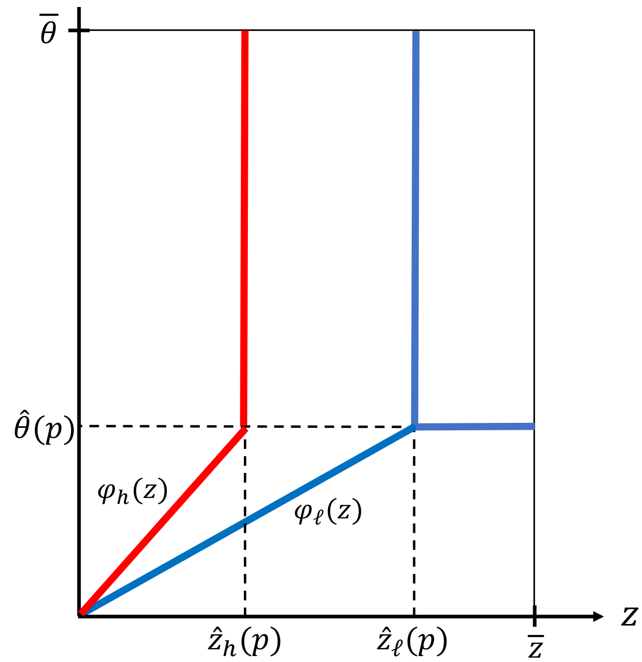



Figure 4: Individuals' decisions under moral hazard

As expected, $\widehat{z}_h(p) \equiv \widehat{z}_{\gamma_h}(p) < \widehat{z}_{\gamma_\ell}(p) \equiv \widehat{z}_\ell(p)$ and $\varphi_h(z) \equiv \varphi_{\gamma_h}(z) > \varphi_{\gamma_\ell}(z) \equiv \varphi_\ell(z)$. Hence, the higher the detection probability or the lower the price, the lower the mass of evaders. Observe that, in this context, the different thresholds do not depend upon the level of deterrence effort $e$, which has an impact only on the likelihood of each configuration.

Since effort is either non-observable or non-contractible, and it is exerted by the firm, it cannot be directly chosen by the regulator; she can only influence its level through transfers. In fact, since the detection probability $\gamma$ is observable, the regulator can condition transfers on its value, and therefore chooses $T = (T_{\gamma_h}, T_{\gamma_l})$.

For a given realization of $\gamma$, the firm's ex-post utility is given by

$$V_\gamma(p, e, T) \equiv pX_\gamma^F + T_\gamma - C(X_\gamma^F + X_\gamma^I) - \psi(e). \tag{12}$$

Now, as $e \in [0, \infty)$, the function $\psi$ satisfies $\lim_{e\to\infty} \psi(e) = \lim_{e\to\infty} \psi'(e) = \infty$. As before, we rewrite the regulator's problem as a function of $p, e$, and the firm's ex-post utilities $V_h$ and $V_\ell$, which leads to

**Lemma 4** *The regulator solves the following problem*[22]

$$\max_{p,e,V_\ell,V_h} \rho(e)W_h(p,e,V_h) + (1-\rho(e))W_\ell(p,e,V_\ell) \tag{13}$$

$$s.t. \;\; V_h \geq 0, \quad (VP_h)$$

$$V_\ell \geq 0, \quad (VP_\ell)$$

$$\rho'(e)[V_h - V_\ell] - \psi'(e) = 0, \quad (IC_{MH})$$

*where*

$$\begin{aligned} W_i(p,e,V_i) = y &+ \int_0^{\widehat{z}_i}\int_{\varphi_i}^{\overline{\theta}} \theta \, dG(\theta)dF(z) + \int_{\widehat{z}_i}^{\overline{z}}\int_{\widehat{\theta}}^{\overline{\theta}} \theta \, dG(\theta)dF(z) \\ &- \gamma_i \int_0^{\widehat{z}_i}\int_{\varphi_i}^{\overline{\theta}} z \, dG(\theta)dF(z) + \lambda p X_i^F(p) - (1+\lambda)\big[C(X_i(p)) + \psi(e)\big] - \lambda V_i, \end{aligned}$$

*and* $V_i = V_{\gamma_i}$, *for* $i \in \{h, \ell\}$.

Problem (13) has a different structure than (6), with an additional incentive constraint $(IC_{MH})$. In order to have a benchmark, we solve (13) under full information, i.e., when effort is observable and contractible, and thus the regulator does not need to induce its level. Therefore, the incentive constraint does not apply, and both participation constraints bind.[23] The following proposition characterizes the solution to this restricted problem.

**Proposition 5** *When effort is observable and contractible, the optimal price is implicitly given by the first-order conditions*

$$\begin{aligned} (1+\lambda)g(\widehat{\theta})\frac{\partial\widehat{\theta}}{\partial p}\mathbb{E}\big[1-F(\widehat{z}_i)\big](p^E - c) + (1+\lambda)\big[1-G(\widehat{\theta})\big]\mathbb{E}\big[\widehat{z}_i f(\widehat{z}_i)\big] \\ = \lambda \mathbb{E} X_i^F(p), \;\; \textit{for } i \in \{h,\ell\} \textit{ when } p^E < \gamma_\ell \overline{z}, \end{aligned} \tag{14}$$

$$\begin{aligned} (1+\lambda)g(\widehat{\theta})\frac{\partial\widehat{\theta}}{\partial p}\big[1-F(\widehat{z}_h)\big](p^E - c) + (1+\lambda)\big[1-G(\widehat{\theta})\big]\big[\widehat{z}_h f(\widehat{z}_h)\big] \\ = \lambda X_h^F(p) \textit{ when } \gamma_\ell \overline{z} \leq p^E < \gamma_h \overline{z}. \end{aligned} \tag{15}$$

[22]For the remainder of this section, we will write the thresholds $\widehat{z}$, $\widehat{\theta}$, and $\varphi$ omitting their dependence to the other variables or parameters of the model.

[23]Recall that solving this full-information benchmark is not equivalent to solve (6), because the detection technology is not comparable between these two problems.

*Moreover, the optimal level of effort $e^E$ is characterized by the following first-order condition*

$$\rho'(e^E)\big[W_h(p^E,e^E) - W_\ell(p^E,e^E)\big] = (1+\lambda)\psi'(e^E). \tag{16}$$

The regulator will not set a price $p^E \geq \gamma_h \overline{z}$, because this would imply that, as $\overline{z} < \widehat{z}_h$, she would never face formal consumers, with the ensuing loss of all firm's income. But depending upon the different parameters of the model, the regulator may find optimal to set a price such that the firm will either always receive formal payments (when $p^E < \gamma_\ell \overline{z}$) or only when the detection probability is $\gamma_h$ (when $\gamma_\ell \overline{z} \leq p^E < \gamma_h \overline{z}$).

In both cases, the intuition of the pricing rule is identical to the one that explained the first-order condition (8). The unique difference between (14) and (15) lies on the fact that, in the former, the regulator needs to identify the expected marginal evader, whereas in the latter, she only has to deal with the marginal evader when the detection probability is high. But in this context, the rule that gives the optimal level of effort is simpler than (9): $e^E$ balances the increase in the likelihood of shifting from $W_\ell(p,e,)$ to $W_h(p,e)$ against its social marginal cost.[24]

When effort is neither observable nor contractible, we assume that the moral hazard incentive constraint $IC_{MH}$ is completely characterized by its first-order condition. The following lemma characterizes the ex-post voluntary participation constraints at the optimum.

**Lemma 5** *If at the optimum $e > 0$, $V_h^E = \frac{\psi'(e)}{\rho'(e)} > V_\ell^E = 0$.*

As is standard in moral hazard problems, optimal regulation implies letting some informational rents to the firm. Indeed, to induce the monopolist to exert any positive level of effort, it must have a stake in the low evasion scenario, when $\gamma = \gamma_h$. As expected, the steeper the effort cost function $\psi$ is, the higher the informational rent should be *ceteris paribus*. But if the slope of the function $\rho$ is also large, a small increase in effort leads to a big jump in the likelihood of $\gamma_h$. Therefore, there is no need to induce the firm to increase effort substantially, which implies that the firm's compensation is attenuated. Using Lemma 4, the next proposition characterizes the optimal price-effort scheme $(p_{MH}^E, e_{MH}^E)$.

**Proposition 6** *The optimal price $p_{MH}^E$ is characterized by the first-order conditions (14) and (15).*

[24] Although it is not possible to prove that, for any pair $(p,e)$ and any configuration of parameters of the model, $W_h(p,e) > W_\ell(p,e)$, such inequality must hold at the optimum. Otherwise, the first-order condition (16) would not be satisfied.

