Società italiana di economia pubblica
XVI CONFERENZA
POLITICHE PUBBLICHE, SVILUPPO E CRESCITA
Pavia, Università, 7 - 8 ottobre 2004

# PARTIAL REGULATION OF QUANTITIES: AN IMPLEMENTATION SCHEME

ETIENNE BILLETTE DE VILLEMEUR, ANNALISA VINELLA

pubblicazione internet realizzata con contributo della
COMPAGNIA di San Paolo

# Partial Regulation of Quantities: An Implementation Scheme


Etienne Billette de Villemeur
University of Toulouse (IDEI and GREMAQ)

Annalisa Vinella
University of Toulouse, GREMAQ
and Dipartimento di Scienze Economiche, University of Bari


Preliminary Version - Please do not cite


**Abstract**

Highly concentrated industries often yield output under-provision, a point of crucial importance in the debate around the liberalization of services of general interest. Having in mind the former monopolies recently opened up to competition, where only the original incumbents, not the entrants, are subject to regulation, we propose a regulatory scheme imposing obligations on the supply of one of the outputs only. We exhibit the conditions that ensure convergence of this multi-period mechanism to welfare-enhancing allocations. The mechanism requires the sole access to book-keeping data and memory of the previous performance.

# 1 Introduction

Highly concentrated industries yield output under-provision. This issue should play a key role in assessing the merits of the generalised liberalization process within the EU. It is indeed often crucial in the case of utilities providing services of general interest.

The present paper proposes a partial regulatory scheme designed to induce larger production than what the market equilibrium would entail. The economic literature has traditionally investigated monopolies where the profit-maximising producer distorts all its activities away from the socially optimal allocation, calling for regulatory intervention over all of them. In reality, it is frequently impossible to control all inefficiency sources. However, a partial correction like imposing obligations on one or some of the productive areas may lead to a significant welfare improvement. It may concern either some of the activities managed by a single operator or the unique activity executed by one amongst several providers. Even for an industry where there is no competition, it is quite common to observe that new services are rarely subject to regulation. Where market liberalization occured, while competing with unregulated firms, the previous monopolist remains indeed often subject to regulation.

Referring to this scenario, we model an industry providing two imperfectly substitute goods. Both multiproduct monopoly and duopoly are considered. We characterize the allocations which realize in case one of the outputs is produced by a provider perfectly instructed to maximise social welfare, the second quantity being chosen according to the supplier's interests. The effects of the asymmetry between objective functions show up as corrective Ramsey-type terms in the mark-up formulas. In general these mixed allocations induce a higher social welfare than the free market equilibria.

The idea that supply is determined by pursuing heterogeneous objectives is inherited from the literature about mixed oligopolies. Biglaiser and Ma (1995) study the optimal price regulation of a Stackelberg incumbent in the presence of an unregulated follower. They concentrate on optimal mechanisms, whereas we focus more on implementation and look for a general mechanism that does not depend on the industry structure. We propose a multi-period scheme of partial regulation designed to implement optimal allocations by imposing obligations on the supply of one output only. The regulated firm, which is subject to a constraint which embodies prices and profits of the previous period, maximizes its profits on a per-period basis. Since it only requires accessibility to the firm's official book-keeping data, the mechanism is manageable for the regulator. It is not meant to extract private information by meeting incentive-compatibility conditions. It provides the right incentives by evolving as a process with memory of the previous performance. In this sense, the regulatory sequence at stake is in the spirit of the converging process analysed by Vogelsang and Finsinger (1979). While the latter are concerned with inducing a multiproduct monopoly to progressively adjust the charges for all offered goods, until the Ramsey prices are enforced, we investigate asymmetric contexts by allowing for the presence of multiple economic agents, still preserving the most stylized relationship between regulator/principal and regulated enterprise/agent.

The paper is organized as follows. Section 2 presents the model. Section 3, 4 and 5 are devoted to the analysis of the ideal policy for the multiproduct monopoly, the Cournot and the Stackelberg duopoly respectively, both in a first and second-best scenario. In Section 6 we propose the output regulatory scheme and investigate the implementation in the three different market structures. Section 7 concludes.

# 2 The Model: Products, Demands, Welfare

We consider the market(s) for two imperfectly substitute products, indexed by $j$ and $k$. The representative consumer surplus, net of total expense, is given by

$$NS\left(q_j, q_k\right) = S\left(q_j, q_k\right) - p_j q_j - p_k q_k. \tag{1}$$

It gives rise to the demand $q_j = Q_j\left(p_j, p_k\right)$ for good $j$, a function of both prices defined implictly by the pair of downward-sloping inverse demand functions $P_j(q_j, q_k) = \left(\partial S/\partial q_j\right)$ and $P_k(q_j, q_k) = \left(\partial S/\partial q_k\right)$. Observe that, under gross substitutability, the cross derivative $\left(\partial^2 S/\partial q_j \partial q_k\right)$ is negative. It follows that the price of one commodity decreases with the quantity of the other commodity supplied on the market. Moreover, the regularity condition $\left(\partial^2 S/\partial q_j \partial q_k\right) = \left(\partial^2 S/\partial q_j \partial q_k\right)$ imposes symmetric cross-price effects for the demand of a well-behaved consumer, hence $\left(\partial P_j/\partial q_k\right) = \left(\partial P_j/\partial q_k\right) < 0$.

