

Optimal regulation of network expansion*

Bert Willems[†] and Gijsbert Zwart[‡]

September 7, 2017

Abstract

We model the optimal regulation of continuous, irreversible, capacity expansion, where a regulated firm has private information about capacity costs, investments are financed from the firm's cash flows, and demand is stochastic. The optimal mechanism can be implemented as a revenue tax that increases with the level of the price cap. If the asymmetric information has large support, then the optimal mechanism consists of a laissez-faire regime for low-cost firms. That is, the firm's price cap corresponds to that of an unregulated monopolist, and it is not taxed. This 'maximal distortion at the top' is necessary to provide information rents, as direct subsidies are not feasible.

Keywords: regulation, real option value, asymmetric information, optimal contracts

JEL classification: D81, D82, L51

*We have benefited from useful comments from seminar participants at Chulalongkorn University, June 2017, Ecoles de Mines de Paris, March 2013, Norwegian University of Science and Technology, June 2016, Thammasat University, June 2017, Tilburg University, April 2014, Toulouse University, January 2017, UCL Louvain La Neuve, December 2016, University of Basel, February 2017, the Symposium in Honour of Jean Tirole, The Hague, December 2014, the IFN Conference on The Performance of Electricity Markets, Stockholm, July 2014, the IAEE conference, Singapore, June 2017, and from discussions with Peter Broer, Claude Crampes, Steffen Hoernig, Bruno Julien, Thomas-Olivier Léautier and Thomas Tangerås. We also thank two anonymous referees and Mark Armstrong, the editor, for their insightful comments.

[†]Tilburg University, Tilec, Center and Toulouse School of Economics; bwillems@tilburguniversity.edu.

[‡]University of Groningen, SOM and Tilburg University, Tilec; g.t.j.zwart@rug.nl.

1 Introduction

Since the nineties many regulated network industries switched from cost-plus (or rate-of-return) to incentive regulation, often under some form of price cap regulation. This switch was motivated by the fear that cost-plus regulated firms would “gold-plate” their networks and over-invest in capital (Averch and Johnson, 1962) and the realization that a price cap provides high-powered incentives for cost-efficiency (Cabral and Riordan, 1989).¹ However, in recent years stakeholders have argued that with high powered incentive regulation, firms postpone socially efficient investments in durable assets, especially in risky environments, and that a different form of regulation is necessary. For instance, in response to large investment needs, the UK electricity and gas regulator OFGEM modernized its price cap mechanism by explicitly taking into account these investment needs.² European energy directives allow specific network investments to be exempted from regulation in order to foster investments if uncertainty is large.³ For the telecom sector, ETNO, the industry association representing European telecom operators, recommends relaxing access regulation, as it sees it as the main reason for European infrastructure investments lagging those in the U.S. (Williamson, Lewin and Wood, 2016).⁴ Also academic scholars recognize that implementing price cap regulation is challenging for durable investments and when uncertainty is important (Guthrie, 2006; Armstrong and Sappington, 2007).

In this article we contribute to this debate by studying the optimal regulation of capacity investments in a dynamic setting in which investment prospects are uncertain. For this we consider a regulated private firm that has to gradually expand its network to cope with a growth in demand for network access, needs to fund its investments from operating profits and has superior information on investments costs. The regulator contracts with the firm about when it should expand capacity (and when it would be better to delay), at which price the capacity should be sold, and which fraction of its revenues it may keep. The regulator acts as a social planner and maximizes the expected discounted sum of consumers’ surplus and the firm’s profit.

In the optimal regulatory mechanism existing capacity is always used efficiently: prices for network access are equal to the short-run marginal cost of transportation as long as there is spare capacity, and prices are above marginal cost when there is congestion. Capacity is expanded, whenever the price for capacity reaches a threshold value. This price threshold increases with investment costs, and is always higher than under the first best symmetric information optimum with demand uncertainty. Hence, investments are delayed.⁵

As we assume that the firm does not receive subsidies, investment costs need to be paid from market revenues. However, any operating profits that remain after those costs have been paid for, could be taxed by the regulator. Under optimal regulation the regulator does *not* tax the firms which reveal to be relatively efficient, whereas the

¹See for instance Sappington (2002) for an overview of the perceived drawbacks of rate of return regulation.

²The UK was one of the first countries to introduce the RPI-X price cap model (Beesley and Littlechild, 1989). After a review, a new set of regulatory principles was introduced, the RIIO-model. This is still a form of price regulation but includes output obligations and additional funds for experimentation (OFGEM, 2010*b*).

³Exemptions can be granted if among others “the level of risk attached to the investment is such that the investment would not take place unless an exemption was granted”. (Regulation EC 714/2009, Art.17 & Directive 2009/73/EC, Art. 36). Until 2015 there were 35 exemption requests, most of which were (partially) granted.

⁴In contrast to the energy sector, the European Commission does not allow regulatory exemptions for the telecommunication sector and it successfully appealed the decision of the German government to grant Deutsche Telekom an exemption. (Commission v. Germany 2009, case number C-424/07, the European Court of Justice.)

⁵Note that the first best investment expansion plan already delays investments to take into account the real option value of network expansion.

tax rate for the less efficient firms increases with their levels of inefficiency. Those low tax rates are necessary to provide the efficient firms with information rents.

If the information asymmetry between the regulator and the firm has large support, then the relatively efficient firms will be allowed to invest as if they were unregulated monopolists, as this provides the largest possible information rents. Hence, a laissez-faire regime is optimal for those firms. In the case with small support, the regulator will bunch the more efficient firms and require identical investment levels for these. Hence, optimal regulation no longer results in an equilibrium with full separation of types.

In our model demand growth is not fully predictable (i.e. stochastic) and network investments are sunk. Hence the firm is continuously forecasting demand and balancing the benefits of expanding capacity now (and obtaining additional revenue) and delaying investments (and obtaining superior information about future demand). In other words, it needs to take into account the real option value of investments (Dixit and Pindyck, 1994). McDonald and Siegel (1986) show that an unregulated monopolist delays investments under uncertainty, and Pindyck (1988) extends this result to a continuous investment model. Although also first-best investment involves a delay, under monopoly, this delay is longer. If a regulator would try to correct this situation with only the price cap instrument at its disposal, then the first best outcome cannot be reached (Dobbs, 2004), as one instrument is used for two goals: efficient investments ex-ante and optimal consumption ex-post. Building on Dobbs, but introducing scale economics for capacity expansion – in which case grouping investments across time is cost efficient – Evans and Guthrie (2012) show that the price cap should be lowered and that it might be efficient to allow some demand rationing to increase the size of subsequent expansions. Roques and Savva (2009) extend Dobbs’ model to a Cournot duopoly with a price cap. Our article also starts from Dobbs’ model but includes asymmetric information, a self-financing constraint and assumes that the regulator has additional instruments to enforce investments. As we assume constant returns to scale in capacity expansion, it is never optimal to group investments or to ration demand, in contrast to Evans and Guthrie (2012).

In order to model the interaction between the regulator and the firm, we rely on the assumption that the firm has superior information about its own investment costs as in the seminal article by Baron and Myerson (1982). Whereas most adverse selection types of models allow for lump-sum transfers to the agent, we impose a self-financing constraint which limits transfers. In Baron and Myerson, the most efficient firm invests at the efficient level, i.e. there is “no distortion at the top”, and gains information rents by receiving a large lump-sum transfer. In our model, information rents can only be obtained by being more profitable in the market, and hence investment levels need to be distorted away from the efficient level. In fact, we find that if the maximum cost level is large, compared to the most efficient firm’s cost, monopoly level investments are optimal for the most efficient firms. So we find what could be called “maximal distortion at the top”. If information asymmetries are small, we find a bunching equilibrium for the most efficient firms. Our model differs in a number of ways from Baron and Myerson. (1) Instead of choosing production output, the firm invests in additional capacity.⁶ (2) The objective of the regulator gives the same weight to consumers’ surplus and the firm’s rents. Instead, the cost of leaving information rents is endogenous, arising from a self-financing constraint. (3) We consider a

⁶Constant elasticity demand in our model ensures that all capacity is subsequently used.

multi-period setting.⁷

We allow the regulator to tax, but not to make subsidies to the firm. This reflects the fact regulators are often legally forbidden to subsidize firms.^{8,9} In our model, we allow for one-way state-contingent payments from the firm to the regulator. This is in contrast with the delegation literature in which any transfer is ruled out, and the regulator decides only on the amount of discretion for the agent.¹⁰ Our results are similar to the principal agent model of Gautier and Mitra (2006). They assume that the regulator can provide lump-sum transfers but is limited by its budget constraint, which is determined exogenously. In our model, the transfers that the firm can receive are determined endogenously by its investment decisions and the revenues those investments generate.

One strand of the literature on incentive regulation and durable investments highlights the lack of commitment by the regulator.¹¹ If a regulator cannot commit to a price level for a sufficiently long period, it will lower prices once investment has taken place as those prices are ex-post efficient. This will lead to hold-up and lower investments ex-ante. In order to address this commitment problem Gans and King (2004) propose a *regulatory holiday* in which the regulator commits not to regulate prices for a *limited duration*, under the implicit assumption that it is easier to commit not to regulate than to commit to a high regulated price. We find that even if the regulator has full commitment power, it might be optimal to provide an exemption from regulation to the most efficient firms, as this provides information rents for firms who want to invest earlier and with larger quantities. Note that the regulatory exemption in our model is not of a limited duration and not unconditional. If a firm invests too late, then a fraction of its operating profits should be taxed. We assume the regulator fully commits to a *long-term, non-renegotiable* contract.¹²

Many network industries are characterized by features similar to the ones of our model. They have capital intensive networks with relatively long-lived assets. Examples include the local-loop in telecommunications markets, low voltage distribution and high voltage transmission networks in power markets. Growing demand by network users both in volume and service quality require continuous upgrades and expansions of switches in local central offices, voltage transformers and new communication equipment in power networks.¹³ Recent

⁷Baron and Besanko (1984) extend Baron and Myerson (1982) to a multi-period setting in which the firm's types might be correlated across time periods. In the extreme, with perfect correlation, as in our model, output is distorted to the same extent for all time periods.