*The optimal level of effort $e^E_{MH}$ is implicitly given by*

$$\rho'(e^E_{MH})\left[W_h(p^E_{MH},e^E_{MH}) - W_\ell(p^E_{MH},e^E_{MH})\right] = (1+\lambda)\psi'(e^E_{MH}) \\ + \lambda\psi'(e^E_{MH})\left[1 + \frac{\rho(e^E_{MH})}{\rho'(e^E_{MH})}\left(\frac{\psi''(e^E_{MH})}{\psi'(e^E_{MH})} - \frac{\rho''(e^E_{MH})}{\rho'(e^E_{MH})}\right)\right]. \quad (17)$$

The only difference between the first-order conditions (16) and (17) is the second term at the right-hand side of the latter, which corresponds to the impact of an increase in effort on the expected rent left to the firm. Such impact depends upon the relative curvatures of the functions that indicate the effort cost and the likelihood of $\gamma_h$, respectively. Moreover, conditional on the optimal level of effort $e^E_{MH}$, the pricing rule is identical to the full information one, when effort is observable and contractible. In other words, the pricing rule is not distorted by the presence of moral hazard.

This last finding is reminiscent of the 'dichotomy property' obtained by Laffont and Tirole (1990). At first glance, this similarity with their result could seem obvious because our total cost specification satisfies the separability condition that ensures that property.[25] But things are not that simple. In Laffont and Tirole's model, effort only has an impact on the firm's costs.[26] In our model, effort pushes up the likelihood of a higher level of enforcement, which leads to more formal purchases. Therefore, effort also affects the expected formal demand of good $x$, and thus one could have anticipated the **pricing rule** to be somewhat distorted to attenuate the cost of inducing the optimal effort level $e$. The intuition for this generalization of the dichotomy property lies on the fact that, to balance the marginal benefits and costs of a price change, the regulator needs to select optimally the expected marginal evader. When $p^E < \gamma_\ell \bar{z}$, effort only affects the identity of this individual through the weights $\rho(e)$ and $1-\rho(e)$. Moreover, when $p^E \geq \gamma_\ell \bar{z}$, the dichotomy is absolute: The first-order condition (15) is not affected by the choice of the level of effort.

The following corollary shows how the optimal price-effort scheme under moral hazard $(p^E_{MH}, e^E_{MH})$ compares with the corresponding one under full information.

**Corollary 3** *With moral hazard, the optimal price $p^E_{MH}$ and the effort level $e^E_{MH}$ are lower than their corresponding full-information levels.*

[25]See Proposition 3 in Laffont and Tirole (1990).
[26]More precisely, effort decreases the ex-post observable cost.

Since moral hazard entails the extra cost of leaving rents to the firm when the detection probability is high, and thus evasion is low, the regulator ends up choosing a lower level of effort than when the latter is observable and contractible. As this implies that changing the price has a smaller marginal benefit, the regulator settles on a lower price as well. Therefore, the presence of moral hazard aggravates the tendency of the regulated price to be set below marginal cost.

# 5 An empirical application

So far, we have shown theoretically that the existence of evaders limits the use of prices to fund natural monopolies. In this section, we try to verify empirically whether this issue is really of a first-order concern when prices are regulated in a context of high evasion, using data from the bus system Transantiago (Santiago, Chile) between 2013 and 2015.

To start with, the first-order condition (8) that characterizes the optimal pricing rule crucially depends upon the distributions $F(z)$ and $G(\theta)$, which are difficult to observe. Therefore, it seems *a priori* challenging to link our results to real data on public services. However, exploiting the modified Ramsey formula (11) allows to undertake this task because it only rests on some sufficient statistics and parameters that can be estimated. Therefore, in the following paragraphs, we use elasticities of evasion and total demand obtained in previous contributions, and an estimation of the marginal cost of public funds in Chile, to compute the optimal price $p^E$.

First, we derive the elasticities $\eta^F$ and $\eta^I$. Each month, the *Ministerio de Transporte y Telecomunicaciones* of Chile estimate, using as the main input direct counting of evaders in a sample of buses stops of Santiago, the share of passengers $E$ that do not validate their smart-card *BIP!*. In our notation, $E = \frac{X^I}{X}$. With this data, Troncoso and de Grange (2017) econometrically estimate a function that relates this share $E$ to the level of the fares and the number of inspections.[27] In particular, in their preferred specification, they obtain the semi-elasticity $\widehat{\beta} = \frac{\partial E}{\partial lnp} = 0.21$.

Next, we relate $E$ with the elasticities $\eta^F$ and $\eta^I$. As

$$\frac{\partial E}{\partial p} = \frac{\partial X^I}{\partial p}\frac{1}{X} - \frac{\partial X}{\partial p}\frac{X^I}{X^2}, \tag{18}$$

[27] The fact that Troncoso and de Grange (2017)'s empirical estimation takes into account the enforcement dimension is important with respect to our application, because expression (11) also depends upon the optimal level of effort.

we obtain, after some manipulations,

$$\widehat{\beta} = p\frac{\partial E}{\partial p} = E(\eta^I - \eta^D), \tag{19}$$

where $\eta^D = \frac{\partial X}{\partial p}\frac{p}{X}$. So, $\eta^I = \eta^D + \frac{0.21}{E}$. According to *Ministerio de Transporte y Telecomunicaciones*, between 2103 and 2015, the average share of evaders had been 24.27% of all passengers. Moreover, we take from de Grange et al. (2013) an average of their estimations of the price elasticity of total demand, $\eta^D = -0.317$.[28] Hence, $\eta^I = 0.548$.

Finally, we only need to compute the elasticity of the formal demand $\eta^F$. As $X = X^F + X^I$, we differentiate this equality with respect the price and we rearrange, obtaining

$$\eta^D = \frac{X^I}{X}\eta^F + \frac{X^I}{X}\eta^I.$$

So, $\eta^F = \frac{\eta^D - E\eta^I}{1-E} = -0.594$. This calibrated value is very similar to estimated price elasticities of bus systems in other cities, like -0.56 in Stockholm (Kholodov et al. 2021), -0.58 in San Francisco (McFadden 1974), and -0.64 in London (Mitrani et al. 2002).

Second, we need to assess the marginal cost of public funds (mcpf), $1 + \lambda$, in Chile. [29] Harrison et al. (2002) estimate that the mcpf of the VAT and the uniform import tariff were 1.076 and 1.185, respectively. More recently, based on the methodology adopted by Auriol and Warlters (2012) (which explicitly deals with the shadow economy), Rodríguez Ylasaca (2012) computes the mcpf for VAT, import and export tariffs, labor and capital income taxes. Altogether, with data for 2010, he obtains 1.12, which is the value we adopt for our basic calibration.

Third, we obtain a measure of the cost of an additional ride, $c$. de Grange et al. (2018) estimate, using data of Transansantiago (Santiago, Chile) from 2013 to 2015, a cost function $TC$ for the bus system. Upon other things, the cost depends on the annual number of trips $q$ (corrected to incorporate evaders). They obtain an average elasticity

$$\epsilon_{TC,q} = \frac{\partial lnTC}{\partial lnq} = \frac{\partial CT}{\partial q}\frac{q}{CT} = 1.21. \tag{20}$$

[28] These authors estimate the total demand for buses in Santiago from March to December 2010, using different specifications of discrete choice models. For each specification, we compute the simple average elasticity between peak and off-peak periods, and then average again among the different specifications.

[29] Worldwide comparable cross-country data from which to retrieve this variable is not available. Kleven and Kreiner (2006), Auriol and Warlters (2012) and Barrios et al. (2013) calculate the mcpf, but only for few countries or a group of them.

During this period of time, the number of initiated trips (validated trips plus the correction for evaders) were $1.82 \times 10^8$. But to compute the number of passengers one has to take into account that Transantiago is an integrated fare system, and thus passengers can take more than one travel with a unique validation of their smart card *BIP!*. Hence, we correct the number obtained in (20) by multiplying it to 1.524 (the average value of travels per validated card from 2013 to 2015, as stated by Directorio de Transporte Publico Metropolitano (2015)). As the average yearly cost was $CT = 8.29 \times 10^7$ thousand of Chilean pesos (CLP), we recover

$$\frac{\partial CT}{\partial q} = \epsilon_{TC,q} \cdot \frac{CT}{q} = 1.21 \times 298.85 = 361.61 \text{ CLP}. \tag{21}$$

Finally, with all these values, we proceed to calculate $p^E$ from (11). As $\frac{X^I}{X^F} = \frac{E}{1-E} = 0.32$ and $\frac{\eta^I}{\eta^F} = -0.923$, $p^E = 310.71$ CLP.