Denote $\varepsilon_{jj} \equiv -\left(\frac{\partial q_j}{\partial p_j}\frac{p_j}{q_j}\right)$ and $\varepsilon_{jk} \equiv \left(\frac{\partial q_j}{\partial p_k}\frac{p_k}{q_j}\right)$ the (absolute value of the) own-price and the cross-price elasticity of demand, respectively. Let $\widehat{\varepsilon}_{jj} \equiv \varepsilon_{jj}\left(\frac{\varepsilon_{jj}\varepsilon_{kk} - \varepsilon_{jk}\varepsilon_{kj}}{\varepsilon_{jj}\varepsilon_{kk} + \varepsilon_{jj}\varepsilon_{jk}}\right)$ be the (standard) super-elasticity and $\eta_{jj} \equiv \left(\frac{-\partial P_j}{\partial q_j}\frac{q_j}{p_j}\right)$ the (absolute value of the) elasticty of the inverse demand function with respect to quantity. The value $1/\varepsilon_{jj}$ defines the markup of a Bertrand duopolist. The value $1/\widehat{\varepsilon}_{jj}$ defines the markup of a multiproduct monopolist. The elasticity $\eta_{jj}$ defines the markup over the commodity $j$ of a Cournot duopolist. It is possible to show that

$$\frac{1}{\varepsilon_{jj}} < \eta_{jj} < \frac{1}{\widehat{\varepsilon}_{jj}}.$$

Social welfare consists in (unweighed) total surplus as measured by gross consumer surplus net of total production costs. The latter depends explicitly on the market structure. In a monopoly, a unique enterprise bears total costs as expressed by the continuously differentiable function $C(q_j, q_k)$. In a duopoly, total costs result from aggregating firms' costs as $C(q_j, q_k) = C_j(q_j) + C_k(q_k)$. Welfare is given by

$$W(q_j, q_k) = S(q_j, q_k) - C(q_j, q_k). \tag{2}$$

The first-best allocation consists in the output pair $\left(q_j^{FB}, q_k^{FB}\right)$ pinning down the marginal cost pricing efficiency point. Notice that the optimal provision of good $j$ is a decreasing function of the supply of product $k$ since $\frac{dq_j}{dq_k} = \frac{\partial P_j}{\partial q_k} / \left(\frac{\partial^2 C}{\partial q_j^2} - \frac{\partial P_j}{\partial q_j}\right) < 0$[1].

# 3 Multiproduct Monopoly

Let us suppose that products are provided by a monopolist exhibiting the previously described cost function. In what follows we will analyse different scenarios, taking outputs as the choice variables in the supplier's programme. We will characterize the corresponding allocations, according to whether decisions are inspired by the logic of business profitability or rather by the welfare-maximisation criterion that would be proper of a benevolent planner. Alternatively this scenario may be thought of as the one which would realize if a perfectly informed benevolent regulator were able to control one quantity. The resulting supply would then be socially optimal, conditional on market structure and unregulated output. Therefore, we will be equivalently speaking about the intervention of a social planner and that of a regulatory body. We will temporarily rule out informational asymmetries, an assumption to be relaxed in the second part of the paper.

## 3.1 Unconstrained Monopoly Problem

The profit-maximising monopolist chooses quantities by solving the programme

$$\underset{\{q_j, q_k\}}{Max} \left\{P_j(q_j, q_k)q_j + P_k(q_j, q_k)q_k - C(q_j, q_k)\right\}. \tag{3}$$

The optimality conditions reveal that the monopoly allocation is socially suboptimal: production is not large enough to entail first-best[2]. In particular, supply is adjusted so that the mark-up on, say, product $j$ equals the sum of own and cross price elasticity, a decision rule which allows the supplier to internalize the impact of both quantities on $p_j$. Indeed, defining the cross-elasticity as $\eta_{jk} \equiv -\left(\frac{\partial P_j}{\partial q_k}\frac{q_k}{p_j}\right)$, the Lerner index is given by

[1] Demand functions are decreasing in prices and, by the second-order condition, $\left(\frac{\partial^2 C}{\partial q_j^2} - \frac{\partial P_j}{\partial q_j}\right) > 0$.

[2] The first-order condition of (3) for, say, $q_j$ writes as $p_j - \frac{\partial C}{\partial q_j} = -\left(\frac{\partial P_j}{\partial q_j}q_j + \frac{\partial P_k}{\partial q_j}q_k\right)$. By the second-order condition of (2), $\left(p_j - \frac{\partial C}{\partial q_j}\right)$ is decreasing in $q_j$; hence, so is $-\left(\frac{\partial P_j}{\partial q_j}q_j + \frac{\partial P_k}{\partial q_j}q_k\right)$.

$$\frac{p_j - \frac{\partial C}{\partial q_j}}{p_j} = \eta_{jj} + \eta_{jk}. \tag{4}$$

The producer trades-off outputs against one another and finds it sub-optimal to vary both provisions in the same direction. Indeed, operation calls for $\frac{dq_j}{dq_k} < 0$ .

## 3.2 Unconstrained First and Second-Best Solution

The monopoly which were perfectly instructed to provide the socially optimal outputs would offer $\left(q_j^{FB}, q_k^{FB}\right)$ and price goods at marginal cost, eventually losing the fixed cost.

Suppose next that the producer is instructed to select one output, namely $q_r$, in a socially desirable way, whereas $q_u$ can be chosen in the private best interest. The subscripts $r$ and $u$ stay for *regulated* and *unregulated* respectively, a notation we will resort to whenever we will be speaking about one controlled and one uncontrolled output. Clearly, decisions about $q_u$ are made contingent on the choice of $q_r$, which the operator internalizes. Having a monopoly which produces according to this rule is equivalent to having an ideal planner who forces the supplier to implement

$$\begin{cases} q_r^{SB} \in \underset{q_r}{Argmax} \left\{ S(q_r, q_u(q_r)) - C(q_r, q_u(q_r)) \right\} \\ q_u^{SB} \in \underset{q_u}{Argmax} \left\{ \pi\left(q_r^{SB}, q_u(q_r^{SB})\right) \right\} \end{cases}. \tag{5}$$