⁸The lack of subsidies can also be motivated by regulatory collusion and commitment problems (Laffont and Tirole, 1993, 1991) or to prevent regulatory competition, and justifies much of the literature on Ramsey pricing. Armstrong and Sappington (2007) indicate that normative models on optimal contracting have limited real-world applications as, among others, transfers to the firms are assumed to be feasible.

⁹European state-aid guidelines require prior authorization of subsidies for energy infrastructure investments, *i.a.* to prevent indirect subsidies to the energy intensive industry. Only when they correct for positive externalities they are allowed (OCJ/C 200/1).

¹⁰If the firm's and regulator's preferences are sufficiently aligned, then the optimal delegation set is a closed interval of decisions that the firm might choose from (Alonso and Matouschek, 2008). If not, the set contains gaps: the firm is forbidden to take intermediate actions. In a Baron and Myerson model without transfers, unimodally distributed costs and linear or constant elastic demand, the regulator cannot do better than a price cap.

¹¹Another strand of literature discusses practical challenges for price cap regulation and long term investments. Joskow (2008) reports problems with measuring cost of capital and setting prices at the end of a regulatory period. In a dynamic Ramsey pricing model, Evans and Guthrie (2005) show that prices should not be based on the ex-post optimized replacement value of assets.

¹²*Renegotiation* generally increases agency costs, as efficient firms require more information rents early on (Bester and Strausz, 2001; Hart and Tirole, 1988; Laffont and Tirole, 1990; Battaglini, 2007). With *Repeated short-term contracts* the agent faces ex-post hold-up, which is costly to compensate ex-ante (Laffont and Tirole, 1988), unless the mutual threat of punishment restores incentives (Salant and Woroch, 1992; Gilbert and Newbery, 1994).

¹³Instead of a growth of demand volume, a different interpretation would be that demand for quality increases over time. Equivalently, we could assume constant demand and network capacity which degrades over time and where the firm builds replacement capacity when necessary. Mathematically this model would be very similar.

technological changes have put those investments requirements to the forefront. Video on demand and cloud computing creates additional pressure on telecommunication networks, the large scale introduction of renewable energy and decentralized production requires substantial upgrade of power networks. Many network firms have been privatized and governments are not keen on subsidizing investments.

The characterization of the telecom sector by the European association of telecom operators highlights many of our assumptions:

The bulk of the investment required to meet policy objectives for the Digital Single Market will need to come from private investment in Europe's access networks. This private investment is a continuous and incremental, rather than a one-off, process. Investment decisions are constrained by the annual cash flows generated by the businesses.... Market players are better placed to make efficient investment decisions than NRAs [National Regulatory Authorities] or governments. They have far more information on both the incremental costs of deploying new technologies and the incremental revenues which might flow from investing (Williamson, Lewin and Wood 2016, Pg.2-3).

In our model we assume that the regulator not only regulates the firm's revenue (by setting which fraction of operating profit the firm is allowed to keep), but also enforces the required investments levels. Such dual requirements are also found in practice. For instance under the new RIIO regulatory model for the UK energy markets, the regulator not only specifies a certain price level, but also agrees on specific output parameters. If firms are unable to reach those output targets they will lose their operating license (OFGEM, 2010a).

We also show that the optimal mechanisms can be implemented as a revenue tax, which depends positively on the maximal amount of congestion which is reflected in high scarcity prices for bottleneck capacity. In practice those congestion prices might not be directly observable or very stochastic. However, often several proxies for congestion levels can be relied on as a basis for regulation (packages lost in telecom networks, redispatch cost in electricity markets).¹⁴

In our article we do not consider one-off (lumpy) investments. Moreover, we consider information asymmetry on a static parameter (costs) and assume that the information asymmetry is not related to the stochastic demand realization, although those features might sometimes be present in practice. We refer to two companion articles for those aspects of regulation: Broer and Zwart (2013) assume that investments are lumpy and only occur once; Arve and Zwart (2014) assume asymmetric information with respect to stochastic parameters instead of static ones, and allow for lump sum transfers to the firm. These models do not exhibit the pooling and monopoly pricing regimes arising from the budget constraint that is central in this article.

Finally, our regulated firm both *builds* and *operates* the network under a *long-term* procurement contract and it faces *substantial risk*. Its regulatory contract therefore has the characteristics of a Private-Public-Partnership (Iossa and Martimort, 2015). Our article complements this literature by highlighting the downsides of having such projects funded by user fees alone.¹⁵

¹⁴See for instance Lesieutre and Eto (2003) for a discussion on measuring congestion in electricity networks.

¹⁵We do not study how agency costs can be reduced by bundling tasks and by relying on private funding. *Bundling investment and operation* improves the agent's incentives to internalize externalities across tasks due to economies of scope in agency costs

2 Model

We consider a continuous-time, continuous investment model of a principal, the regulator, contracting with a monopolist to make irreversible investments $dQ(t)$ to expand network capacity $Q(t)$, as in Dobbs (2004). Capacity is a continuous variable, and capacity expansions come at a constant marginal cost c . There is no depreciation of capacity.¹⁶ Initial capacity at $t = 0$, the time of contracting, is $Q(0) = 0$, i.e. we consider a greenfield project.¹⁷ Marginal cost c is drawn from a cumulative distribution $F(c)$ with full support $[c_L, c_H]$, and density $f(c) > 0$. A sufficient condition that the solution of the first-order conditions actually corresponds to a maximum, is that the density is downward sloping $f'(c) \leq 0$ on its support $[c_L, c_H]$. We will assume this.¹⁸

At each moment in time, the capacity $Q(t)$ is sold to users at a price $p(t)$ (we will drop the t -dependence of price and other variables in subsequent notation).

The demand for network capacity has constant elasticity,

$$p = AQ^{-\gamma} \quad (1)$$

where $0 < \gamma < 1$ is the inverse of demand elasticity. The associated flow of gross consumer surplus from using Q is then $AQ^{1-\gamma}/(1-\gamma)$.¹⁹

The demand shift parameter A is stochastic, and satisfies a geometrical Brownian motion,

$$dA = \mu A dt + \sigma A dz, \quad (2)$$

where μ and $\sigma > 0$ are the associated drift and volatility parameters. As A grows over time, demand for capacity will increase, making capacity investment more valuable. Let A_0 be the demand shift $A(0)$ at $t = 0$. We will assume that demand A and capacity Q are observable and verifiable by the regulator, but the realization of investment cost c is private information.

Irreversible investments, $dQ(t) \geq 0$, at time t , will be governed by an investment rule, which will be part of the contract between the regulator and the monopolist. Such an investment rule will specify that capacity is to be expanded as soon as demand reaches some threshold value. We denote by $\bar{A}(Q)$ the threshold value for A , given Q , at which investment occurs.²⁰

The principal's objective is to maximize expected total welfare which is the difference of gross consumers' (Martimort and Pouyet, 2008; Iossa and Martimort, 2012), or because reallocating property rights reduces hold-up (Hart, 2003; Bennett and Iossa, 2006). In our model *operational costs* are common knowledge and normalized to zero. *Private funding* can lower agency costs if investors have superior information (Iossa and Martimort, 2012, 2015) or improve project selection (Maskin and Tirole, 2008; de Bettignies and Ross, 2009).

¹⁶It would be a straightforward extension of the model to assume a constant depreciation rate.

¹⁷Alternatively, we demand that the regulator also needs to remunerate the monopolist for its existing investments in efficient capacity.

¹⁸Below we will show that our condition will only be relevant for costs above a certain (endogenously determined) threshold level. If local first order conditions are not sufficient, then the optimal contract will need additional ironing as in Guesnerie and Laffont (1984).

¹⁹Note that with such constant elasticity demand, it is always optimal to use all available capacity.

²⁰Note that because there is no exogenous dependence on time other than through A , the optimal policy, as well as total welfare, cannot explicitly depend on time t .

surplus and investment costs. The flow of total welfare, given an investment threshold $\bar{A}(Q)$ and costs c , equals

$$U(A(t), Q(t)|\bar{A}, c) dt = \frac{A(t)Q(t)^{1-\gamma}}{1-\gamma} dt - c dQ(t). \quad (3)$$

The continuation value of total expected welfare, then equals the expected discounted sum of these welfare flows,

$$W(A(t), Q(t)|\bar{A}, c) = \mathbb{E}_A \left[\int_t^\infty e^{-r(\tau-t)} U(A(\tau), Q_{\bar{A}}(\tau), c) d\tau \right] \quad (4)$$

where \mathbb{E} denotes the expectation over the future demand shock paths $A(t)$. $Q_{\bar{A}}(\tau)$ for $\tau > t$ represents the future capacity path, which will be determined by the evolution of A and the investment threshold $\bar{A}(Q)$. r is the risk-free rate, and we assume $r > \mu$.

The regulator contracts with the monopolist to achieve optimal investment. A contract specifies the investment rule $\bar{A}(Q)$, as well as a monetary transfer to the agent that remunerates him.

We focus on a regulatory contract that is written at time $t = 0$, when demand $A = A_0$, for a greenfield investment, $Q = 0$. Let $W_0(\bar{A}, c)$ be the expected welfare of a greenfield investment according to investment rule \bar{A} , at expansion cost c :

$$W_0(\bar{A}, c) \equiv W(A_0, 0|\bar{A}, c).$$

The regulator designs a menu of contracts to maximize expected welfare of greenfield investments across all types c ,

$$\mathcal{W} = \int_{c_L}^{c_H} W_0(\bar{A}, c) dF(c).$$

where the chosen investment rule will typically vary for the different cost types.