Although this is a crude estimation, based on a stylized framework, we can make the following observations. First, this estimation is an upper bound of the optimal price. The reason is that, as private companies are in charge of detecting evaders, regulators face in fact moral hazard, a situation that we know further depresses the optimal fare. Second, the calibrated optimal price $p^E$ is 49.4% lower than the observed average fare of Transantiago between 2013 and 2015, which was 613.61 CLP. This difference is quantitatively significant, and quite robust. Indeed, if we consider all possible combinations of the estimated parameters that appear in the contributions mentioned in this section, the highest computed value of the optimal price $p^E$ is 377.62 CLP, implying that the abovementioned difference only reduces to 38.46%. Hence, this 'worst case' scenario confirms that the difference between the observed and the computed fare is important. Fourth, in the basic calibration, the optimal price $p^E$ is 14.08% below the assessment of the marginal cost $c$. This result persists qualitatively for all parameter configurations that we consider, supporting empirically the assertion that, in contexts of high evasion, pricing below marginal cost is not just a theoretical possibility, but a relevant feature of the optimal regulation of natural monopolies.

# 6 Conclusions

In this paper, we provide a normative model in which regulation of a natural monopoly under the threat of evasion is the focus, and in particular, we analyze the tension between enforcement and assuring the firm's participation via socially costly transfers or prices. We highlight the main channels through which we believe evasion alters the received theory of natural monopoly regulation, stressing the relation between prices, marginal costs of production, and the deadweight loss of taxation.

Our results generate some important policy implications. First, even when the tax system is not distorted and thus the firm's cost could be entirely covered with transfers, the mere existence of individuals that may evade prevents the regulator price at marginal cost but also to provide the good for free, and calls for the use of deterrence effort. Second, we proved that evasion really dampens the use of prices as feasible instruments to mitigate financial deficits of regulated firms. Indeed, under plausible circumstances, it may be optimal to regulate the price below the marginal cost, for any value of the marginal cost of public funds. Third, the presence of moral hazard aggravates this tendency towards low prices, tendency that we confirm in the empirical application of our model.

We conclude by emphasizing some limitations of our analysis and present some possible extensions. So far, we have followed the traditional welfare approach in deterrence models, assuming that the regulator weights equally formal consumers and evaders. But this may be troublesome and politically costly to defend, so it can be of interest to inquire what the regulator should do if evaders' welfare is not taken into account in the measurement of the total surplus. A priori, one could have expected that such a change may trigger an increase in deterrence effort and thus, also on the price. Surprisingly, in Appendix N we show that, in fact, the optimal price should decrease, as a way to prevent formal buyers to start evading, in which case their utility from consumption would be lost. This result reinforces our view that evasion exerts a noticeable downward pressure upon regulated prices.

Finally, we have ignored that firms may have different efficiency levels to detect evaders. Incorporating this adverse selection feature into the model obviously modifies our results. But, more importantly, it could enable us to go one step ahead, and study the regulator's decision of granting the monopoly when it faces a pool of diverse potential entrants in a context of pervasive evasion. This analysis can help regulators to improve the design of bidding mechanisms for the provision of public services when such misbehavior is an institutional concern.

# References

Armstrong, M. and D. Sappington (2007). Recent developments in the theory of regulation. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*

*Volume 3*, pp. 1557–1700. North Holland, Elsevier, London, United Kingdom.

Armstrong, M. and J. Vickers (1993). Price discrimination, competition and regulation. *The Journal of Industrial Economics 41*(4), 335–359.

Auriol, E., C. Crampes, and A. Estache (2021). *Regulating Public Services Bridging the Gap Between Theory and Practice*. Cambridge University Press, Cambridge, United Kingdom.

Auriol, E. and M. Warlters (2012). The marginal cost of public funds and tax reform in Africa. *Journal of Development Economics 97*(1), 58–72.

Ayal, S., J. Celse, and G. Hochman (2021). Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion. *Organizational Behavior and Human Decision Processes 166*, 9–19.

Barrios, S., J. Pycroft, and B. Saveyn (2013). The marginal cost of public funds in the EU: The case of labour versus green taxes. In M. Marino, M. Tasso, and P. Tommasino (Eds.), *Fiscal Policy and Growth*, pp. 403–431. Banca d'Italia, Roma.

Baumol, W. and D. Bradford (1970). Optimal departures from marginal cost pricing. *American Economic Review 60*(3), 265–283.

Bijleveld, C. (2007). Fare dodging and the strong arm of the law An experimental evaluation of two different penalty schemes for fare evasion. *Journal of Experimental Criminology 3*(2), 183–199.

Boiteux, M. (1956). Sur la gestion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica 24*(1), 22–40.

Bucciol, A., F. Landini, and M. Piovesan (2013). Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization 93*, 248–257.

Buehler, S., D. Halbheer, and M. Lechner (2017). Payment evasion. *The Journal of Industrial Economics 65*(4), 804–832.

Caillaud, B., R. Guesnerie, P. Rey, and J. Tirole (1988). Government Intervention in Production and Incentives Theory: A Review of Recent Contributions. *RAND Journal of Economics 19*(1), 1–26.

Calvo, M. (2015). Análisis desagregado del comportamiento de usuarios de transporte público utilizando datos masivos. Memoria, Departamento de Ingeniería Civil, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

Chander, P. and L. Wilde (1998). A General Characterization of Optimal Income Tax Enforcement. *The Review of Economic Studies 65*(1), 165–183.

Dahlby, B. (2008). *The Marginal Cost of Public Funds: Theory and Applications*. The MIT Press, Cambridge, Massachusetts.

Dai, Z., F. Galeotti, and M. Villeval (2019). Fare-dodging in the lab and the moral cost of dishonesty. In A. Bucciol and N. Montinari (Eds.), *Dishonesty in Behavioral Economics*, pp. 245–265. Academic Press, Elsevier, London.

de Grange, L., F. González, J. Muñoz, and R. Troncoso (2013). Aggregate estimation of the price elasticity of demand for public transport in integrated fare systems: The case of Transantiago. *Transport Policy 29*, 178–185.

de Grange, L., R. Troncoso, and I. Briones (2018). Cost, production and efficiency in local bus industry: An empirical analysis for the bus system of Santiago. *Transportation Research Part A 108*, 1–11.

Egu, O. and P. Bonnel (2020). Can we estimate accurately fare evasion without a survey? Results from a data comparison approach in Lyon using fare collection data, fare inspection data and counting data. *Public Transport 12*, 1–26.

Estache, A. and L. Wren-Lewis (2009). Toward a Theory of Regulation for Developing Countries: Following Jean-Jacques Laffont's Lead. *Journal of Economic Literature 47*(3), 729–770.

Harrison, G., T. Rutherford, and D. Tarr (2002). Trade Policy Options for Chile: The Importance of Market Access. *The World Bank Economic Review 16*(1), 49–79.

Kholodov, Y., E. Jenelius, O. Cats, N. van Oort, N. Mouter, M. Cebecauer, and A. Vermeulen (2021). Public transport fare elasticities from smartcard data: Evidence from a natural experiment. *Transport Policy 105*, 35–43.

Kleven, H. and T. Kreiner (2006). The marginal cost of public funds: Hours of work versus labor force participation. *Journal of Public Economics 90*, 1955–1973.

Laffont, J.-J. (1994). The New Economics of Regulation Ten Years After. *Econometrica 62*(3), 507–537.

Laffont, J.-J. (2005). *Regulation and Development*. Cambridge University Press, Cambridge, United Kingdom.

Laffont, J.-J. and J. Tirole (1985). Using Cost Observation to Regulate Firms. Working paper 368, Department of Economics, MIT.

Laffont, J.-J. and J. Tirole (1990). The Regulation of Multiproduct Firms Part I: Theory. *Journal of Public Economics 43*(1), 1–36.

Laffont, J.-J. and J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation*. The MIT Press, Cambridge, Massachusetts, United States.

Lee, J. (2011). Uncovering San Francisco, California, Muni's Proof-of-Payment Patterns to Help Reduce Fare Evasion. *Transportation Research Record 2216*(1), 75–84.

McFadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics 3*(4), 303–328.

Meade, J. (1944). Price and output policy of state enterprise. *The Economic Journal 54*(215/216), 321–328.

Mitrani, A., I. Kincaid, D. Edwards, and G. Hobbs (2002). London Underground and Bus Demand Analysis 1970—2000. Report, London Transport.