The producer behaves as a follower *vis à vis* the regulator, who is able to optimize $q_r$ anticipating the former's strategy. The upscript $SB$ is affixed because outputs $\left(q_r^{SB}, q_u^{SB}\right)$ constitute a second-best allocation: in what follows it will be clear that some rents still exist on *both* markets, except under some specific circumstances. This is the case because only one instrument (output $q_r$) is available to foster social welfare. Provisions are still traded-off against one another. Indeed, the relation

$$\frac{dq_u}{dq_r} = -\left( \frac{p_r - \frac{\partial C}{\partial q_r}}{p_u - \frac{\partial C}{\partial q_u}} \right), \tag{6}$$

resulting from the first-order condition of (5), is negative. (6) reveals that it is never feasible to select $q_r$ such that marginal cost pricing entails on the unregulated market. Moreover, even for the controlled good, setting $p_r = \frac{\partial C}{\partial q_r}$ constitutes the optimal solution just in case $\frac{dq_u}{dq_r} = 0$, that is when output $u$ is inelastic to $q_r$. More generally, with $\frac{dq_u}{dq_r} \neq 0$, the supplier is provided an incentive to expand $q_u$ by letting the marginal welfare of $q_r$ be non-zero at the second-best equilibrium[3]. Intuitively, further offer of $q_r$ would be beneficial to society. Yet it is necessary to prevent the

[3] While the first-best solution satisfies the equality $\frac{\partial W}{\partial q_r} = 0$, the second-best one requires $\frac{\partial W}{\partial q_r} + \frac{\partial W}{\partial q_r}\frac{dq_u}{dq_r} = 0$. Provided that $\frac{\partial W}{\partial q_r} > 0$ and $\frac{dq_u}{dq_r} < 0$, we can only have $\frac{\partial W}{\partial q_r} > 0$.

commercialization of the additional units in order to support good $u$, which is also welfare-enhancing.

One more way to characterize this second-best environment is to look at the Lerner index for the controlled good as appearing in the Ramsey-type formula

$$\frac{p_r - \frac{\partial C}{\partial q_r}}{p_r} = (\eta_{uu} + \eta_{ur}) \left( -\frac{dq_u}{dq_r} \frac{p_u}{p_r} \right) {}^{4}. \tag{7}$$

(7) confirms that some rent remains even on the controlled production; indeed, the mark-up would be null only if so were $\frac{dq_u}{dq_r}$. Interestingly, it depends on the elasticities of $p_u$: since the operator does not make independent decisions, the gain obtained on the instrumented production needs be sized accounting for the impact on the free supply. Theoretically, the mark-up in (7) may exceed that in (4). This is the case when, from a social perspective, the quantity of good $r$ that is supplied at the market equilibrium is relatively too large; then increasing the associated monopoly rent provides an incentive to reduce $q_r$ and to enlarge $q_u$. Overall two different effects show up. As the price for the controlled commodity increases, the margin accruing to its producer rises while output reduces, which is detrimental to consumers. On the other hand, making good $r$ more expensive induces a substitution effect in favour of good $u$, which is increasing in the degree of substitutability and beneficial in terms of aggregate output. Therefore, the second-best mark-up depends on the quantity elasticity of the uncontrolled commodity as well as on the sensitivity of its price to variations in the controlled output scale.

Both in the first and in the second-best scenario, the solution allocation does not need to ensure budget balance for the operator. In the short run, the eventual presence of a fixed cost component is not relevant to the analysis. However, a firm that is unable to cover all its costs may be finally led to quit at least one market. Reasonably enough, it is not in the social interest to shut down a regulated monopoly since the activities it executes are likely to have poor substitution opportunities. It is necessary to let profits be non-negative, as we illustrate in the following Subsection.

## 3.3 Solutions under Break-Even Constraint

Firm's losses can be prevented by maximising (2) under the constraint that the producer breaks-even. Letting $\lambda \geq 0$ the associated multiplier, resulting conditions differ according to whether one or both quantities are chosen in a socially optimal way. In the former case, the Lerner index for good $j$ takes the familiar form

[4] Based on (8), we have replaced the mark-up on good $u$, the ratio $\left( \frac{p_u - \frac{\partial C}{\partial q_u}}{p_u} \right)$, with $(\eta_{uu} + \eta_{ur})$.

$$\frac{p_j - \frac{\partial C}{\partial q_j}}{p_j} = \left(\frac{\lambda}{1+\lambda}\right)(\eta_{jj} + \eta_{jk}). \tag{8}$$

In the latter case, moving to the notation previously illustrated, the mark-up for commodity $r$ writes as

$$\begin{aligned} \frac{p_r - \frac{\partial C}{\partial q_r}}{p_r} &= (\eta_{uu} + \eta_{ur})\left(-\frac{dq_u}{dq_r}\frac{p_u}{p_r}\right) + \\ &+ \left(\frac{\lambda}{1+\lambda}\right)\left[\eta_{rr} + \eta_{ru}\left(1 + \frac{dq_u}{dq_r}\frac{q_r}{q_u}\right) + \eta_{uu}\left(-\frac{dq_u}{dq_r}\frac{p_u}{p_r}\right)\right]. \end{aligned} \tag{9}$$

If compared to (7), (9) exhibits an additional term, weighed by $\left(\frac{\lambda}{1+\lambda}\right) < 1$. If the budget constraint is binding, then the optimal rent on good $r$ needs be augmented to account for the sensitivity of both $p_r$ and $p_u$ to variations in quantities. Rearranging (9) reveals that the term containing the elasticities of $p_r$ has a weight equal to $\frac{\lambda}{1+\lambda} < 1$, those embodying the elasticities of $p_u$ have weights $\frac{1+2\lambda}{1+\lambda} > 1$ and 1 respectively. Therefore, the Lerner index crucially depends on the indirect effects caused on the remunerativeness of the uncontrolled product. This suggests that, from the regulator's viewpoint, playing with product substitutability may constitute an additional instrument, which somehow compensates for the loss of control due to the possibility of regulating just one quantity.