When choosing from the regulatory menu, the monopolist's goal is profit maximization. Profits are determined by the expected present value of the total remunerations minus costs of capacity expansion. We assume these remunerations to be payments from the regulator to the firm, financed out of the proceeds of the sale of capacity. We denote total expected present value of future remunerations promised to the monopolist at time t , with current state $(A(t), Q(t))$, by $T(A(t), Q(t))$.

We can then state the following ongoing budget constraints: under a given investment rule $\bar{A}(Q)$, total expected future remunerations T cannot exceed the total expected proceeds from the capacity sale,

$$T(A(t), Q(t)) \leq \mathbb{E}_A \left[\int_{\tau=t}^\infty e^{-r(\tau-t)} p(A(\tau), Q_{\bar{A}}(\tau)) Q_{\bar{A}}(\tau) d\tau \right]. \quad (5)$$

In addition, we impose the ongoing participation constraint on the monopolist that total expected profits should be non-negative at any time period, for any given cost c :

$$\Pi(A(t), Q(t)|\bar{A}, c) = T(A(t), Q(t)) - \mathbb{E}_A \left[\int_{\tau=t}^\infty e^{-r(\tau-t)} c dQ_{\bar{A}}(\tau) \right] \geq 0. \quad (6)$$

3 First-best and monopoly benchmarks

As a benchmark, we first explore the first-best outcome (as analyzed in Pindyck, 1988). In the absence of asymmetric information, the principal sets $\bar{A}(Q)$, the threshold value for A given Q , at which investment occurs to optimize total continuation welfare W in equation (4). The standard method of solving for W is first to note that in the region $A < \bar{A}(Q)$ where no investment occurs, W satisfies a Bellman equation (Dixit and Pindyck, 1994) (See figure 1),

$$rW = \frac{AQ^{1-\gamma}}{1-\gamma} + \mu A \frac{\partial W}{\partial A} + \frac{1}{2}\sigma^2 A^2 \frac{\partial^2 W}{\partial A^2}. \quad (7)$$

Imposing the boundary condition that W vanishes when $A \rightarrow 0$, the general solution to this differential equation takes the form,

$$W(A, Q|\cdot) = \frac{AQ^{1-\gamma}}{(1-\gamma)(r-\mu)} + g(Q)A^\lambda, \quad (8)$$

where $g(Q)$ is any function of Q , and λ is the positive solution to the fundamental quadratic $r = \mu\lambda + \frac{1}{2}\sigma^2\lambda(\lambda-1)$. In this expression, the first term represents the expected present value from using existing capacity Q (without any future expansions), whereas the second term is the value of the option to expand capacity beyond its current level if demand rises.

Next, we solve for $g(Q)$ by imposing the boundary condition at the point of investment $\bar{A}(Q)$, that the marginal benefit of increasing Q should equal the marginal cost of investment,

$$\frac{\partial W}{\partial Q}(\bar{A}(Q), Q|\cdot) = c.$$

Substituting for W , we find a condition on the derivative of $g(Q)$,

$$\frac{\partial g(Q)}{\partial Q} = \bar{A}(Q)^{-\lambda} \left(c - \frac{\bar{A}(Q)Q^{-\gamma}}{r-\mu} \right).$$

We impose that as Q goes to infinity, there is no longer any (option) value to further investment ($g(Q) \rightarrow 0$), to find

$$g(Q) = \int_Q^\infty \bar{A}(q)^{-\lambda} \left(\frac{\bar{A}(q)q^{-\gamma}}{r-\mu} - c \right) dq, \quad (9)$$

which specifies, jointly with equation (8), total welfare W , given an investment threshold $\bar{A}(Q)$. Note that welfare W is an increasing function of $g(Q)$. The optimal investment threshold then follows from point-wise maximization of the integrand, and is given by

$$\bar{A}^c(Q)Q^{-\gamma} = \frac{\lambda}{\lambda-1}(r-\mu)c \equiv \bar{p}^c(c), \quad (10)$$

or in other words, investing whenever price p reaches the level \bar{p}^c .²¹ Superscript c refers to the ‘‘competitive’’ benchmark. The price at which optimal investment occurs exceeds the annualized costs by the factor $\lambda/(\lambda-1) >$

²¹The fact that optimal investment occurs at a fixed price level, independent of the level of the demand shock A , is related to the demand function being of the form $Q = AD(p)$ for some function D . Elasticity of demand at a given price level is then independent of A .

1, which itself depends on the parameters of the stochastic process and in particular grows as volatility σ increases. This is a reflection of the well-known option value of delaying investment (McDonald and Siegel, 1986).

At $t = 0$, when capacity $Q = 0$, there will be a one-off investment $Q_0^c = \left(\frac{A_0}{\bar{p}^c}\right)^{1/\gamma}$ to bring initial price to the threshold price. Hence, total welfare at $t = 0$ is equal to

$$W_0(\bar{A}^c, c) = W(A_0, Q_0^c | \bar{A}^c, c) - cQ_0^c = \frac{\gamma}{1-\gamma} \frac{\gamma\lambda}{\gamma\lambda-1} cQ_0^c.$$

[Figure 1 about here.]

As a second benchmark, it will be relevant to consider the investment rule that an unregulated monopolist receiving all revenues from selling capacity would choose. Define the firm's profit flow as

$$\pi(A(t), Q(t) | \bar{A}, c) dt = p(A(t), Q(t))Q(t) dt - c dQ(t) = A(t)Q(t)^{1-\gamma} dt - c dQ(t), \quad (11)$$

and the associated total expected continuation value of the firm as

$$V(A(t), Q(t) | \bar{A}, c) = \mathbb{E}_A \left[\int_t^\infty e^{-r(\tau-t)} \pi(A(\tau), Q_{\bar{A}}(\tau) | \bar{A}, c) d\tau \right]. \quad (12)$$

The unregulated monopolist will then choose an investment threshold $\bar{A}(Q)$ that maximizes that expected value. The analysis is similar to the total welfare maximization, with firm value taking the form

$$V(A(t), Q(t) | \cdot) = \frac{A(t)Q(t)^{1-\gamma}}{r-\mu} + A(t)^\lambda \int_Q^\infty \bar{A}(q)^{-\lambda} \left(\frac{\bar{A}(q)q^{-\gamma}(1-\gamma)}{r-\mu} - c \right) dq, \quad (13)$$

analogously to expressions (8, 9) for total welfare. We can again use point-wise maximization to find the profit maximizing investment policy. This is to invest as soon as prices rise to the monopoly price level

$$\bar{p}^m(c) = \frac{\lambda}{\lambda-1} \frac{(r-\mu)c}{1-\gamma}.$$

This expression differs from the welfare optimizing price \bar{p}^c in (10) by the $1-\gamma$ factor, representing the standard Lerner markup, $(\bar{p}^m - \bar{p}^c)/\bar{p}^m = \gamma$.

Finally, it is useful to evaluate total firm value V under the first-best investment rule, *i.e.* invest whenever price reaches \bar{p}^c . Substituting the corresponding threshold $\bar{A}^c(Q)$ in the firm's value function (13), we find that the firm just breaks even, including the costs of the initial investment to bring capacity from $Q = 0$ to a level consistent with the threshold price. In other words, with symmetric information on costs, the regulator can ask the firm to invest according to the first-best rule, and remunerate it using the proceeds of the capacity sales, hence satisfying both the ongoing budget constraint (5), and the firm's participation constraint (6) at $t = 0$. It is not difficult to verify that also the firm's ongoing participation constraint holds in this case; we will check that more generally in the case with adverse selection in section 4. We summarize these benchmark results as

Proposition 1. *Compared to a welfare optimizing social planner, a monopolist delays investment in capacity. That is, it waits until demand has risen to higher levels before investing. Threshold prices that trigger investment are*

$$\bar{p}^m = \frac{\lambda}{\lambda - 1} \frac{r - \mu}{1 - \gamma} c, \quad (14)$$

$$\bar{p}^c = \frac{\lambda}{\lambda - 1} (r - \mu) c. \quad (15)$$

for the monopolist and the social planner respectively. When a greenfield firm invests at the competitive threshold \bar{p}^c , total expected revenues from selling capacity at market clearing prices equal total costs,

$$V(A_0, 0 | \bar{A}^c, c) = 0. \quad (16)$$

4 Optimal regulation under adverse selection

In this section we turn our attention to regulation with asymmetric information on the firm's capacity expansion cost c . We consider the regulator offering the firm a menu of contracts for a capacity expansion schedule that may depend on demand (or price) realizations, which are observable and contractable. In return, the regulator offers a transfer fee T to the firm. The fee has to be financed out of the expected revenues of the capacity sale and therefore has to satisfy budget constraint (5). Also, the firm should earn a non-negative profit, as reflected by participation constraint (6). We shall first focus on the budget and participation constraints at time $t = 0$, and hence only consider the $t = 0$ expected value of all future transfers to the firm, $T_0 = T(A_0, 0)$. At the end of this section we shall see that we can structure the timing of the fee such that also the ongoing budget and participation constraints are met. The regulator maximizes expected total welfare, which is the difference between gross consumers' surplus and investment costs.

Without asymmetric information on costs, we saw in section 3 that the regulator can achieve the first-best investment levels (invest when prices reach threshold price \bar{p}_c), while respecting budget and participation constraints. With private information on costs, the contracts offered will need to respect incentive compatibility as well, and therefore will need to leave information rents to the firms. In view of the budget constraint, distorting the contracts from the first-best scheme is now optimal.