Ng, Y.-K. and M. Weisser (1974). Optimal pricing with a budget constraint: the case of the two-part tariff. *The Review of Economic Studies 41*(3), 337–345.

Ramos, R. and H. Silva (2023). Fare evasion in public transport: How does it affect the optimal design and pricing? *Transportation Research Part B: Methodological 176*, Forthcoming.

Ramsey, F. (1927). A contribution to the theory of taxation. *The Economic Journal 37*(145), 47–61.

Rana, A., M. Ngulube, and V. Foster (2022). Global Electricity Regulatory Index (GERI) 2022. ESMAP report, World Bank.

Rodríguez Ylasaca, E. (2012). Análisis del Costo Marginal de los Fondos Públicos para Chile. Master Thesis, Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

Sherman, R. (1989). *The regulation of monopoly*. Cambridge University Press, Cambridge.

Silva, E. and C. Kahn (1993). Exclusion and moral hazard The case of identical demand. *Journal of Public Economics 52*(2), 217–235.

Slemrod, J. (2019). Tax Compliance and Enforcement. *Journal of Economic Literature 57*(4), 904–954.

Smith, T. (2004). Electricity theft: a comparative analysis. *Energy Policy 32*(18), 2067–2076.

Srinagesh, P. (1991). Mixed linear-nonlinear pricing with bundling. *Journal of Regulatory Economics 3*, 251–263.

Sterner, A. and S. Sheng (2013). The effect of social stigma on fare evasion in Stockholm's public transport. *Journal of Transport Literature 7*(4), 50–74.

Troncoso, R. and L. de Grange (2017). Fare evasion in public transport: A time series approach. *Transportation Research Part A: Policy and Practice 100*, 311–318.

The World Bank (2014). Electric power transmission and distribution losses (% of output). data retrieved from World Development Indicators, https://data.worldbank.org/indicator/EG.ELC.LOSS.ZS.

Yurtseven, C. (2015). The causes of electricity theft: An econometric analysis of the case of Turkey. *Utilities Policy 37*, 70–78.

## Appendix A: Proof of Lemma 1

i) Let $\lambda > 0$. Assume that $VP$ does not bind at the optimum. If so, we can always decrease $V^E$ by a small $\epsilon > 0$ such that $VP$ still holds. This change leads to an increase in expected welfare $W(p, e, V)$, which is a contradiction.
ii) If $\lambda = 0$, there is a continuum of optimal policies $(p^E, e^E, \tilde{V})$, with $\tilde{V} \geq 0$, all generating the same level of social welfare. For the sake of completeness, we take $\tilde{V} = 0$ ■

## Appendix B: Non-optimality of a configuration with no formal consumers

If $p \geq e\overline{z}$, formal demand vanishes. Hence, total consumption $X(p, e) = X^I(p, e) = \int_0^{\overline{z}} \int_{\varphi(e,z)}^{\overline{\theta}} dG(\theta)dF(z)$, which does not depend upon the price $p$. Under these circumstances, the social welfare is

$$\widehat{W}(e) = y + \int_0^{\overline{z}} \int_{\varphi(e,z)}^{\overline{\theta}} \theta \, dG(\theta)dF(z) - \int_0^{\overline{z}} \int_{\varphi(e,z)}^{\overline{\theta}} ez \, dG(\theta)dF(z) \\ - (1+\lambda)\left(K + c \int_0^{\overline{z}} \int_{\varphi(e,z)}^{\overline{\theta}} dG(\theta)dF(z) + \psi(e)\right). \quad (22)$$

Without any loss of generality, we assume that the regulator sets the lowest price compatible with this case, namely $p^E = e\overline{z}$. Moreover,

$$\lim_{p \to e\overline{z}} W(p, e) = \widehat{W}(e).$$

Finally, for any any value of the effort $e$, we compute

$$\lim_{p \to e\overline{z}} \frac{\partial W(p, e)}{\partial p} = -(1+\lambda)\overline{z}f(\overline{z})\left[1 - G\left(\frac{e\overline{z}}{b}\right)\right] < 0,$$

because $\overline{z} > 0$, $f(\overline{z}) > 0$, and $e\overline{z} \leq \overline{z} < \overline{\theta}$. We thus conclude that the regulator always prefers to set $p^E < e^E\overline{z}$ ■

## Appendix C: Proof of Proposition 1

The inequality $p^E < c$ is straightforward: By replacing $\lambda = 0$ in (8), it is the unique way to satisfy this first-order condition.

Then, we want to show that, when $\lambda = 0$, the regulator never sets $p^E = e^E = 0$. When $p = e = 0$, all individuals are indifferent between consuming formally at a null price or evading its payment without being detected. Moreover, they are all strictly worse off if they drop out of the market. Therefore, expected welfare is

$$W(0,0) = y + \theta^e - (K + c), \tag{23}$$

where $\theta^e = \int_0^{\overline{\theta}} \theta dG(\theta)$.

From this limit case, we next try to find a welfare-enhancing policy-deviation. Let be a sequence of rays $e = \frac{n}{\overline{z}}p$, with $n \in ]1, \infty[$. When the price-effort scheme lies on one of these rays, $\widehat{z} = \frac{\overline{z}}{n}, \widehat{\theta} = p$, and $\varphi = \frac{n}{\overline{z}}pz$. Thus, the regulator faces a configuration with formal consumers, and expected welfare is

$$\begin{aligned}\widetilde{W}_n(p) = y + \int_0^{\frac{\overline{z}}{n}} \int_{\frac{n}{\overline{z}}pz}^{\overline{\theta}} \theta\, dG(\theta) dF(z) + \int_{\frac{\overline{z}}{n}}^{\overline{z}} \int_p^{\overline{\theta}} \theta\, dG(\theta) dF(z) \\ - \int_0^{\frac{\overline{z}}{n}} \int_{\frac{n}{\overline{z}}pz}^{\overline{\theta}} \frac{n}{\overline{z}} pz\, dG(\theta) dF(z) - \left( c \cdot \left[ X_n^F(p) + X_n^I(p) \right] + \psi\left( \frac{n}{\overline{z}}p \right) \right),\end{aligned} \tag{24}$$

where $X_n^F = \left[1 - F\left(\frac{\overline{z}}{n}\right)\right] [1 - G(p)]$ and $X_n^I = \int_0^{\frac{\overline{z}}{n}} \int_{\frac{n}{\overline{z}}pz}^{\overline{\theta}} dG(\theta) dF(z)$. We then compute

$$\begin{aligned}\frac{\partial \widetilde{W}_n(p)}{\partial p} = -pg(p) \left[ 1 - F\left( \frac{\overline{z}}{n} \right) \right] - \frac{n}{\overline{z}} \int_0^{\frac{\overline{z}}{n}} z \left[ 1 - G\left( \frac{n}{\overline{z}} pz \right) \right] dF(z) + cg(p) \left[ 1 - F\left( \frac{\overline{z}}{n} \right) \right] \\ + c \frac{n}{\overline{z}} \int_0^{\frac{\overline{z}}{n}} zg\left( \frac{n}{\overline{z}} pz \right) dF(z) - \psi'\left( \frac{n}{\overline{z}} p \right) \frac{n}{\overline{z}}.\end{aligned} \tag{25}$$

When $p \to 0$ along these rays, the equilibrium features a mass $F\left(\frac{\overline{z}}{n}\right)$ of evaders (i.e., individuals that evade the payment of a null price and do not suffer any subjective cost), and a mass $1 - F\left(\frac{\overline{z}}{n}\right)$ of formal consumers (i.e., individuals that pay a null price). We also have

$$\lim_{p \to 0} \frac{\partial \widetilde{W}_n(p)}{\partial p} = \frac{n}{\overline{z}} \int_0^{\frac{\overline{z}}{n}} z dF(z) \left( cg(0) - 1 \right) + cg(0) \left[ 1 - F\left( \frac{\overline{z}}{n} \right) \right]. \tag{26}$$

As the first term in (26) depends upon parameters of the model, and thus can be negative, we cannot be sure that marginal increases in $p$ and $e$ along *all* these rays are welfare-improving. The following lemma proves that, no matter the values of $c$ and $g(0)$, there exists some rays where our result holds.