# 4 Duopoly *à la* Cournot

Turning next to a multiple-agent scenario, we analyse a duopoly in which firms compete *à la* Cournot, each offering one of the imperfectly substitute products. The technology is represented by the continuously differentiable cost function $C_j(q_j), \forall j$.

## 4.1 Cournot Equilibrium

Suppliers simultaneously and independently solve the programme

$$\underset{q_j}{Max}\left\{P_j(q_j, q_k)\, q_j - C_j(q_j)\right\}, \forall j, k^5. \tag{10}$$

Firm $j$'s reaction function $q_j(q_k) = -\frac{\partial P_j}{\partial q_j} \Big/ \left(p_j - \frac{\partial C_j}{\partial q_j}\right)$ shows that, *ceteris paribus*, production decreases in marginal cost, hence it is positively related to technological

[5] We implicitly take the conditions for the existence of a pure-strategy equilibrium (i.e. concavity of the objective functions) as met. Under the stability condition $\left|\frac{\partial q_j(q_k)}{\partial q_k}\right| < 1$, uniqueness follows as well.

efficiency. Recalling the definition of $\eta_{jj}$, from the best reply we obtain

$$\frac{p_j - \frac{\partial C_j}{\partial q_j}}{p_j} = \eta_{jj}. \tag{11}$$

(11) suggests that, for any given level of the rival's production, duopolist $j$ offers the quantity at which mark-up and elasticity of $p_j$ to $q_j$ are equal. The rent does not exceed the one which would be extracted in a multiproduct price-setting context, as we have $\eta_{jj} < \frac{1}{\widehat{\varepsilon}_{jj}}$ (see Section 2). Moreover, the vector of Cournot payoffs belongs to the interior of the profit frontier and the resulting welfare distortion is not as large as the one which would realize, *ceteris paribus*, in a monopoly. Nevertheless producers are profiting from output under-provision. An aggregate increase would enhance welfare, though first-best quantities are not necessarily both larger than the Cournot ones. Optimal output distribution between competitors depends on technologies as well as on demand parameters[6].

## 4.2 Unconstrained Second-Best Solution

Suppose next that firm $r$ is instructed to provide the no-rent output

$$q_r^{SB} \in Arg \max_{q_r} \left\{ S\left(q_r, q_u\right) - C_r\left(q_r\right) - C_u(q_u) \right\}. \tag{12}$$

As in the monopoly case, we are faced with a second-best allocation in that the private condition $\left(\frac{p_u - \frac{\partial C_u}{\partial q_u}}{p_u}\right) = \eta_{uu}$ still holds. Due to the asymmetry in pursued objectives, marginal cost pricing occurs exclusively on market $r$, whereas the other production remains sub-optimal. In the limit case that the controlled competitor enjoys a significant technological advantage, good $u$ may even disappear: intuitively, firm $u$ is performing so badly, that it should not produce at all, whereas it actually does in the Cournot game[7]. In this event, the second-best solution restores efficiency by requiring shut down of the free provider's business and by letting a controlled monopoly produce the first-best output.

Interestingly, the current scenario is very similar in nature to that of a multi-product monopoly facing independent demands where, though unique, the producer is unable to internalize choices due to the absence of links between markets. From the social viewpoint, what may rather make a difference is that, while the monopolist executes both productions by means of a unique technology, the duopolists

[6] For instance, let us take $P_j(q_j, q_k) = \alpha - \beta q_j - \gamma q_k$ and $C_j(q_j) = c_j q_j + F$. Then, for $q_j^{FB}$ to be larger than $q_j^c$, we must have $\frac{\alpha - c_j}{3\beta\gamma} > \frac{\alpha - c_k}{2\beta^2 + \gamma^2}$. This condition does not allow for definitive conclusions.

[7] If the socially optimal reaction function, determined for firm $r$ by solving (12), lies above enterprise $u$'s reply function at all output levels, there is no crossing point of the two curves in the quantity-plane.

operate with separated (and possibly different) cost functions. Therefore, eventual economies of scale and/or scope are more exploitable under the former. However, this is hardly the case for the real-world industries the present study refers to; indeed, were the available economies very important, the recent introduction of competition would be hardly justified.

Observe that extracting the whole rent from the partially regulated monopoly is only possible with independent markets, whereas it is always feasible and optimal in the Cournot game. In this sense, whenever a regime of partial output control is available, a duopoly is preferable to the monopoly. However, this comes at a cost: due to the myopic behaviour of the Cournot agent, the benefit in terms of $q_r$ partially dissipates through the penalty imposed on $q_u$. In other words, the social planner has no way to account for and even indirectly affect the strategy governing the supply of good $u$.

## 4.3 Second-Best Solution under Break-Even Constraint

At the second-best solution, firm $r$ may not be able to break-even. The benevolent planner may preserve its long-run viability by selecting output compatibly with budget balance. The similarity with a monopoly serving independent demands remains. Indeed, using the multiplier $\lambda \geq 0$, the solution satisfies

$$\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_r} = \left( \frac{\lambda}{1+\lambda} \right) \eta_{rr}, \tag{13}$$

which coincides with (9) as soon as all output interaction terms in (9) are null. *Mutatis mutandis,* previous comments apply. Firstly, for future production to be ensured, firm $r$ needs to preserve a portion $\left(\frac{\lambda}{1+\lambda}\right)$ of the mark-up it would obtain by pursuing private interests; the larger the price reduction induced by output increments, the higher the Lerner index required for budget balancing. Secondly, despite welfare maximisation guides the choice of one output, it does not affect the strategy followed for the other, hinging on the property that the equilibrium is just *some* crossing point between reaction functions.