In analyzing the optimal scheme, we follow the standard procedure in optimal contract design, and focus without loss of generality on direct revelation, incentive compatible mechanisms. The regulator offers a menu of contracts to the firm, consisting of a pairs of transfers and investment thresholds $\{(T_0, \bar{A}(Q))\}(\hat{c})$, that depend on the firm's reported cost \hat{c} . We denote the components of these contracts by $T_0(\hat{c})$ and $\bar{A}(Q, \hat{c})$. The firm, by reporting costs \hat{c} , chooses the best option from this menu, which we design such that truthful reporting, $\hat{c} = c$, is optimal.

We saw that both first best and profit maximization require investing when prices reach threshold level, \bar{p}^c and \bar{p}^m respectively, where those thresholds did not depend on capacity. Here, for expositional simplicity, we

assume likewise that the optimal threshold $\bar{A}(Q, \hat{c})$ under adverse selection takes such a form, i.e. there will exist a threshold price $\bar{p}(\hat{c})$ such that $\bar{A}(Q, \hat{c})Q^{-\gamma} = \bar{p}(\hat{c})$. In the appendix we demonstrate that this is indeed optimal. We can then equivalently express the menu of contracts as a set of transfer fees and threshold prices, $\{(T_0(\hat{c}), \bar{p}(\hat{c}))\}$.

At current demand level A_0 , a contract specifying price threshold $\bar{p}(\hat{c})$ will require a greenfield firm to immediately invest capacity $Q_0 = (\frac{A_0}{\bar{p}(\hat{c})})^{1/\gamma}$ to make current price equal the threshold price, so that $A_0 = \bar{A}(Q_0, \hat{c})$. In addition the firm will have to increase capacity as demand A grows to ensure prices remain below the threshold. Total rents R of accepting the contract $(T_0(\hat{c}), \bar{p}(\hat{c}))$, for a firm with actual costs c , are then given by

$$\begin{aligned} R(A_0, c, \hat{c}) &= T_0(\hat{c}) - cQ_0(\bar{p}(\hat{c})) - A_0^\lambda \int_{Q_0(\bar{p}(\hat{c}))}^{\infty} \bar{A}(q, \hat{c})^{-\lambda} c \, dq \\ &= T_0(\hat{c}) - cQ_0(\bar{p}(\hat{c})) \frac{\gamma\lambda}{\gamma\lambda - 1}. \end{aligned} \quad (17)$$

In the first line, we used the expected present value of the costs of future expansions, computed analogously to the cost component of the firm's continuation value V from (13). In the second line, we substituted $\bar{A}(Q, \hat{c}) = \bar{p}(\hat{c})Q^\gamma$, to evaluate the integral.

Incentive compatibility now requires that the firm optimizes this value if it truthfully reveals its costs, $\hat{c} = c$, choosing fee and threshold $(T_0(c), \bar{p}(c))$ from the menu of contracts. Writing the resulting profits from this optimization as $\Pi(c) = R(A_0, c, \hat{c} = c)$, we derive the following necessary conditions for incentive compatibility.

Lemma 1. *Incentive compatibility requires that total greenfield profits Π vary with costs c as*

$$\frac{d\Pi}{dc} = -Q_0(\bar{p}(c)) \frac{\gamma\lambda}{\gamma\lambda - 1}, \quad (18)$$

and that the investment price threshold $\bar{p}(c)$ is non-decreasing in costs c .

We will analyze the welfare optimizing choice of contracts under incentive compatibility constraint (18), as well as the $t = 0$ budget and participation constraints, and ignore the monotonicity requirement on the threshold for the moment. After finding an optimal threshold, we will verify that monotonicity indeed holds. As a last step, we look at the structure of the fees across time and verify that also ongoing constraints can be met.

As a first step let us write the welfare function and the budget constraint in terms of the threshold price \bar{p} . The regulator's objective is to maximize the total welfare, averaged over the possible realizations of costs c . For a given cost c and threshold price \bar{p} , welfare includes the costs of a one-off lumpy investment Q_0 to bring price to the threshold at the current value of the demand shift A , the expected welfare generated by this investment Q_0 , as well as the expected additional welfare from future network expansions (i.e. real option value). From

equation (8),

$$\begin{aligned}
W_0(\bar{p}, c) &= -cQ_0 + \frac{A_0Q_0^{1-\gamma}}{(1-\gamma)(r-\mu)} + A_0^\lambda \int_{Q_0}^\infty \bar{A}(q, c)^{-\lambda} \left(\frac{\bar{A}(q, c)q^{-\gamma}}{r-\mu} - c \right) dq, \\
&= \underbrace{\left(\frac{\gamma\lambda}{\gamma\lambda-1} \right) \left(\frac{\lambda-1}{\lambda} \right) \frac{\bar{p}}{r-\mu} \frac{Q_0}{1-\gamma}}_{\text{Cons Surplus}(\bar{p})} - \underbrace{\left(\frac{\gamma\lambda}{\gamma\lambda-1} \right) cQ_0}_{\text{Cost}(c, \bar{p})} \\
&= \frac{\gamma\lambda}{\gamma\lambda-1} Q_0 \left(\frac{\bar{p}}{r-\mu} \frac{\lambda-1}{\lambda} \frac{1}{1-\gamma} - c \right), \tag{19}
\end{aligned}$$

where again $Q_0 = (\frac{A_0}{\bar{p}})^{1/\gamma}$ and we used $\bar{p} = \bar{A}(Q)Q^{-\gamma}$.

At this point, it is useful to note that in the absence of the budget constraint, the regulator can achieve the first-best outcome as in Loeb and Magat (1979), by setting competitive threshold prices $\bar{p}(c) = \bar{p}^c(c)$, allowing the monopolist to keep all consumer payments, and providing an additional subsidy equal to the monopolist's information rents

$$\frac{\gamma\lambda}{\gamma\lambda-1} \int_c^{c_H} Q_0(\bar{p}^c(c')) dc',$$

so as to satisfy incentive compatibility (18).²² We now proceed to explore the optimal regulation in the presence of the budget constraint.²³

The budget constraint is that for any cost c , total profits, Π , cannot exceed total revenues minus costs. Using the expression for continuation value of revenues and costs (equation 13) and again substituting \bar{p} , we can write this as

$$\begin{aligned}
\Pi(c) &\leq -cQ_0 + \frac{A_0Q_0^{1-\gamma}}{r-\mu} + A_0 \int_{Q_0}^\infty \bar{A}(q, c)^{-\lambda} \left(\frac{\bar{A}(q, c)q^{-\gamma}(1-\gamma)}{r-\mu} - c \right) dq, \\
&= \underbrace{\left(\frac{\gamma\lambda}{\gamma\lambda-1} \right) \left(\frac{\lambda-1}{\lambda} \right) \frac{\bar{p}}{r-\mu} Q_0}_{\text{Revenue}(\bar{p})} - \underbrace{\left(\frac{\gamma\lambda}{\gamma\lambda-1} \right) cQ_0}_{\text{Cost}(c, \bar{p})} \\
&= \frac{\gamma\lambda}{\gamma\lambda-1} Q_0 \left(\frac{\bar{p}}{r-\mu} \frac{\lambda-1}{\lambda} - c \right). \tag{20}
\end{aligned}$$

Summing up, we can now state the regulator's optimization program in terms of an optimal control problem, with state variable $\Pi(c)$ and control $\bar{p}(c)$, and the Hamiltonian

$$\mathcal{H}(c) = \frac{\gamma\lambda Q(\bar{p})}{\gamma\lambda-1} \left(f(c) \left[\frac{\bar{p}(\lambda-1)}{(r-\mu)\lambda(1-\gamma)} - c \right] - \nu + \phi \left[\frac{\bar{p}(\lambda-1)}{(r-\mu)\lambda} - c \right] \right) - \phi\Pi(c), \tag{21}$$

with $f(c)$ the density of the distribution of costs, co-state variable $\nu(c)$ the multiplier of the incentive constraint

²²Recall that with competitive threshold prices, expected production cost equals market revenue (Proposition 1). So, the firm's expected profit equals the additional subsidy.

²³An alternative approach would be to allow for an exogenous social cost associated to leaving rents to the firm. If the regulator optimizes $W - \alpha\Pi$, for some constant social penalty for such rents $\alpha > 0$, a standard computation (as outlined in, e.g., Laffont and Tirole, 1993) results in an optimal threshold price

$$\bar{p}(c) = \frac{\lambda}{\lambda-1} (r-\mu) \left(c + \alpha \frac{F(c)}{f(c)} \right) = \bar{p}^c(c) \left(1 + \frac{\alpha F(c)}{c f(c)} \right),$$

under the assumption that the 'virtual costs' $c + \alpha F/f$ are monotone. The term $\alpha F(c)$ reflects the marginal social cost of leaving rents to a mass $F(c)$ of more efficient firms. Our model concerns the case where the cost of leaving those rents is endogenous, and depends on the shadow price of the budget constraints of more efficient firms.

(18), and $\phi(c)$ the multiplier for the budget constraint (20). The resulting first-order conditions for the optimum are

$$\frac{\partial \mathcal{H}}{\partial \bar{p}} = 0 \quad (22)$$

$$\frac{\partial \mathcal{H}}{\partial \Pi} = -\frac{d\nu}{dc} \quad (23)$$

We still have to impose the $t = 0$ participation constraint for all types. As usual, this can be achieved by requiring that at the upper boundary of the support of c , profits are zero, $\Pi(c_H) = 0$. As $\Pi(c)$ is decreasing in costs by lemma 2, this makes sure that all types get non-negative rents. At the lower boundary, we have either $\nu(c_L) = 0$, or price is at its monopoly level $\bar{p}^m(c_L)$.