**Lemma 6**

$$\lim_{n \to \infty} \left( \lim_{p \to 0} \frac{\partial \widetilde{W}_n(p)}{\partial p} \right) = cg(0) > 0. \tag{27}$$

**Proof.** Let's denote by $x = \frac{\bar{z}}{n}$. Applying l'Hospital's rule, we obtain

$$\lim_{n\to\infty} \frac{n}{\bar{z}} \int_0^{\frac{\bar{z}}{n}} z dF(z) = \lim_{x\to 0} \frac{\int_0^x z dF(z)}{x} = \lim_{x\to 0} \frac{x f(x)}{1} = 0. \tag{28}$$

As $\frac{\bar{z}}{n} \to 0$, the result is immediate ■

If $n$ is big enough, deviating from $p = e = 0$ along a ray of slope $\frac{n}{\bar{z}}$ increases welfare. This immediately proves that, at the optimum, $p^E > 0$ and $e^E > 0$ ■

## Appendix D: Proof of Lemma 3

First, we compute

$$\frac{\partial^2 W(p,e)}{\partial\lambda\partial p} = \frac{1}{\lambda}\left[\frac{p^E - c}{p^E}\widehat{\theta} g(\widehat{\theta})[1 - F(\widehat{z})] + \widehat{z} f(\widehat{z})[1 - G(\widehat{\theta})]\right] > 0, \tag{29}$$

$$\frac{\partial^2 W(p,e)}{\partial\lambda\partial e} = \frac{1}{1+\lambda}\int_0^{\widehat{z}} z f(z)\left[1 - G(\varphi)\right] dz > 0, \tag{30}$$

and

$$\frac{\partial^2 W(p,e)}{\partial e\partial p} = -(1+\lambda)\left[1 - G(\widehat{\theta})\right] f(\widehat{z}) \frac{\partial \widehat{z}}{\partial e}\left[1 + \widehat{z}\frac{f'(\widehat{z})}{f(\widehat{z})} + \widehat{z}\frac{f(\widehat{z})}{1 - F(\widehat{z})}\right] > 0. \tag{31}$$

The sign of (29) directly follows from the first-order condition (8). Together with the fact that $\frac{\partial \widehat{z}}{\partial e} < 0$, the following lemma enables us to prove the sign of (31).

**Lemma 7** *At the optimum,* $\widehat{z}\frac{f'(\widehat{z})}{f(\widehat{z})} > -1$.

**Proof.** Since $f$ is log-concave, $\frac{f'(z)}{f(z)}$ decreases with $z$, reaching its minimal value when $z = \bar{z}$.
If $f'(\bar{z})/f(\bar{z}) \geq 0$, $f'(z)/f(z) \geq 0$ for all $z$, and thus the inequality mentioned in the lemma is obviously satisfied. Otherwise, if $f'(\bar{z})/f(\bar{z}) < 0$, we need to consider the following two sub-cases:

- $\frac{f'(\bar{z})}{f(\bar{z})} < 0 < \frac{f'(\widehat{z})}{f(\widehat{z})}$, which implies the result.
- $\frac{f'(\bar{z})}{f(\bar{z})} \leq \frac{f'(\widehat{z})}{f(\widehat{z})} < 0$, which yields $\bar{z}\frac{f'(\bar{z})}{f(\bar{z})} \leq \widehat{z}\frac{f'(\widehat{z})}{f(\widehat{z})}$. Thus, by Assumption 1, $\widehat{z}\frac{f'(\widehat{z})}{f(\widehat{z})} > -1$ ■

With these results, it is straightforward to obtain

$$\frac{dp^E}{d\lambda} = \frac{1}{\text{SOC}}\left( -\frac{\partial^2 W(p,e)}{\partial\lambda\partial p} \cdot \frac{\partial^2 W(p,e)}{\partial e^2} + \frac{\partial^2 W(p,e)}{\partial\lambda\partial e} \cdot \frac{\partial^2 W(p,e)}{\partial e\partial p} \right) > 0, \tag{32}$$

and

$$\frac{de^E}{d\lambda} = \frac{1}{\text{SOC}}\left( -\frac{\partial^2 W(p,e)}{\partial p^2} \cdot \frac{\partial^2 W(p,e)}{\partial\lambda\partial e} + \frac{\partial^2 W(p,e)}{\partial\lambda\partial p} \cdot \frac{\partial^2 W(p,e)}{\partial e\partial p} \right) > 0, \tag{33}$$

because we have assumed that the second-order conditions SOC for a maximum hold.

## Appendix E: Proof of Proposition 2

We can rewrite (8) as

$$\lambda\left[X^F + p^E\frac{\partial X^F}{\partial p}\right] - (1+\lambda)c\left(\frac{\partial X^F}{\partial p} + \frac{\partial X^I}{\partial p}\right) = \Gamma, \tag{34}$$

where

$$\Gamma \equiv \widehat{\theta}g(\widehat{\theta})\int_{\widehat{z}}^{\overline{z}} f(z)dz + \widehat{z}f(\widehat{z})\int_{\widehat{\theta}}^{\overline{\theta}} g(\theta)d\theta. \tag{35}$$

Straightforward manipulations yield

$$(1+\lambda)(p^E - c)\frac{p^E}{p^E}\frac{1}{X^F}\frac{\partial X^F}{\partial p} = \frac{\Gamma}{X^F} + \frac{p^E}{X^F}\frac{\partial X^F}{\partial p} - \lambda + (1+\lambda)\frac{c}{X^F}\frac{\partial X^I}{\partial p}. \tag{36}$$

If we denote by $\eta^F \equiv \frac{p^E}{X^F}\frac{\partial X^F}{\partial p}$ the price elasticity of the total formal demand, rearranging (36) obtains

$$\begin{aligned}\frac{p^E - c}{p^E} &= \frac{\Gamma}{(1+\lambda)X^F\eta^F} + \frac{1}{1+\lambda} - \frac{\lambda}{(1+\lambda)\eta^F} + \frac{c}{X^F}\frac{\partial X^I \backslash \partial p}{\eta^F} \\ &= \frac{\Gamma}{(1+\lambda)p^E\frac{\partial X^F}{\partial p}} + \frac{1}{1+\lambda} - \frac{\lambda}{(1+\lambda)\eta^F} + \frac{c}{p^E}\frac{\partial X^I \backslash \partial p}{\partial X^F \backslash \partial p}.\end{aligned} \tag{37}$$

As

$$X^F = \int_{\widehat{z}}^{\overline{z}} f(z)dz \int_{\widehat{\theta}}^{\overline{\theta}} g(\theta)d\theta,$$

$$\frac{\partial X^F}{\partial p} = -g(\widehat{\theta})\frac{\partial \widehat{\theta}}{\partial p}\int_{\widehat{z}}^{\overline{z}} f(z)dz - f(\widehat{z})\frac{\partial \widehat{z}}{\partial p}\int_{\widehat{\theta}}^{\overline{\theta}} g(\theta)d\theta. \tag{38}$$

Hence,

$$\begin{aligned} p^E\frac{\partial X^F}{\partial p} &= p^E\left( - g(\widehat{\theta})\frac{\partial \widehat{\theta}}{\partial p}\int_{\widehat{z}}^{\overline{z}} f(z)dz - f(\widehat{z})\frac{\partial \widehat{z}}{\partial p}\int_{\widehat{\theta}}^{\overline{\theta}} g(\theta)d\theta\right) \\ &= -\widehat{\theta}g(\widehat{\theta})\int_{\widehat{z}}^{\overline{z}} f(z)dz - \widehat{z}f(\widehat{z})\int_{\widehat{\theta}}^{\overline{\theta}} g(\theta)d\theta \\ &= -\Gamma \end{aligned} \tag{39}$$

So finally the first two expressions of the right-hand side of (37) cancel out, and thus

$$\frac{p^E - c}{p^E} = -\frac{\lambda}{(1+\lambda)\eta^F} + \frac{c}{p^E}\frac{\partial X^I \backslash \partial p}{\partial X^F \backslash \partial p}. \tag{40}$$

If we also denote by $\eta^I \equiv \frac{p^E}{X^I}\frac{\partial X^I}{\partial p}$ the price elasticity of the total informal demand, rearranging (40) yields (11) ■

## Appendix F: Proof of Corollary 1

First, we nest the scenario when all individuals have identical reputation cost $z_1 < \overline{z}$ with the general case where $z$ follows a continuous distribution $F(z)$. To do that, we define the following cumulative distribution function for $z$,

$$\Omega(z) = \epsilon F(z) + (1-\epsilon)\delta_{z_1}(z), \tag{41}$$

where $\epsilon \in [0,1]$ and $\delta_{z_1}(z)$ is the Dirac delta function at $z = z_1$.