At this stage of our investigation, a key issue is whether the presence of a second provider playing Cournot is preferable to the absence of competition, whenever production is disciplined by a perfectly designed partial regulatory regime. A clear-cut conclusion in favour of either market structure may be hardly drawn without additional elements. However, it should be noticed that once the social programme is constrained by budget balance, the regulated duopolist may be induced to approach the allocation which, *ceteris paribus*, would be proper of a Stackelberg leader. A second policy phase might then begin, the first-mover being instructed to offer the second-best output.

# 5 Duopoly *à la* Stackelberg

The potential evolution from a Cournot to a Stackelberg market provides one reason for the latter to be studied. Additionally, it is highly realistic and combines features of the Cournot and monopolistic scenario while introducing the novelty of move sequentiality.

## 5.1 Stackelberg Equilibrium

The follower $f$ maximizes its profit *à la* Cournot, which involves $\left(\frac{p_f - \frac{\partial C_f}{\partial q_f}}{p_f}\right) = \eta_{ff}$. The leader $l$ anticipates and internalizes the rival's behaviour so that its mark-up

$$\frac{p_l - \frac{\partial C_l}{\partial q_l}}{p_l} = \eta_{ll} + \eta_{lf} \frac{dq_f}{dq_l} \frac{q_l}{q_f} \tag{14}$$

is a modified version of (4). In particular, in (14) $\eta_{ll}$ sums up to the cross elasticity as multiplicatively corrected by a term of quantity interaction. The presence of the latter is not innocuous. Indeed, having $\frac{dq_f}{dq_l} < 0$ implies the known result that, *ceteris paribus,* the mark-up of a Stackelberg leader is smaller than that of a monopolist and associated to a wider supply. Nevertheless it is characterized by positive rents.

## 5.2 Unconstrained Second-Best Solution

Let us suppose that the ideal planner can induce the leader to offer the socially optimal quantity. Considering the structure of the industries inspiring our investigation, it looks reasonable to assume that the regulator targets the first-mover, that is the former monopolist endowed with the largest market share even after competitors' entry. With our familiar notation, at the regulated firm's optimum we have

$$\frac{dq_u}{dq_r} = -\left(\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_u - \frac{\partial C_u}{\partial q_u}}\right), \tag{15}$$

which closely reproduces (6). In the monopoly scenario we could only establish that $\frac{dq_u}{dq_r} < 0$, here we are able to conclude something more by relying on the stability condition $\left|\frac{dq_u}{dq_r}\right| < 1$. At the second-best solution, the price of the controlled product cannot exceed its marginal cost to the same extent the rival good does, since it must be the case that $\left(p_r - \frac{\partial C_r}{\partial q_r}\right) < \left(p_u - \frac{\partial C_u}{\partial q_u}\right)$. This inequality is easily interpretable: as the margin on good $u$ increases, the condition becomes less stringent for the margin on good $r$ meaning that, in order to countervail the profitability effect which keeps the supply of $q_u$ excessively small, it is necessary to make the mark-up over $q_r$ relatively more attractive.

At this point we need to repropose a remark already made for the monopoly: exploiting the substitutability between quantities may help provide incentives for production balance, an instrument which is instead unavailable with a Cournot market structure. The additional benefit when firms play Stackelberg is that the social planner may refer to the stability condition as an order of magnitude for the correct use of this tool. Pushing further the parallel with the monopoly, the Lerner index for firm $r$

$$\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_r} = \eta_{uu} \left( -\frac{dq_u}{dq_r} \frac{p_u}{p_r} \right) \tag{16}$$

looks very close to the mark-up in (7), except that it only accounts for $\eta_{uu}$, rather than for $(\eta_{uu} + \eta_{ur})$ [8]. In other words, in the partially regulated Stackelberg game, the second-best mark-up for the controlled provider needs be proportional only to the quantity elasticity of the good sold by the rival; passing from (7) to (16), $\eta_{ur}$ disappears since the second production is now delegated to a myopic follower taking the choice of $q_r$ as exogenous.

In general, as long as a duopoly is not wasteful in terms of non-exploitable economies and duplication of fixed costs, a Stackelberg market yields higher welfare than a multiproduct monopoly. *Ceteris paribus*, both the mark-up on the free and on the controlled segment are lower. Indeed, the profitability of good $r$ is to be sized accounting for that of its (imperfect) substitute; therefore, in order to properly foster either provision, having a smaller margin on commodity $u$ allows to reduce also the rent on item $r$. On the other hand, the zero mark-up property of Cournot competition reappears on the Stackelberg regulated segment just in the event that demands are independent. But then two separated markets, exhibiting monopolistic structure, replace the genuine duopoly.

## 5.3 Second-Best Solution under Break-Even Constraint

In the precedent Subsection we have not been concerned with the budget balance issue for the regulated enterprise. As soon as the profit non-negativity constraint is added to the planner's programme, optimization with respect to the regulated quantity yields

$$\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_r} = \left( \frac{1}{1+\lambda} \right) \eta_{uu} \left( -\frac{dq_u}{dq_r} \frac{p_u}{p_r} \right) + \left( \frac{\lambda}{1+\lambda} \right) \left( \eta_{rr} + \eta_{ru} \frac{dq_u}{dq_r} \frac{q_r}{q_u} \right). \tag{17}$$

It is interesting to compare the second-best Lerner index in (17) to (9). Once again we are faced with the sum of two terms, having weights $\left(\frac{1}{1+\lambda}\right)$ and $\left(\frac{\lambda}{1+\lambda}\right)$

[8] Recalling that *de facto* the unregulated follower is a Cournot player, (20) has been obtained by replacing the ratio $\left( \frac{p_u - \frac{\partial C_u}{\partial q_u}}{p_u} \right)$ with $\eta_{uu}$.