We note at this point that the stochastic dynamic problem has effectively decoupled from the asymmetric information problem. The Hamiltonian of the system is equivalent to that of a static contracting problem, with the firm selecting its output from the menu offered by the principal, subject to a no-subsidy constraint. The information from the dynamics of the model is encoded in the scaling factors on price and quantity. We will therefore now proceed to solve this essentially static model; later, when we return to the *ongoing* participation and budget constraints, the temporal structure of the model will of course play a role again.

The solution to the first-order equations is as follows

Lemma 2. *Optimal threshold prices fall in one of three regimes:*

- *Regime I, the markup regime: the budget constraint does not bind, $\phi = 0$ and $\nu > 0$ is constant. In this regime,*

$$\bar{p}(c) = \bar{p}^c(c) \left(1 + \frac{\nu}{cf(c)} \right). \quad (24)$$

- *Regime II, the bunching regime: the budget constraint binds, \bar{p} is a constant in between competitive and monopoly prices $\bar{p}^m(c) \geq \bar{p} \geq \bar{p}^c(c)$, and*

$$\nu(c) = \nu(c_0) \frac{\bar{p}^m(c_0) - \bar{p}}{\bar{p}^m(c) - \bar{p}} + \frac{1}{1 - \gamma} \frac{\int_{c_0}^c (\bar{p} - \bar{p}^c(c')) f(c') dc'}{\bar{p}^m(c) - \bar{p}}. \quad (25)$$

for some c_0 within the interval in which this regime holds.

- *Regime III, the laissez-faire regime: the budget constraint binds, price is at the monopoly level,*

$$\bar{p}(c) = \bar{p}^m(c), \text{ and } \frac{\nu(c)}{f(c)} = \frac{\gamma c}{1 - \gamma}. \quad (26)$$

We see that as long as $f(c)$ is non-increasing, prices are non-decreasing in each of the regimes.

The optimal strategy is then a combination of two of the above regimes, joined together such that threshold price \bar{p} , co-state variable ν and profit Π are continuous on the regime boundary:

Proposition 2. *Optimal regulation involves either the laissez-faire regime (regime III) for the lowest cost types,*

followed by the markup regime (Regime I) for higher types; or bunching at constant price (Regime II) for the lowest types, followed by the markup regime (Regime I).

The laissez-faire case obtains if there exists a cost level $c_m \in [c_L, c_H]$ such that

$$\frac{\gamma c_m}{1 - \gamma} = \int_{c_m}^{c_H} \left((1 - \gamma) \frac{c}{c_m} + \gamma \frac{f(c_m)}{f(c)} \right)^{-1/\gamma} dc. \quad (27)$$

If such a c_m cannot be found, the optimum will involve bunching at constant price for low cost levels.

In the next section, we provide an illustration of the optimum strategy for the particular case of uniform cost distribution $f(c)$, when we can do the integrations explicitly. We will see that if the asymmetry of information, measured as c_H/c_L , is large enough, we will have the laissez-faire case for the low-cost types, whereas for small information asymmetry the bunching solution applies.

The intuition for the result is that, as usual, the optimal design balances the minimization of distortions away from optimality (investing too little too late) with rent extraction. As is standard, distortions for high-cost types are required to reduce the incentive for lower types to mimick high-cost types: the higher price caps make sure that investments are delayed longer, which is more costly for low cost types who suffer a reduction in output over a larger price cost margin.

In our model, leaving rents to the firms is costly because of the budget constraint: rents can only be paid for by allowing above competitive prices for low-cost types, which is why, necessarily, there will also be distortions for those types. There is a maximum to the rents that can be afforded, however: one cannot generate more rents than those created by the monopoly investment level. In situations with large rents, we therefore expect monopoly outcomes on the lower end of the cost distribution.

So far, we focused on the ex ante participation constraint: total ex ante expected rents are sufficient that all cost types earn non-negative expected rents under the contract. It is in fact possible to structure the payments of those rents over time such that also ongoing participation constraints are satisfied for each type: at any later moment total future expected income under the previously accepted contract are sufficient to cover future expected costs.

One implementation of the optimal contract that satisfies ongoing participation constraints is a non-linear revenue tax schedule $\tau(\bar{p})$. The firm announces at which maximum threshold price level \bar{p} it will invest, and is allowed to keep a fixed proportion $1 - \tau(\bar{p})$ of the revenues of selling capacity. It is obvious that such an implementation schedule also satisfies an ongoing budget constraint, as payments to the firm are a fraction of actual revenues at all times.

Proposition 3. *Suppose the regulator remunerates the firm by allowing the firm to keep all revenues up to a revenue tax $\tau(\bar{p})$ which depends on the firm's chosen investment threshold \bar{p} . The non-linear tax schedule is*

defined as:

$$\begin{aligned}\Pi(c) &= (1 - \tau(\bar{p}(c))) \cdot \text{Revenue}_0(\bar{p}(c)) - \text{Cost}_0(\bar{p}(c), c) \\ &= \frac{\gamma\lambda}{\gamma\lambda - 1} Q_0(\bar{p}(c)) \left[(1 - \tau(c)) \frac{1}{r - \mu} \frac{\lambda - 1}{\lambda} \bar{p}(c) - c \right]\end{aligned}\quad (28)$$

where $\Pi(c)$ and $\bar{p}(c)$ are the profits and investment threshold from the optimal mechanism. Given this non-linear tax level $\tau(\bar{p})$, the firm chooses the second best expansion strategy $\bar{p}(c)$ and earns non-negative expected profits at any moment $t > 0$, i.e. the ongoing participation constraints are satisfied.

The results follow directly from the fact that with a proportional tax, the firm's objective function at $t > 0$ is of the same form of that of a greenfield firm, up to the constant sunk benefit corresponding to installed capacity $Q(t)$. Indeed the participation constraint is hardest to satisfy for the green-field investor without installed capacity. Of course, for low c where the budget constraint binds with equality, the revenue tax is zero, and the firm keeps all revenues. Only in the markup regime, the higher c values, do we have positive tax rates $\tau(\bar{p})$. In the next section we will investigate the exact form of $\tau(\bar{p})$ for the particular example with uniform cost distribution.

5 Example

For an illustration of the optimal regulation, we explore the case with uniform distribution on investment costs, $f(c) = \frac{1}{c_H - c_L}$. In that case, we can do the required integrations analytically, and use that to solve the model. As we will see, if the range of costs, measured by c_H/c_L , is not too large, we will have an optimum in which lower-cost firms bunch at a constant price (Regime II), whereas for higher costs firms, the budget constraint will not be binding and we are in the markup regime. Conversely, if cost uncertainty is large, the lowest cost types will end up being unregulated and setting monopoly prices instead.

To find the optimum we need to join together the various pricing regimes in a continuous fashion. By proposition 2 we have either a laissez-faire regime for low costs $c \in [c_L, c_m]$, and the markup regime prevails for higher cost levels. Or, if we cannot match profits from the two regimes by solving (27), we will instead have a bunching regime for the low types.

Focusing first on the laissez-faire situation, we can work out (27) explicitly as f is constant in the uniform case. Doing the integration leads to the following

Proposition 4. *With uniform cost distribution and a large support $\frac{c_L}{c_H} \leq \xi \equiv (1 - \gamma)(\gamma^{\frac{\gamma}{\gamma-1}} - \gamma)^{-1}$, we have the laissez-faire regime with monopoly pricing for $c \in [c_L, c_m]$, and a constant markup on competitive prices for $c \in [c_m, c_H]$, with $c_m = c_H \xi$.*

If the range of costs is small, so that $c_L > c_H \xi$, we cannot match profits at any $c_m \in [c_L, c_H]$ satisfying (27), but instead have bunching at constant price for low cost levels, $c < c_b$, whereas again we find constant markups for $c > c_b$. To find c_b , as well as the price and markup levels, we can again use continuity of $\Pi(c)$, $\bar{p}(c)$ and $\nu(c)$,

and combine the incentive compatibility equation (18) with the expression for ν in the bunching regime, (25). The resulting conditions on the transition level c_b and the bunching price are as in the following proposition.

Proposition 5. *With uniform cost distribution and a small support $\frac{c_L}{c_H} \geq \xi$, we have bunching for $c \in [c_L, c_b]$, and the markup regime (with constant markup) for $c \in [c_b, c_H]$, with transition point c_b and bunching price $\bar{p}_b \equiv \frac{(r-\mu)\lambda}{\lambda-1} \tilde{p}_b$ determined jointly by the solution to*

$$\begin{aligned} \tilde{p}_b - c_b &= \frac{\tilde{p}_b(c_b - c_L) - \frac{1}{2}(c_b^2 - c_L^2)}{c_b - \tilde{p}_b(1 - \gamma)} \\ \tilde{p}_b^{-\frac{1}{\gamma}}(\tilde{p}_b - c_b) &= \frac{\gamma}{\gamma - 1} \left((\tilde{p}_b - c_b + c_H)^{\frac{\gamma-1}{\gamma}} - \tilde{p}_b^{\frac{\gamma-1}{\gamma}} \right). \end{aligned}$$

The first equation in the proposition follows from integrating ν between c_L and c_b , and requiring that its end value equals the markup in the markup regime. The second equation follows from making sure the profits at that transition point (where the budget constraint holds with equality) coincide with the integral of the incentive compatibility constraint, and profits at c_H are zero.

We plot the results of the two solutions, one for high cost uncertainty (high c_H/c_L), figure 2, and one for low cost uncertainty (c_H/c_L nearer to one), figure 3. In the first case, we have monopoly pricing up to c_m , and in the second we see bunching at constant price for low realizations of costs.

[Figure 2 about here.]