**Case 1** When $p > ez_1$ (and thus $\widehat{z} > z_1$), the social welfare is

$$\begin{aligned} W_1(p,e) &= y + \int_0^{\widehat{z}}\int_{\varphi}^{\overline{\theta}} \theta dG(\theta)d\Omega(z) + \int_{\widehat{z}}^{\overline{z}}\int_{\widehat{\theta}}^{\overline{\theta}} \theta dG(\theta)d\Omega(z) \\ &\quad - \int_0^{\widehat{z}}\int_{\varphi}^{\overline{\theta}} ezdG(\theta)d\Omega(z) + \lambda pX^F(p,e) - (1+\lambda)\big[C(X(p,e)) + \psi(e)\big] \\ &= y + \epsilon\int_0^{\widehat{z}}\int_{ez}^{\overline{\theta}} \theta dG(\theta)dF(z) + (1-\epsilon)\int_{ez_1}^{\overline{\theta}} \theta dG(\theta) + \epsilon\int_{\widehat{z}}^{\overline{z}}\int_{\widehat{\theta}}^{\overline{\theta}} \theta dG(\theta)dF(z) \\ &\quad -\epsilon\int_0^{\widehat{z}}\int_{ez}^{\overline{\theta}} ezdG(\theta)dF(z) - (1-\epsilon)z_1\int_{ez_1}^{\overline{\theta}} edG(\theta) + \epsilon\lambda p[1-F(\widehat{z})][1-G(\widehat{\theta})] \\ &\quad -\epsilon(1+\lambda)c\Big([1-F(\widehat{z})][1-G(\widehat{\theta})] + \int_0^{\widehat{z}}\int_{ez}^{\overline{\theta}} dG(\theta)dF(z)\Big) \\ &\quad -(1-\epsilon)(1+\lambda)c\int_{ez_1}^{\overline{\theta}} dG(\theta) - (1+\lambda)\psi(e). \end{aligned} \tag{42}$$

Taking limits in (42), we obtain

$$\begin{aligned} W_1^{lim}(e) \equiv \lim_{\epsilon\to 0} W_1(p,e) = \; & y + \int_{ez_1}^{\overline{\theta}} \theta dG(\theta) - z_1\int_{ez_1}^{\overline{\theta}} edG(\theta) \\ & -(1+\lambda)c\int_{ez_1}^{\overline{\theta}} dG(\theta) - (1+\lambda)\psi(e). \end{aligned} \tag{43}$$

**Case 2** When $p \le ez_1$ (and thus $\widehat{z} \le z_1$), the social welfare is

$$\begin{aligned} W_2(p,e) &= y + \int_0^{\widehat{z}}\int_{\varphi}^{\overline{\theta}} \theta dG(\theta)d\Omega(z) + \int_{\widehat{z}}^{\overline{z}}\int_{\widehat{\theta}}^{\overline{\theta}} \theta dG(\theta)d\Omega(z) \\ &\quad - \int_0^{\widehat{z}}\int_{\varphi}^{\overline{\theta}} ezdG(\theta)d\Omega(z) + \lambda pX^F(p,e) - (1+\lambda)\big[C(X(p,e)) + \psi(e)\big] \\ &= y + \epsilon\int_0^{\widehat{z}}\int_{ez}^{\overline{\theta}} \theta dG(\theta)dF(z) + (1-\epsilon)\int_{\widehat{\theta}}^{\overline{\theta}} \theta dG(\theta) + \epsilon\int_{\widehat{z}}^{\overline{z}}\int_{\widehat{\theta}}^{\overline{\theta}} \theta dG(\theta)dF(z) \\ &\quad -\epsilon\int_0^{\widehat{z}}\int_{ez}^{\overline{\theta}} ezdG(\theta)dF(z) + \epsilon\lambda p[1-F(\widehat{z})][1-G(\widehat{\theta})] + (1-\epsilon)\lambda p[1-G(\widehat{\theta})] \\ &\quad -\epsilon(1+\lambda)c\Big([1-F(\widehat{z})][1-G(\widehat{\theta})] + \int_0^{\widehat{z}}\int_{ez}^{\overline{\theta}} dG(\theta)dF(z)\Big) \\ &\quad -(1-\epsilon)(1+\lambda)c\int_{\widehat{\theta}}^{\overline{\theta}} dG(\theta) - (1+\lambda)\psi(e). \end{aligned} \tag{44}$$

Taking limits in (44), we obtain

$$\begin{aligned} W_2^{lim}(p,e) &\equiv \lim_{\epsilon\to 0} W_2(p,e) \\ &= y + \int_{\widehat{\theta}}^{\overline{\theta}} \theta dG(\theta) + \lambda p\big[1-G(\widehat{\theta})\big] \\ &\quad -(1+\lambda)c\int_{\widehat{\theta}}^{\overline{\theta}} dG(\theta) - (1+\lambda)\psi(e). \end{aligned} \tag{45}$$

Then, we show that the regulator always prefers to chose a price-effort scheme $(p,e)$

satisfying $p \leq ez_1$. As $W_1^{lim}(e)$ does not depend upon the price $p$, it is straightforward to show that, for any level of effort $e$,

$$\begin{aligned} \lim_{p\to ez_1} W_2^{lim}(p,e) = \; & y + \int_{ez_1}^{\overline{\theta}} \theta dG(\theta) + \lambda p\left[1 - G(ez_1)\right] - (1+\lambda)c\int_{ez_1}^{\overline{\theta}} dG(\theta) \\ & -(1+\lambda)\psi(e) \\ & > W_1^{lim}(e). \end{aligned} \tag{46}$$

Second, to obtain the optimal price-effort scheme, the regulator solves the following problem,

$$\max_{p,e} W_2^{lim}(p,e) \quad \text{subject to} \quad p \leq ez_1. \tag{47}$$

The Lagrangean of this problem is

$$\begin{aligned} \mathcal{L} = \; & y + \int_{\widehat{\theta}}^{\overline{\theta}} \theta dG(\theta) + \lambda p\left[1 - G(\widehat{\theta})\right] - (1+\lambda)c\int_{\widehat{\theta}}^{\overline{\theta}} dG(\theta) \\ & -(1+\lambda)\psi(e) - \alpha\left[p - ez_1\right], \end{aligned} \tag{48}$$

where $\alpha$ is the multiplier of the inequality constraint. Recalling that $\widehat{\theta} = p$, the first-order conditions of problem (47) are

$$\frac{\partial \mathcal{L}}{\partial p} = -(1+\lambda)(p-c)g(\widehat{\theta}) + \lambda\left[1 - G(\widehat{\theta})\right] - \alpha = 0, \tag{49}$$

$$\frac{\partial \mathcal{L}}{\partial e} = -(1+\lambda)\psi'(e) + \alpha z_1 = 0, \tag{50}$$

and the complementary-slackness condition is

$$\alpha.\left[p - ez_1\right] = 0. \tag{51}$$

**Lemma 8** *At the optimum of problem (47), $p^E = e^E z_1$ and $\alpha > 0$.*

**Proof.** Assume that the inequality constraint $p \leq ez_1$ is slack at the optimum, and thus that $\alpha = 0$. If so, (50) implies that $e = 0$, and $p < 0$. But this yields to a contradiction in (49). A similar argument can be used to show that assuming that, at the optimum, the inequality constraint binds and that $\alpha = 0$ also yields to a contradiction ■

Hence, rearranging (49), (50), and (51) the optimal price-effort scheme is implicitly characterized by the following conditions,

$$\frac{p^E - c}{p^E} = -\frac{\lambda}{1+\lambda}\frac{1 - G(\widehat{\theta})}{p^E g(\widehat{\theta})} - \frac{\psi'(e^E)}{z_1 p^E g(\widehat{\theta})} \tag{52}$$

$$p^E = e^E z_1. \tag{53}$$

As in this case the total formal demand is

$$X^F(p) = \left[1 - G(\widehat{\theta})\right] \int_{\widehat{z}}^{\bar{z}} [\epsilon f(z)dz + (1-\epsilon)\delta_{z_1}(z)] = \left[1 - G(\widehat{\theta})\right]\left[1 - \epsilon F(\widehat{z})\right],$$

$$\frac{\partial X^F}{\partial p} = -g(\widehat{\theta})\left[1 - \epsilon F(\widehat{z})\right] - \epsilon f(\widehat{z})\frac{\partial \widehat{z}}{\partial p}\left[1 - G(\widehat{\theta})\right].$$

Thus, we can define the elasticity of formal demand

$$\eta^F = \frac{p}{X^F(p)}\frac{\partial X^F}{\partial p} = \frac{p}{\left[1 - G(\widehat{\theta})\right]\left[1 - \epsilon F(\widehat{z})\right]}\frac{\partial X_F}{\partial p}. \tag{54}$$