respectively. Which one is larger depends on the value $\lambda$ takes. $\lambda \in (0,1)$ means that the constraint, though binding, is not very stringent and the associated shadow cost relatively small. Then $\left(\frac{1}{1+\lambda}\right) > \left(\frac{\lambda}{1+\lambda}\right)$ and so the terms containing the elasticity of good $r$ are relatively more important, suggesting that the constrained second-best mark-up on the controlled commodity is to be set mainly accounting for the sensitivity of its own price to either quantity. Conversely, $\lambda > 1$ realizes when the constraint is stringent; then it is relatively more relevant to consider the elasticity of $p_u$ to $q_u$ as well as the margin on market $u$. In particular, as $\left(\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_r}\right)$ is increasing in $\eta_{uu}$ with $-\frac{dq_u}{dq_r} > 0$, the larger the rent on good $u$, the higher the mark-up to be given up on good $r$. Market $u$ being highly profitable and production quite small, the corresponding $q_r$ is quite large and associated to a low mark-up. If the leader's technology includes significant fixed costs, then such a contained margin is likely not to suffice for break-even and it needs be augmented accordingly.

The nature of the programme recalls that of a monopoly, except for the differences in the mark-up formula following from the constraint. The planner has to ensure that the monopolist makes overall non-negative profits by keeping revenues from both goods at least as large as total costs. In a Stackelberg duopoly, break-even concerns only arise for the controlled supplier: the revenues emerging exclusively from market $r$ should fully cover the corresponding costs, as neither the demand for nor the costs of good $u$ are present in the constraint. Concerns about enterprise $u$ do not arise as it risks shut down only if significantly inefficient, in which case not going ahead with its production is actually in the social interest. This was already the case for the Cournot unregulated competitor but the follower is less exposed to losses since its profit needs be preserved non-negative in order to maintain $\left(\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_r}\right)$ as large as budget balance requires.

Ensuing conclusion is that, when regulating a Stackelberg leader under break-even constraint, the planner may rely on some kind of cross-subsidization between markets, by balancing output magnitudes (and mark-ups) through the control of a single quantity. It is known that *ceteris paribus* the Stackelberg market equilibrium makes consumers better off than the Cournot allocation and, *a fortiori*, the monopoly solution (the latter as long as technologies do not impose severe waste). In this perspective, it constitutes preferable starting point for a regulatory policy. Similar appreciation may be devoted to the Cournot game to the extent that the Cournot competitor easily evolves into a leader. Furthermore, the analysis so far performed seems to suggest that the Stackelberg scenario (possibly heritage of a former Cournot game) may offer more instruments and ensure wider manageability to a perfectly informed benevolent planner engaged in a partial regulatory regime. The latest result is promising in terms of the market outcomes attainable in some of the industrial sectors inspiring our investigation. Indeed, the leader-follower scenario well suites those ancient monopolies where the incumbent enterprise has preserved the first-mover advantage over the entrant(s).

# 6 Quantity Regulation: An Implementation Scheme

Real-world situations are generally characterized by the presence of informational asymmetries between regulatory bodies and enterprises. Gaps may concern several features, typical example being the technology and so the production costs, which are made publicly available only *ex post*. As a result, it is unlikely that the ideal policy so far presented can be implemented just by instructing the targeted firm to pursue social objectives. It is rather necessary to design some regulatory mechanism that is able to provide private incentives toward desired allocations. In what follows we propose a scheme which adapts to the plurality of environments analysed above. The regulated firm is required to repeatedly solve the private programme subject to a constraint embodying both profits and prices realized in the previous period of activity. The mechanism proves to be manageable as the authority only needs to access the book-keeping data published by the controlled enterprise. By preserving memory of the previously entailed performance, the process allows for progressive improvements in social welfare.

We will analyse the implementation of our scheme in the four basic scenarios we have referred to for the ideal policy. Firstly, we will investigate both a complete and a partial version as designed for the monopoly. We will subsequently move to the Cournot and Stackelberg games, for which we will only be concerned with the partial regime.

## 6.1 Complete Regulation of Monopoly Quantities

Let us suppose that, at each subsequent period $t$, the regulatory authority allows the monopolist to select a quantity pair $(q_j, q_k)$ within the set

$$R^t = \left\{ (q_j, q_k) \left| q_j p_j^{t-1} + q_k p_k^{t-1} - C(q_j, q_k) \geqslant \pi^{t-1}(1+\delta) \right. \right\}, \tag{18}$$

where $p_j^{t-1}$, $p_k^{t-1}$ and $\pi^{t-1}$ are, respectively, market prices and profits as resulting from the firm's accounting data published at the end of period $(t-1)$. Under (18) the firm chooses outputs such that, were the market prices of the previous period applied to the newly offered quantities, its profit would be at least as large as the one previously obtained, as corrected by the proportionality term $\delta$. The restriction on the feasible range involves no direct or explicit obligation in terms of production scale on either quantity. Nevertheless, we name this mechanism *complete* because past prices are applied to *both* outputs.

The regulator may not (and usually does not) know the function $C(\cdot)$, which is *ex ante* private information. She rather observes the prices realized on the market as well as the profit obtained by the enterprise at time $(t-1)$. The constraint actually encompasses both elements: the latter for restricting the producer's choices, the former for taking into account the technological conditions of production.

The $\delta$ parameter is the regulatory choice variable. Its introduction prevents the producer from just replicating its previous performance. To see this, suppose that the firm realizes a positive profit during the period preceding the first implementation; reasonably enough, it will be the maximum achievable profit associated to the optimal private quantities. Setting $\delta = 0$ provides the supplier with an obvious incentive to repeat the previous offers. On the opposite, as the aim is to induce larger provision of goods and lower prices, the regulator needs to impose $\delta > 0$ for the mechanism not to be trivial and to provoke desirable effects. Clearly $\delta$ should be properly selected; for instance, if it is too large, the constraint becomes excessively stringent and break-even is prevented. Moreover, as we will clarify at later stage, the stability of the emerging solution is perturbed. The benefits obtained during the early periods of implementation dissipates if the regulated supplier is no longer able to satisfy its obligations or if the process initially approaches desirable allocations, but is subsequently shifted away and finally diverges.