With a direct revelation mechanism, the regulator offers the menu of contracts $\{(\Pi(c), \bar{p}(c))\}$ and the firm truthfully announces its type c . As we saw, alternatively, the regulator could offer a menu of contracts $\{\tau(\bar{p})\}$ in which the firm announces at which maximum threshold price level \bar{p} it will invest, and the regulator taxes a fraction $\tau(\bar{p})$ of the revenue of selling capacity. The tax rate $\tau(\bar{p})$ is determined implicitly by equation (28). For the markup regime we find

$$\tau(\bar{p}) = \frac{\bar{p}(c_H) - \bar{p}^c(c_H)}{\bar{p}(c)} - \frac{\gamma}{1 - \gamma} \left(1 - \left(\frac{\bar{p}(c)}{\bar{p}(c_H)} \right)^{\frac{1-\gamma}{\gamma}} \right),$$

matching to $\tau = 0$ at the transition point with the laissez-faire or bunching regimes. Figures 4 and 5 plot the resulting set of pairs of threshold prices \bar{p} versus required taxes τ . With large cost asymmetry c_H/c_L , we have the range of monopoly prices for low costs, accompanied with zero taxation. With smaller c_H/c_L , we have the single bunching price for the lower cost realizations. To benefit from the zero tax rate the firm needs to invest early, at a relatively low threshold price. Alternatively, the firm could invest later, which implies accepting a higher tax rate.

[Figure 3 about here.]

6 Discussion and conclusion

In this article we have derived the optimal regulation of network expansion, by combining the real option and principal-agent literature. We show that a regulatory exemption might sometimes be optimal, not to prevent hold-up problems created by a lack of commitment power by the regulator as in Gans and King (2004), but because of the combination of a self-financing constraint and information asymmetry: The information rents of efficient cost types need to be collected in the market, which require higher prices for network access and delayed investments. However, such an exemption is not a blanket authorization for the firm to invest whenever it feels fit. It implies the requirement, in accordance with the relatively low expansion costs of the efficient firm, for sufficiently early investment. If the support of the asymmetric information is small, the efficient firm requires less information rents and a laissez-faire regime is no longer optimal. Instead the regulator will bunch the regulatory contracts for the most efficient firms, obliging them to invest at a price below the monopoly price.

Note that whether a laissez-faire regime is socially optimal, does *not* depend on the level of demand uncertainty, and the riskiness of investments.²⁴ This stands in contrast with the requirement in EU energy markets, that the *risk* should be too high for investments to incur without exemption. In our model, the laissez-faire regime is a reward for the low cost firm, who invests earlier than the high cost firm.

In the optimal regulation, inefficient firms are subject to an investment requirement, *i.e.* they are obliged to invest whenever the price for capacity reaches a threshold level. It is well known from the literature that when demand is stochastic, a price cap cannot be used to both limit the rents of the regulated firm and to incentivize timely investments. Instead the regulator needs to rely on a combination of instruments such as for instance in the UK where the new regulation sets a price cap but also sets an output obligation on the firms. Alternatively, as we show, the regulator can penalize a firm investing late by increasing its tax level.

Formally, our model considers greenfield investments. The same analysis holds, however, when regulated firms are allowed by law to recoup sunk investments cost of previously built assets, as long as these are at or below the regulated level. Such a principle of no regulatory takings is common. If, on the other hand, some of the investments have already been fully paid off at the time of contracting, the participation constraint is relaxed, and less information rents need to be paid to investors. The laissez-faire regime is then less likely to be optimal. We also assumed that capacity does not depreciate. This could easily be adjusted by appropriate shifts in μ and r , as in Dobbs (2004).

References

Alonso, R and Matouschek, N. "Optimal Delegation." *The Review of Economic Studies*, Vol. 75 (2008), pp. 259–293.

Armstrong, M. and Sappington, D. E. M. "Chapter 27: Recent Developments in the Theory of Regulation."

²⁴Recall that the firm is perfectly informed about its own investment costs, so the only source of risk is demand uncertainty.

- In M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization Vol. 3*, Amsterdam: North-Holland, 2007.
- Arve, M. and Zwart, G. “Optimal Procurement and Investment in New Technologies under Uncertainty.” SSRN Paper no. 2471089, Social Science Research Network, 2014.
- Averch, H. and Johnson, L. L. “Behavior of the Firm under Regulatory Constraint.” *The American Economic Review*, Vol. 52 (1962), pp. 1052–1069.
- Baron, D. P. and Besanko, D. “Regulation and information in a continuing relationship.” *Information Economics and Policy*, Vol. 1 (1984), pp.267–302.
- Baron, D. P. and Myerson, R. B. “Regulating a Monopolist with Unknown Costs.” *Econometrica*, Vol. 50 (1982), pp. 911–930.
- Battaglini, M. “Optimality and renegotiation in dynamic contracting.” *Games and Economic Behavior*, Vol. 60 (2007), pp. 213–246.
- Beesley, M. E. and Littlechild, S. C. “The Regulation of Privatized Monopolies in the United Kingdom.” *The RAND Journal of Economics*, Vol. 20 (1989), pp. 454–472.
- Bennett, J. and Iossa, E. “Building and managing facilities for public services.” *Journal of Public Economics*, Vol. 90 (2006), pp. 2143–2160.
- Bester, H. and Strausz, R. “Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case.” *Econometrica*, Vol. 69 (2001), pp. 1077–1098.
- Broer, P. and Zwart, G. “Optimal Regulation of Lumpy Investments.” *Journal of Regulatory Economics*, Vol. 44 (2013), pp.177–196.
- Cabral, L. M. B. and Riordan, M. H. “Incentives for Cost Reduction under Price Cap Regulation.” *Journal of Regulatory Economics*, Vol. 1 (1989), pp. 93–102.
- de Bettignies, J-E. and Ross, T. W. “Public–private partnerships and the privatization of financing: An incomplete contracts approach.” *International Journal of Industrial Organization*, Vol. 27 (2009), pp. 358–368.
- Dixit, R. K. and Pindyck, R. S. *Investment under Uncertainty*. Princeton: Princeton University Press, 1994.
- Dobbs, I. M. “Intertemporal Price Cap Regulation under Uncertainty.” *The Economic Journal*, Vol. 114 (2004), pp. 421–440.
- Evans, L. and Guthrie, G. “Price-cap Regulation and the Scale and Timing of Investment.” *The RAND Journal of Economics*, Vol. 43 (2012), pp. 537–561.
- Evans, L. T. and Guthrie, G. A. “Risk, Price Regulation, and Irreversible Investment.” *International Journal of Industrial Organization*, Vol. 23 (2005), pp. 109–128.
- Gans, J. S. and King, S. P. “Access Holidays and the Timing of Infrastructure Investment.” *Economic Record*,

- Vol. 80 (2004), pp. 89–100.
- Gautier, A. and Mitra, M. “Regulating a Monopolist with Limited Funds.” *Economic Theory*, Vol. 27 (2006), pp. 705–718.
- Gilbert, R. J. and Newbery, D. M. “The Dynamic Efficiency of Regulatory Constitutions.” *The RAND Journal of Economics*, Vol. 25 (1994), pp. 538–554.
- Guesnerie, R. and Laffont, J-J. “A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-managed Firm.” *Journal of Public Economics*, Vol. 25 (1984), pp. 329–369.
- Guthrie, G. “Regulating Infrastructure: The Impact on Risk and Investment.” *Journal of Economic Literature*, Vol. 44 (2006), pp. 925–972.
- Hart, O. “Incomplete Contracts and Public Ownership: Remarks, and an Application to Public-Private Partnerships.” *The Economic Journal*, Vol. 113 (2003), pp. 69–76.
- Hart, O. D. and Tirole, J. “Contract Renegotiation and Coasian Dynamics.” *The Review of Economic Studies*, Vol. 55 (1988), pp.509–540.
- Iossa, E. and Martimort, D. “Risk allocation and the costs and benefits of public–private partnerships.” *The RAND Journal of Economics*, Vol. 43 (2012), pp.442–474.
- Iossa, E. and Martimort, D. “The Simple Microeconomics of Public-Private Partnerships.” *Journal of Public Economic Theory*, Vol. 17 (2015), pp. 4–48.
- Joskow, P. L. “Incentive Regulation and its Application to Electricity Networks.” *Review of Network Economics*, Vol. 7 (2008), pp. 547–560.
- Laffont, J-J. and Tirole, J. “The Dynamics of Incentive Contracts.” *Econometrica*, Vol. 56 (1988), pp. 1153–1175.
- Laffont, J-J. and Tirole, J. “Adverse Selection and Renegotiation in Procurement.” *The Review of Economic Studies*, Vol. 57 (1990), pp. 597–625.
- Laffont, J-J. and Tirole, J. “The Politics of Government Decision-Making: A Theory of Regulatory Capture.” *The Quarterly Journal of Economics*, Vol. 106 (1991), pp.1089–1127.
- Laffont, J-J. and Tirole, J. *A Theory of Incentives in Procurement and Regulation*. Cambridge: MIT Press, 1993.
- Lesieutre, B. C. and Eto, J. H. “Electricity transmission congestion costs: A review of recent reports.” LBNL-54049, Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, 2003.
- Loeb, M. and Magat, W. A. “A Decentralized Method for Utility Regulation.” *The Journal of Law & Economics*, Vol. 22 (1979), pp. 399–404.
- Martimort, D. and Pouyet, J. “To build or not to build: Normative and positive theories of public–private partnerships.” *International Journal of Industrial Organization*, Vol. 26 (2008), pp.393–411.