Taking limits

$$X_1^F(p) = lim_{\epsilon \to 0} X^F(p) = 1 - G(\widehat{\theta}), \tag{55}$$

and

$$\eta_1^F = lim_{\epsilon \to 0}\eta^F = \frac{-pg(\widehat{\theta})}{1 - G(\widehat{\theta})}.$$

So we can rewrite (52) as

$$\frac{p^E - c}{p^E} = -\frac{\lambda}{1+\lambda}\frac{1}{\eta_1^F} - \frac{\psi'(e^E)}{z_1 p^E g(\widehat{\theta})}. \tag{56}$$

Finally, considering $z_1 \to \infty$, which requires $\bar{z} \to \infty$, we have that $p^E \to p^*$ and $e^E \to 0$ ■

## Appendix G: Proof of Corollary 2

First, we characterize the price set by an unregulated monopoly that faces evaders. The monopoly solves

$$\max_{p,e} \Pi(p,e) = \quad pX^F(p,e) - c\big(X^F(p,e) + X^I(p,e)\big) - \psi(e).$$

Rearranging $\frac{\partial \Pi(p,e)}{\partial p}$, we obtain

$$X^F(p,e)\eta^F\left[\frac{p-c}{p} + \frac{1}{\eta^F} - \frac{c}{p}\frac{X^I(p,e)}{X^F(p,e)}\frac{\eta^I}{\eta^F}\right]. \tag{57}$$

The price set by an unregulated monopoly that faces evaders $p_M^E$ is implicitly characterized by equalizing the expression in square brackets in (57) to 0, which replicates expression (11) when $\lambda \to \infty$. Hence, the optimal price $p^E \to p^E_M$ when $\lambda \to \infty$.

Next, let's define $p_M$ as the price set by an unregulated monopoly that only deals with formal buyers, which is implicitly characterized by

$$\frac{p_M - c}{p_M} = -\frac{1}{\eta}, \tag{58}$$

where, in the context of our model, the demand elasticity is

$$\eta = \frac{-pg(p)}{1 - G(p)}$$

Finally, we assess the sign of (57), evaluated at $p_M$. As in this case

$$\eta^F = -\frac{\widehat{z}(p_M) f(\widehat{z}(p_M))}{1 - F(\widehat{z}(p_M))} - \frac{p_M g(p_M)}{1 - G(p_M)},$$

$$\frac{p_M - c}{p_M} + \frac{1}{\eta^F} - \frac{c}{p_M} \frac{X^I(p_M, e)}{X^F(p_M, e)} \frac{\eta^I}{\eta^F} > \underbrace{\frac{p_M - c}{p_M} + \frac{1}{\eta}}_{=0} - \frac{c}{p_M} \frac{X^I(p_M, e)}{X^F(p_M, e)} \frac{\eta^I}{\eta^F} > 0.$$

Hence, at $p_M$, the sign of (57) is strictly negative (because $\eta^F < 0$ and $\eta^I > 0$), which implies that $p^E_M < p_M$. Then, as $p^E$ increases with $\lambda$, the inequality always holds ■

## Appendix H: Proof of Proposition 3

By simply observing (11) when $\lambda \to 0$, $p^E < c$ because the first term vanishes, while the second term is negative. Hence, applying a continuity argument, we know that this result would hold in an interval of values of $\lambda$ close to 0.

When $\lambda \to \infty$, $(p^E, e^E) \to (p^E_{\ell im} = p^E_M, e^E_{\ell im})$. We then find

$$sign\left[ \lim_{\lambda \to \infty} \frac{\partial W(p,e)}{\partial p} \bigg|_{p=c} \right] = sign\left[ X^F + p\frac{dX^F}{dp} \bigg|_{p=c} - c\frac{dX}{dp} \bigg|_{p=c} \right] = sign\big(1 - F(\widehat{z}) - \widehat{z} f(\widehat{z}))\big).$$

If $1 - F(\widehat{z}) - \widehat{z} f(\widehat{z}) < 0$, $p^E_{\ell im} < c$. As $p^E$ increases with $\lambda$, we have proved that $p^E < c$ for all values of $\lambda$.

On the other hand, if $1 - F(\widehat{z}) - \widehat{z} f(\widehat{z}) > 0$, $p^E_{\ell im} > c$. In this case, there exists $\tilde{\lambda}$ such that, for all $\lambda > \tilde{\lambda}$, $p^E > c$.

Note that

$$1 - F(\widehat{z}) - \widehat{z} f(\widehat{z}) < 0 \Longleftrightarrow \frac{1 - F\left(\frac{c}{e^E_{\ell im}}\right)}{\frac{c}{e^E_{\ell im}} f\left(\frac{c}{e^E_{\ell im}}\right)} < 1,$$

As the left-hand side of the second inequality decreases with $\frac{c}{e^E_{\ell im}}$, the worst case scenario is for its lowest value, namely when $e^E_{\ell im} = 1$. So, a sufficient condition that ensures that $p^E_{\ell im} < c$ is

$$\frac{1 - F(c)}{cf(c)} < 1. \tag{59}$$

If we denote by $\widetilde{c}$ the value that satisfies (59) with equality, the sufficient condition boils down to $c > \widetilde{c}$ ■

## Appendix I: Proof of Proposition 4

Consider two cumulative distribution functions $F_1(z)$ and $F_2(z)$, with corresponding densities $f_1(z)$ and $f_2(z)$. Let $F_1(z)$ dominate $F_2(z)$ in the hazard rate order. Let $\widetilde{c}_1$ and $\widetilde{c}_2$ be implicitly defined by $1 - F_1(\widetilde{c}_1) = \widetilde{c}_1 f_1(\widetilde{c}_1)$ and $1 - F_2(\widetilde{c}_2) = \widetilde{c}_2 f_2(\widetilde{c}_2)$, respectively. Next, we show that $\widetilde{c}_1 \geq \widetilde{c}_2$.

Assume that $\widetilde{c}_1 < \widetilde{c}_2$. Using their definition, we obtain

$$\frac{1 - F_1(\widetilde{c}_1)}{\widetilde{c}_1 f_1(\widetilde{c}_1)} = \frac{1 - F_2(\widetilde{c}_2)}{\widetilde{c}_2 f_2(\widetilde{c}_2)},$$

which implies that

$$\frac{\frac{1 - F_1(\widetilde{c}_1)}{f_1(\widetilde{c}_1)}}{\frac{1 - F_2(\widetilde{c}_2)}{f_2(\widetilde{c}_2)}} = \frac{\widetilde{c}_1}{\widetilde{c}_2} < 1. \tag{60}$$

Moreover, as $F_1(z)$ dominates $F_2(z)$ in the hazard rate order, and their corresponding hazard rates increase with $z$ (because both densities are log-concave),

$$1 < \frac{\frac{1 - F_1(\widetilde{c}_1)}{f_1(\widetilde{c}_1)}}{\frac{1 - F_1(\widetilde{c}_2)}{f_1(\widetilde{c}_2)}} < \frac{\frac{1 - F_1(\widetilde{c}_1)}{f_1(\widetilde{c}_1)}}{\frac{1 - F_2(\widetilde{c}_2)}{f_2(\widetilde{c}_2)}}. \tag{61}$$

Combining (60) and (61) yields to a contradiction. Hence, $\widetilde{c}_1 \geq \widetilde{c}_2$ ■

## Appendix J: Proof of Lemma 4

The writing of the maximand and the participation constraints faced by the regulator are identical to those in the previous section. But in this case, as the level of effort has to be induced by the regulator via transfers, it follows that

$$e \in \underset{\tilde{e}}{\operatorname{argmax}} \ \mathbb{E}\left[V_i(p, \tilde{e}, T)\right] \quad (IC_{MH}),$$

where expectations are taken with respect to the likelihood of $\gamma_h$ and $\gamma_\ell$, and $T = (T_h, T_\ell)$.