In what follows we will investigate how the process works. For the moment, we will not detail over the associated limiting properties and postpone the discussion about the convergence issue. We will temporarily concentrate on the effects caused on the monopoly performance and, when needed, take long-run convergence as a fact. We will resort to a similar strategy also for the environments analysed in next Subsections. Letting $\lambda^t$ the Lagrange multiplier of the regulatory constraint at period $t$ and optimizing with respect to $q_j$ yields

$$\left(\frac{p_j - \frac{\partial C}{\partial q_j}}{p_j}\right) + \lambda^t \left(\frac{p_j^{t-1} - \frac{\partial C}{\partial q_j}}{p_j}\right) = \eta_{jj} + \eta_{jk}. \tag{19}$$

Interestingly, the left-hand side of (19) is a weighed sum of two elements: the Lerner index, recurrently appearing in the Ramsey formulas, and some "modified" mark-up, determined by the price charged at period $(t-1)$ and weighed with the shadow cost of the constraint. Naturally enough, the monopolist is induced to reduce the mark-up according to the obligation cost. Once the right-hand side of (19) is taken to be fixed, the smaller $\lambda^t$ and/or $p_j^{t-1}$, the larger $\left(\frac{p_j - \frac{\partial C}{\partial q_j}}{p_j}\right)$ is allowed to be.

Implicit in the scheme design is the idea that, as time goes, quantities and prices adjust till a point where production cannot ulteriorly expand and prices just cover costs. Therefore, as the constraint becomes more and more binding, the progressive reduction effect on the value of $p_j^{t-1}$ is countervailed by an increase in $\lambda^t$, which prevents the growth of $\left(\frac{p_j - \frac{\partial C}{\partial q_j}}{p_j}\right)$. The current price $p_j$ is supposed to coincide with $p_j^{t-1}$ as soon as no further rent can be extracted. Then the mark-up becomes

$$\left(\frac{p_j - \frac{\partial C}{\partial q_j}}{p_j}\right) = \left(\frac{1}{1+\lambda^t}\right)(\eta_{jj} + \eta_{jk}), \tag{20}$$

which is similar to (8), except that the term $\left(\frac{\lambda}{1+\lambda}\right)$ is replaced by $\left(\frac{1}{1+\lambda^t}\right)$ embodying the per-period multiplier of the regulatory constraint[9]. Using the contour of the feasible set, namely $q_k = \frac{1}{p_k^{t-1}}\left[C(q_j, q_k) - p_j^{t-1} q_j + \pi^{t-1}(1+\delta)\right]$, we obtain

$$\frac{dq_k}{dq_j} = -\left(\frac{p_j^{t-1} - \frac{\partial C}{\partial q_j}}{p_k^{t-1} - \frac{\partial C}{\partial q_k}}\right). \tag{21}$$

This expression suggests that $\frac{dq_k}{dq_j} < 0$. Indeed, as the mechanism is meant to induce progressive price reduction over time, old prices (reasonably higher than the ones to appear in period $t$) are supposed to either exceed or equal current marginal costs[10]. This result reveals that, as one quantity increases, the second should decrease, for the regulatory constraint to be met. Hence, the structure of the constraint is well posed: (21) perfectly recalls (6), preserves the same sign for $\frac{dq_k}{dq_j}$ and only differs in that the slope of the set contour is here determined by the *lastly realized* prices together with the current marginal production costs. The regulator can be confident that the imposed obligation orients activity management in a socially desirable way.

As an illustration, we will hereafter consider the case where the technology is represented by the linear cost function $C(q_j, q_k) = c_j q_j + c_k q_k + F$, with $c_j > 0$, $c_k > 0$, $F > 0$. The loss of generality is somewhat compensated by the tractability of the adopted specification. Moreover, the linear stylization coupled with huge fixed costs is *de facto* backed by the existence of real-world providers (typically, in public utility sectors) exhibiting reasonably close technologies. Under these circumstances, feasible allocations are those belonging to the region above the lower contour of this set, a straight line of equation

$$q_k = \left(\frac{1}{p_k^{t-1} - c_k}\right)\left[\pi^{t-1}(1+\delta) + F - \left(p_j^{t-1} - c_j\right) q_j\right]. \tag{22}$$

Observe that, *ceteris paribus*, the vertical intercept is increasing in $\pi^{t-1}$ but

[9] We find it appropriate to perform the comparison with (8) because the producer should be guaranteed a non-negative profit when subject to the regulatory scheme.

[10] In particular, if the technology is preserved constant over periods, it must be the case that the conditions $p_i^{t-1} \geq p_i^t \geq \frac{\partial C}{\partial q_i}, \forall i = j, k$, are satisfied for the mechanism to correctly work. Otherwise, the constraint imposes losses to the operator and the risk of exit arises.

decreasing in $p_k^{t-1}$. Moreover the variations in the slope are determined by the previously realized prices, together with the present (and past, absent technological innovations) marginal costs. Hence, the admitted output area changes according to three main elements: the monopoly rent at $(t-1)$, the regulatory parameter $\delta$, the relative margin $\left(\frac{p_j^{t-1}-c_j}{p_k^{t-1}-c_k}\right)$.