- Maskin, E. and Tirole, J. “Public–private partnerships and government spending limits.” *International Journal of Industrial Organization*, Vol. 26 (2008), pp. 412–420.
- McDonald, R. and Siegel, D. “The Value of Waiting to Invest.” *Quarterly Journal of Economics*, Vol. 101 (1986), pp.707–727.
- OFGEM. “Handbook for Implementing the RIIO Model.” Technical report OFGEM, 2010a.
- OFGEM. “RIIO: A New Way to Regulate Energy Networks; Final Decision.” Technical report OFGEM, 2010b.
- Pindyck, R. S. “Irreversible Investment, Capacity Choice, and the Value of the Firm.” *American Economic Review*, Vol. 78 (1988), pp. 969–85.
- Roques, F. A. and Savva, N. “Investment under Uncertainty with Price Ceilings in Oligopolies.” *Journal of Economic Dynamics and Control*, Vol. 33 (2009), pp. 507–524.
- Salant, D. J. and Woroch, G. A. “Trigger Price Regulation.” *The RAND Journal of Economics*, Vol. 23 (1992), pp.29–51.
- Sappington, D. E. M. “Price Regulation and Incentives.” In M. Cave, S. Majumda and I. Vogelsang, eds., *Handbook of Telecommunications Economics Vol. 1 Structure Regulation and Competition*. Amsterdam: Elsevier, 2002.
- Williamson, B., Lewin, D. and Wood, S. Fostering Investment and Competition in the Broadband Access. A report for ETNO, Plum Consultancy, London, 2016.

A Proofs

Proof of proposition 1 The expressions for \bar{p}^m and \bar{p}^c (equations 14 and 15) are derived in the text. The zero profit result for a greenfield firm (equation 16), follows from substituting $A_0 = \bar{p}^c Q_0^\gamma$ and $\bar{A}(q) = \bar{p}^c q^\gamma$ in the expression for firm value, equation (13), and integrating. We then find

$$V - cQ_0 = \frac{\gamma\lambda Q_0}{\gamma\lambda - 1} \left(\frac{\bar{p}^c(\lambda - 1)}{(r - \mu)\lambda} - c \right),$$

and by the definition of \bar{p}^c (15) this vanishes.

Proof of lemma 1 We have

$$R(c, \hat{c}) = T_0(\hat{c}) - cQ_0(\bar{p}(\hat{c})) \frac{\gamma\lambda}{\gamma\lambda - 1},$$

and $\Pi(c) = R(c, \hat{c} = c)$. If $\hat{c} = c$ optimizes R , we can use the envelope theorem to find

$$\frac{d\Pi(c)}{dc} = \left. \frac{\partial R(c, \hat{c})}{\partial c} \right|_{\hat{c}=c} = -Q_0(\bar{p}(c)) \frac{\gamma\lambda}{\gamma\lambda - 1}.$$

To verify that $\bar{p}(c)$ is non-decreasing in c , we note that truthful revelation for a firm with type c requires that

$R(c, c) - R(c, \hat{c}) \geq 0$ for any \hat{c} . Equivalently, for a firm with type \hat{c} it must be that $R(\hat{c}, \hat{c}) - R(\hat{c}, c) \geq 0$. Hence, combining both expressions, for any c, \hat{c} , we must have that:

$$(R(c, c) - R(\hat{c}, c)) - (R(c, \hat{c}) - R(\hat{c}, \hat{c})) \geq 0,$$

or equivalently:

$$\int_{\hat{c}}^c \left(\frac{\partial R}{\partial c'}(c', c) - \frac{\partial R}{\partial c'}(c', \hat{c}) \right) dc' = \frac{\gamma\lambda}{\gamma\lambda - 1} [Q_0(\bar{p}(\hat{c})) - Q_0(\bar{p}(c))] (c - \hat{c}) \geq 0.$$

It then follows that for $\hat{c} < c$, $Q_0(\bar{p}(\hat{c})) \geq Q_0(\bar{p}(c))$, or as demand is downward sloping $\bar{p}(\hat{c}) \leq \bar{p}(c)$.

Proof of lemma 2 The first-order condition of the Hamiltonian for Π (23) gives the dynamics for $\nu(c)$,

$$\frac{d\nu}{dc} = \phi.$$

In cost-regions where the budget constraint (20) does not bind, its multiplier is zero, $\phi = 0$, and hence ν is constant. Using the short-hand

$$\tilde{p} = \frac{\bar{p}(\lambda - 1)}{(r - \mu)\lambda},$$

we can write the first-order condition of the Hamiltonian for \bar{p} (22) as:

$$f(c)(\tilde{p} - c) - \nu + \phi(\tilde{p}(1 - \gamma) - c) = 0.$$

With $\phi = 0$ and $\nu = \nu^{cte}$ constant, this leads to

$$\tilde{p}(c) = c + \frac{\nu^{cte}}{f(c)}$$

in this *markup regime I*. Note that with a zero mark-up, $\nu^{cte} = 0$, the threshold price $\tilde{p}(c)$ is equal to the competitive threshold price and the firm's profit Π is equal to zero by proposition 1. This would violate the firm's incentive constraint (information rents need to be positive for all $c < c_H$ by lemma 1) and can thus not be optimal. Hence, $\nu^{cte} > 0$.

In the cost-regions where the budget constraint (20) does bind, the firm's profit is equal to the revenue from capacity sales $\Pi = \frac{\gamma\lambda Q_0(\tilde{p})}{\gamma\lambda - 1}(\tilde{p} - c)$. From this we can derive the total derivative of profits as a function of the firm's type c :

$$\frac{d\Pi}{dc} = \frac{\partial\Pi}{\partial\tilde{p}} \frac{d\tilde{p}}{dc} - \frac{\gamma\lambda Q_0(\tilde{p})}{\gamma\lambda - 1}$$

Given the incentive compatibility condition (18), the first term on the right hand side is zero, we must then have that either \tilde{p} is constant, or $\frac{\partial\Pi}{\partial\tilde{p}} = 0$, and hence monopoly pricing.

In the case of constant \tilde{p} , we are in the *bunching regime II*. From the first order condition on \tilde{p} , with

$d\nu/dc = \phi$, we then have

$$f(c)(\tilde{p} - c) - \nu + \frac{d\nu}{dc}(\tilde{p}(1 - \gamma) - c) = 0,$$

with constant \tilde{p} . This is a differential equation for $\nu(c)$, for which the solution is equation (25).

Finally, in the *laissez-faire regime III*, we have $\tilde{p}^m = \frac{c}{1-\gamma}$, so that the first-order equation reduces to

$$\frac{\nu(c)}{f(c)} = \tilde{p}^m - c = \frac{\gamma c}{\gamma - 1}. \quad (29)$$

Proof of proposition 2 With $f(c)$ non-increasing, $\frac{\nu(c)}{f(c)}$ is non-decreasing in all three regimes. For the highest cost realization c_H , we have that the information rents are zero, $\Pi(c_H) = 0$, and therefore a non-binding budget regime ($\phi = 0$), so we will have the markup regime. Suppose that for lower cost levels, we have a region $[\underline{c}, \bar{c}]$ in which the bunching regime at a constant price \hat{p} applies, $c_H > \bar{c} > \underline{c} > c_L$. As the lower boundary is assumed to be strictly larger than c_L , the bunching regime is connected from below to one of the other regimes (laissez-faire or markup regime). Hence, at this point $c = \underline{c}$, we have $\hat{p} - \underline{c} = \frac{\nu(\underline{c})}{f(\underline{c})}$, because that relation holds in both laissez-faire and markup regimes. Similarly, the bunching regime is connected to the laissez-faire or markup regime from above, so we must have $\hat{p} - \bar{c} = \frac{\nu(\bar{c})}{f(\bar{c})}$. But this leads to a contradiction as $\bar{c} > \underline{c}$ and $\nu(c)/f(c)$ is non-decreasing. Hence we cannot have $\underline{c} > c_L$, and hence if the bunching regime occurs it is optimal for certain cost level, it is also the case for all lower cost levels.

We therefore have either the laissez-faire regime in a lower cost segment, $c \in [c_L, c_m]$ for some c_m , or bunching in a lower cost segment $c \in [c_L, c_b]$, for some c_b .

To find out whether the laissez-faire solution can apply, we need to verify that price, co-state variable ν and profit Π are continuous at the boundary point c_m between the laissez-faire and the mark-up regimes.

In the constant markup regime on $[c_m, c_H]$, $\nu = \nu^{cte}$ is a constant, and profit $\Pi(c)$ is determined by the boundary condition $\Pi(c_H) = 0$ and the incentive compatibility condition which determines $d\Pi/dc$. On the other side, in the laissez-faire regime, $\nu(c_m) = f(c_m)\frac{\gamma c_m}{1-\gamma}$, and $\Pi(c_m) = \frac{\gamma\lambda}{\gamma\lambda-1}\frac{\gamma c_m}{1-\gamma}Q_0^m$. Matching ν from both regimes at c_m gives

$$\frac{\nu^{cte}}{f(c_m)} = \frac{\gamma c_m}{1-\gamma},$$

which determines ν^{cte} as a function of c_m . The price in the mark-up regime is then determined by $\tilde{p} = c + \frac{\nu^{cte}}{f(c)}$, where $\tilde{p} = \frac{\lambda-1}{\lambda(r-\mu)}\bar{p}$. To match profits, we use incentive compatibility on Π to write

$$\Pi(c_m) = \frac{\gamma\lambda}{\gamma\lambda-1} \int_{c_m}^{c_H} Q_0(\tilde{p}(c')) dc'$$

which produces the condition in the proposition.