As we adopt the first-order approach to solve this moral hazard problem, we assume that $IC_{MH}$ is completely characterized by

$$\rho'(e)\left[\left(pX_h^F + T_h - c\cdot(X_h^F + X_h^I)\right) - \left(pX_\ell^F + T_\ell - c\cdot(X_\ell^F + X_\ell^I)\right)\right] - \psi'(e) = 0. \quad (62)$$

If we substract $\psi(e)$ in both terms in curly brackets, and use the definition of the ex-post utilities $V_h, V_\ell$, we obtain

$$\rho'(e)\left[V_h - V_\ell\right] - \psi'(e) = 0 \blacksquare$$

## Appendix K: Proof of Proposition 5

We prove the first statement of the proposition using an argument similar than the one employed before. If

$$p^E \geq \gamma_h \bar{z}, \quad (63)$$

the regulator never faces formal consumers (because $\widehat{z}_h > \bar{z}$), and thus $p^E$ can take any value that satisfies (63). For the sake of simplicity, assume that $p^E = \gamma_h \bar{z}$ and denote by $\mathbb{E}W_0(e)$,

$$\mathbb{E}W_0(e) \equiv y + \mathbb{E}\left[\int_0^{\bar{z}}\int_{\varphi_i}^{\bar{\theta}}[\theta - \gamma_i z - (1+\lambda)c]\,dG(\theta)dF(z)\right] - (1+\lambda)\psi(e) \text{ for } i \in \{h, \ell\} \quad (64)$$

It is straightforward to show that

$$\lim_{p\to\gamma_h\bar{z}} \mathbb{E}W(p,e) = \mathbb{E}W_0(e) \quad (65)$$

Next, for any level of deterrent effort $e$, we compute

$$\lim_{p\to\gamma_h\bar{z}} \frac{\partial \mathbb{E}W(p,e)}{\partial p} = -(1+\lambda)[1 - G(\widehat{\theta})]\bar{z}f(\bar{z}) < 0. \quad (66)$$

Thus, the regulator always sets $p^E < \gamma_h \bar{z}$. The remainder of the proposition is proved by simply differentiation $\blacksquare$

## Appendix L: Proof of Lemma 5

Consider that, at the optimum, $e^E > 0$. Assume first that constraints $VP_\ell$ and $VP_h$ are slack. If so, we can always decrease $V_\ell^E$ and $V_h^E$ by a small $\epsilon > 0$ such that both constraints still hold. Moreover, since the constraint $(IC_{MH})$ is completely characterized by its first-order condition

$$\rho'(e)\big[V_h(p,e) - V_\ell(p,e)\big] - \psi'(e) = 0, \tag{67}$$

then the firm's ex-post utilities in the new regulatory scheme also satisfies $(IC_{MH})$. These reductions in ex-post utilities lead to an increase in the expected welfare, which is a contradiction.

Now assume that $VP_\ell$ and $VP_h$ bind. This implies that (67) becomes $-\psi'(e) < 0$ and so the profit maximizing level of effort is $e^E = 0$. This contradicts the initial assumption that $e^E > 0$.

Finally, assume that constraint $VP_h$ binds, while $VP_\ell$ is slack at the optimum. Again, (67) becomes $-\rho'(e)V_\ell - \psi'(e) < 0$. Thus, the profit maximizing level of effort is $e^E = 0$, which is a contradiction.

The unique alternative that does not lead to a contradiction is the one conjectured in the lemma ■

## Appendix M: Proof of Corollary 3

When the detection technology is random, let's consider the following fictitious program,

$$\max_{p,e} \mathcal{L} \equiv \quad \alpha\, \mathbb{E}W + (1-\alpha)\, \mathbb{E}W^{MH}, \tag{68}$$

where $\mathbb{E}W^{MH}$ and $\mathbb{E}W$ are the expected welfares of the implementable problems (i.e., the problems incorporating the analysis of the binding participation and incentive constraints), with moral hazard and under full information, respectively. Notice that when $\alpha = 0$ ($\alpha = 1$), $\mathcal{L}$ coincides with $\mathbb{E}W^{MH}$ ($\mathbb{E}W$). So we can undertake comparative statics with respect to $\alpha$. Given our differentiability assumptions, we compute the following cross derivatives,

$$\begin{aligned}
&\mathcal{L}_{\alpha p} = 0 \\
&\mathcal{L}_{\alpha e} = \lambda\psi'(e)\left[1 + \frac{\rho(e)}{\rho'(e)}\left(\frac{\psi''(e)}{\psi'(e)} - \frac{\rho''(e)}{\rho'(e)}\right)\right] \geq 0 \\
&\mathcal{L}_{pe} = \alpha\frac{\partial^2\mathbb{E}W}{\partial p \partial e} + (1-\alpha)\frac{\partial^2\mathbb{E}W^{MH}}{\partial p \partial e} = 0
\end{aligned} \tag{69}$$

As the function $\mathcal{L}$ is supermodular with respect to $(p, e, \alpha)$, Topkis' Monotonicity Theorem implies that $p$ and $e$ increase with $\alpha$, which proves the corollary ■

## Appendix N: Optimal regulation when evaders have no social weight

The following proposition compares the optimal pricing rule $(p^*, e^*)$ with the one that obtains when evaders have the same social weight than formal consumers.

**Proposition 7** *When the regulator does not consider evaders' welfare, the optimal price $p^*$ is lower and the level of effort $e^*$ is higher than their corresponding levels when evaders are fully incorporated in the social welfare.*

**Proof.** When evaders' welfare does not count in the total surplus, the social welfare is

$$\widetilde{W}(p,e) = \quad y + \int_{\widehat{z}}^{\overline{z}} \int_{\widehat{\theta}}^{\overline{\theta}} \theta b dG(\theta) dF(z) + \lambda p X^F(p,e) - (1+\lambda)\big(C(X(p,e)) + \psi(e)\big). \tag{70}$$

Let's compute the following derivatives

$$\left.\frac{\partial \widetilde{W}}{\partial p}\right|_{\substack{p=p^E,\\ e=e^E}} = -(1+\lambda) g(\widehat{\theta}) \frac{\partial \widehat{\theta}}{\partial p}[1-F(\widehat{z})](p^E - c) - f(\widehat{z}) \frac{\partial \widehat{z}}{\partial p}[1-G(\widehat{\theta})]\Pi + \lambda X^F, \tag{71}$$

$$\left.\frac{\partial \widetilde{W}}{\partial e}\right|_{\substack{p=p^E,\\ e=e^E}} = -f(\widehat{z}) \frac{\partial \widehat{z}}{\partial e}[1-G(\widehat{\theta})]\Pi + \int_0^{\widehat{z}} f(z) g(\varphi) \frac{\partial \varphi}{\partial e} dz.(1+\lambda)c = -(1+\lambda)\psi'(e^E), \tag{72}$$

where $\Pi = \left[ \lambda p^E + \frac{\int_{\widehat{\theta}}^{\overline{\theta}} \theta b dG(\theta)}{1-G(\widehat{\theta})} \right]$.

By definition of $\widehat{\theta}$,

$$\Pi = \lambda p^E + \frac{\int_{\widehat{\theta}}^{\overline{\theta}} \theta b dG(\theta)}{1-G(\widehat{\theta})} = (1+\lambda) p^E + \frac{\int_{\widehat{\theta}}^{\overline{\theta}} (\theta b - p^E) dG(\theta)}{1-G(\widehat{\theta})} > (1+\lambda) p^E.$$

Thus, comparing (71) with the first order condition (8), we conclude that

$$\left.\frac{\partial \widetilde{W}}{\partial p}\right|_{\substack{p=p^E,\\ e=e^E}} < 0.$$

Similarly,

$$-f(\widehat{z})\frac{\partial\widehat{z}}{\partial e}\left[1-G(\widehat{\theta})\right]\Pi+\int_0^{\widehat{z}}f(z)g(\varphi)\frac{\partial\varphi}{\partial e}dz.(1+\lambda)c$$

$$=-f(\widehat{z})\frac{\partial\widehat{z}}{\partial e}\left[1-G(\widehat{\theta})\right](1+\lambda)p^E-f(\widehat{z})\frac{\partial\widehat{z}}{\partial e}\left[1-G(\widehat{\theta})\right]\left[\frac{\int_{\widehat{\theta}}^{\overline{\theta}}(\theta b-p^E)dG(\theta)}{1-G(\widehat{\theta})}\right]$$

$$+\int_0^{\widehat{z}}f(z)g(\varphi)\frac{\partial\varphi}{\partial e}dz.(1+\lambda)c$$

$$>-f(\widehat{z})\frac{\partial\widehat{z}}{\partial e}\left[1-G(\widehat{\theta})\right](1+\lambda)p^E-\int_0^{\widehat{z}}zf(z)\left[1-G(\varphi)\right]dz+\int_0^{\widehat{z}}f(z)g(\varphi)\frac{\partial\varphi}{\partial e}dz.(1+\lambda)c$$

because $\frac{\partial\widehat{z}}{\partial e}<0$. Again, comparing (72) with the first order condition (9), we conclude that

$$\left.\frac{\partial\widetilde{W}}{\partial e}\right|_{\substack{p=p^E,\\ e=e^E}}>0\ \blacksquare$$