Proof of proposition 3 The *ongoing participation constraint* requires that at $t \geq 0$, expected future transfers

offset expected future expansion costs. It is sufficient to check the ongoing participation constraint at times t when price reaches the investment threshold, i.e. $A(t)Q(t)^{-\gamma} = \bar{p}(c)$. We need to verify that

$$(1 - \tau(c)) \frac{\gamma\lambda}{\gamma\lambda - 1} Q(t) \frac{1}{r - \mu} \frac{\lambda - 1}{\lambda} \bar{p}(c) - \frac{1}{\gamma\lambda - 1} cQ(t) \geq 0,$$

where the first term represents the after-tax expected revenues at state $A(t), Q(t)$, and the second term the costs of continuing expansion from this point onward. Notably, compared to equation (20), costs do now not include the startup costs $cQ(t)$ to bring capacity from greenfield to level $Q(t)$. These expected future revenues and expansion costs in state $\{A(t), Q(t)\}$ at $t > 0$, are equal to those of a greenfield investor at $t = 0$ with an initial demand shift $A'_0 = A(t)$ and an additional capital endowment $cQ(t)$. The ongoing participation constraint therefore rewrites to

$$\left\{ (1 - \tau(\bar{p})) \cdot \text{Revenue}_0(\bar{p}) - \text{Cost}_0(c, \bar{p}) \right\} \frac{Q(t)}{Q_0} + cQ(t) \geq 0.$$

The first factor in curly brackets reflects the participation constraint of the greenfield firm and is therefore non-negative. The second term is always positive.

Proof of proposition 4 We can directly apply proposition 2 to the uniform case, with $f(c) = 1/(c_H - c_L)$ constant. Matching ν/f of both regimes at c_m gives

$$\frac{\nu^{cte}}{f} = \frac{\gamma c_m}{1 - \gamma},$$

which determines ν^{cte} as a function of c_m . Matching profits at the boundary c_m between the laissez-faire and markup regimes, as in (27), we can then work out the required value of c_m by doing the integration.

Proof of proposition 5 In the alternative case of the bunching regime for low cost $c \in [c_L, c_b]$ and the markup regime for higher costs $c \in [c_b, c_H]$, we again find the boundary point c_b by matching price \bar{p} , co-stage variable ν and profit Π . The co-state variable ν is defined in bunching region by a differential equation and the boundary condition $\nu(c_L) = 0$. Profit in the mark-up regime is defined by the incentive compatibility and the boundary condition $\Pi(c_H) = 0$.

We first compute $\nu(c_b)$ from the differential equation for ν in the bunching region, with boundary condition $\nu(c_L) = 0$. Doing the integration, this gives

$$\frac{\nu(c_b)}{f} = \frac{\tilde{p}_b(c_b - c_L) + \frac{1}{2}(c_L^2 - c_b^2)}{c_b - \tilde{p}_b(1 - \gamma)}$$

and this should equal $\tilde{p}_b - c_b = \frac{\nu^{cte}}{f}$ by matching to the markup region. This is the first equation of the proposition.

From incentive compatibility in the mark-up regime we find

$$\Pi(c_b) = \frac{\gamma\lambda}{\gamma\lambda - 1} \int_{c_m}^{c_H} Q_0(\bar{p}(c')) dc',$$

where $\tilde{p}(c) = c + \frac{\nu^{cte}}{f} = c + \tilde{p}_b - c_b$. Matching this to the profits at c_b from the binding budget constraint gives the second equation of the proposition.

B Constant price thresholds

In this appendix we establish that also in the adverse selection case, the investment threshold occurs at constant price, $\bar{p}(c) = \bar{A}(Q, c)Q^{-\gamma}$. In terms of $\bar{A}(Q, c)$, we have welfare

$$W_0(\bar{A}(Q, c), c) = -cQ_0 + \frac{A_0Q_0^{1-\gamma}}{(1-\gamma)(r-\mu)} + A_0^\lambda \int_{Q_0}^{\infty} \bar{A}(q, c)^{-\lambda} \left(\frac{\bar{A}(q, c)q^{-\gamma}}{r-\mu} - c \right) dq,$$

with $Q_0(\bar{A}, c)$ defined by $\bar{A}(Q, c) = A_0$. Similarly, the incentive and budget constraints are

$$\begin{aligned} \frac{d\Pi}{dc} &= -Q_0 - A_0^\lambda \int_{Q_0}^{\infty} \bar{A}(q, c)^{-\lambda} dq, \\ \Pi(c) &\leq -cQ_0 + \frac{A_0Q_0^{1-\gamma}}{r-\mu} + A_0^\lambda \int_{Q_0}^{\infty} \bar{A}(q, c)^{-\lambda} \left(\frac{\bar{A}(q, c)q^{-\gamma}(1-\gamma)}{r-\mu} - c \right) dq. \end{aligned}$$

Combining these, we can then write the Hamiltonian

$$\begin{aligned} \mathcal{H} &= \frac{A_0Q_0^{1-\gamma}}{(1-\gamma)(r-\mu)} (f + (1-\gamma)\phi) \\ &\quad + A_0^\lambda \int_{Q_0}^{\infty} \bar{A}(q, c)^{-\lambda} \left(\frac{\bar{A}(q, c)q^{-\gamma}}{r-\mu} (f + (1-\gamma)\phi) - (fc + \nu + \phi c) \right) dq \\ &\quad - (fc + \nu + \phi c)Q_0 - \phi\Pi(c). \end{aligned}$$

To optimize, we now need to use variational calculus on the function $\bar{A}(Q, c)$. Such a variation also induces a concomitant variation δQ_0 so as to keep $\bar{A}(Q, c) = A_0$ verified. It is now straightforward to see that the δQ_0 terms in the variation vanish, leaving us only with the integral,

$$A_0^\lambda \int_{Q_0}^{\infty} \delta \bar{A}(q, c) \bar{A}(q, c)^{-\lambda-1} \left[(1-\lambda) \frac{\bar{A}(q, c)q^{-\gamma}}{r-\mu} (f + (1-\gamma)\phi) + \lambda(fc + \nu + \phi c) \right] dq = 0$$

Because this holds for any variation $\delta \bar{A}(Q, c)$, we see that $\bar{A}(Q, c)Q^{-\gamma}$ is independent from Q , and we regain the first-order equations from the main text.

List of Figures

| | | |
|---|---|----|
| 1 | Expected welfare function $W(A, Q \bar{A}, c)$ and its boundary conditions. | 26 |
| 2 | Large range of costs: Laissez-faire and mark-up regimes | 27 |
| 3 | Small range of costs: Bunching regime and markup regime | 27 |
| 4 | Taxes versus price threshold with laissez-faire regime | 28 |
| 5 | Taxes versus price threshold with bunching regime | 28 |

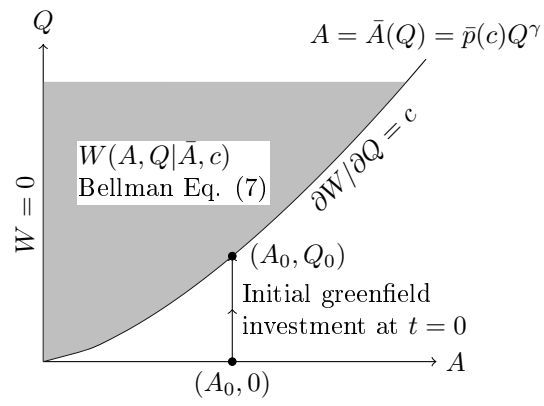


Figure 1: Expected welfare function $W(A, Q | \bar{A}, c)$ and its boundary conditions.

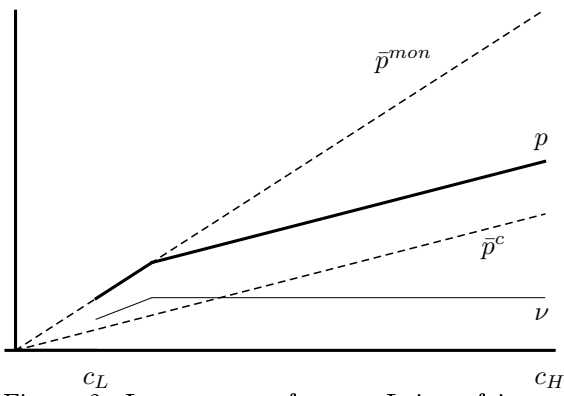


Figure 2: Large range of costs: Laissez-faire and mark-up regimes

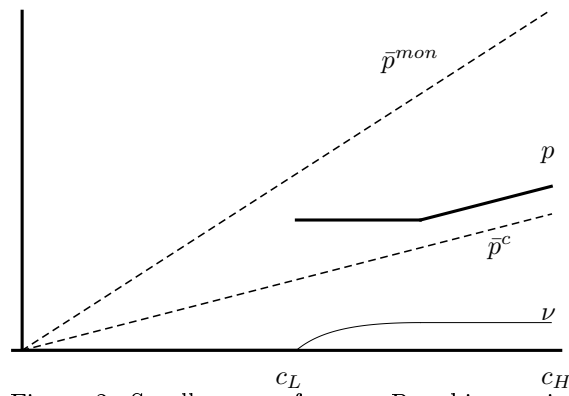


Figure 3: Small range of costs: Bunching regime and markup regime

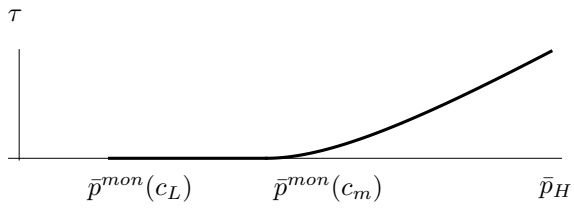


Figure 4: Taxes versus price threshold with laissez-faire regime

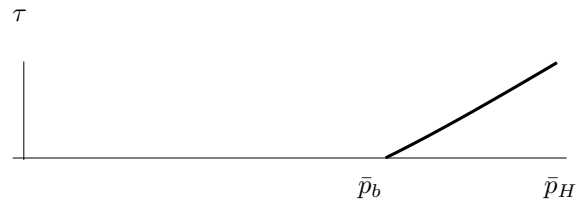


Figure 5: Taxes versus price threshold with bunching regime