The first two elements determine the position of the line in the plane. *Ceteris paribus* the intercept increases in $\delta$: the larger $\delta$, the wider the range of interdicted pairs close to the origin of the axes, but also the quicker and more significant reduction in both $\pi^{t-1}$ and $p_k^{t-1}$. Decreasing $\pi^{t-1}$ implies shifting the intercept closer to the origin, reducing $p_k^{t-1}$ provokes the opposite movement; overall, the former is likely to be larger (in absolute value) than the latter. This observation provides the regulator with a useful lesson: selecting $\delta$ small is preferable as it helps contain the impact linked to $\pi^{t-1}$, under which the forbidden area tends to shrink and the constraint to slack.

The third element identified above may even be innocuous. Take, for instance, $P_j(q_j, q_k) = \alpha - \beta q_j - \gamma q_k$. Then the ratio $\left(\frac{p_j^{t-1}-c_j}{p_k^{t-1}-c_k}\right)$ remains basically constant over time. One way to roughly verify this claim is to simulate the monopoly optimization procedure under the regulatory constraint, by taking the unconstrained performance as the starting scenario. In Table 1 we collect some of the results obtained by assuming the following parameter values: $\alpha = 400$, $\beta = 25$, $\gamma = 10$, $c_j = 8$, $c_k = 15$, $F = 700$, $\delta = 0.1$. In the row associated to Period 0 we report the unconstrained monopoly solution to be inserted into the constraint for Period 1. Figures provide a consistent example of the evolution under the regime. As they show, in a completely linear environment where marginal costs never change, the same optimal relative margin evaluated at past prices realize in all periods. Hence, no variation occurs in the slope of the contour line as time goes.

| | $q_j$ | $q_k$ | $p_j$ | $p_k$ | $\pi$ | $W$ | $\left(\frac{p_j^{t-1}-c_j}{p_k^{t-1}-c_k}\right)$ |
|---|---|---|---|---|---|---|---|
| *Period 0* | 5.6667 | 5.4333 | 204 | 207.5 | 1456.6 | 1534.9 | - |
| *Period 1* | 6.0494 | 5.8004 | 190.7603 | 194.4967 | 1446.7 | 2675.6 | **1.018** |
| *Period 2* | 6.4570 | 6.1911 | 176.6637 | 180.6519 | 1414.6 | 2814.7 | **1.018** |
| .... | .... | .... | .... | .... | .... | .... | .... |
| *Period 8* | 9.5815 | 9.1870 | 68.5926 | 74.5107 | 427.2918 | 3510.1 | **1.018** |
| .... | .... | .... | .... | .... | .... | .... | .... |

Table 1. Evolution of the regulatory process in a linear monopoly environment

Unfortunately, this element of predictability (and so of controllability) is not a robust property. For instance, it is missing with a convex technology of the kind

$C(q_j, q_k) = c_j q_j^2 + c_k q_k^2 + F$, in which case the constraint identifies the area within the ellipse of equation

$$\frac{\left(q_j - \frac{p_j^{t-1}}{2c_j}\right)^2}{\left[\frac{\left(p_j^{t-1}\right)^2}{4c_j^3} + \frac{\left(p_k^{t-1}\right)^2}{4c_k^2 c_j} - \frac{\pi^{t-1}(1+\delta)+F}{c_j}\right]} + \frac{\left(q_k - \frac{p_k^{t-1}}{2c_k}\right)^2}{\left[\frac{\left(p_j^{t-1}\right)^2}{4c_j^2 c_k} + \frac{\left(p_k^{t-1}\right)^2}{4c_k^3} - \frac{\pi^{t-1}(1+\delta)+F}{c_k}\right]} = 1. \qquad (23)$$

(23) suggests that, as the previously charged prices reduce, the centre tends to move toward the origin of the Cartesian axes, though break-even requirements prevent from getting too close. Indeed, given the technology, prices must be larger enough than marginal costs. On the other hand, the countervailing effects linked to past prices and profits reappear and determine the size of the feasible set. A reduction in $p_j^{t-1}$ and $p_k^{t-1}$ makes it smaller, a decrease in $\pi^{t-1}$ causes the opposite effect. Recalling that the ellipse tends to re-centre closer to the origin at each subsequent period, we would like its interior to remain sufficiently, though not excessively, big: like this it would still include relatively large outputs, but possibly exclude too little ones. Observe that keeping $\delta$ small may do the job in a double direction: it slows down the speed at which the movement of the centre occurs and contains (but does not annihilate) the shrinking effect induced by the reduction in $\pi^{t-1}$. Provided that, under convexity of the cost function, too big outputs may be sub-optimal, this may suffice to a desirable result.

As a conclusion, the variable $\delta$ constitutes the truly driving instrument for the global process. Whether the technology is linear or convex, a simple and robust criterion seems to be available for the mechanism to be appreciably managed.

## 6.2 Partial Regulation of Monopoly Quantities

Suppose next that, at each period $t$, the regulatory constraint embodies only one of the previously realized prices, together with $\pi^{t-1}$. The feasible set is given by

$$R^t = \left\{(q_r, q_u) \mid q_r p_r^{t-1} + P_u(q_r, q_u)q_u - C(q_r, q_u) \geqslant \pi^{t-1}(1+\delta)\right\}. \qquad (24)$$

Observe that in (24) $q_u$ is multiplied by its current market price. This asymmetry in the constraint structure is meant to implement a partial policy which regulates the production of $q_r$ and preserves $q_u$ uncontrolled. Though the restriction is not directly targeted to the former output, some degree of freedom is left in terms of $q_u$ by allowing $P_u(\cdot)$ to enter the inequality. This is the correct way to interpret both the adjective *partial* and the *regulated/unregulated* dichotomy.

As the complete regulatory scheme, also the partial one has a well posed structure. Indeed, along the set contour, one has $\frac{dq_u}{dq_r} = -\frac{\left(p_r^{t-1} - \frac{\partial C}{\partial q_r} + \frac{\partial P_u}{\partial q_r} q_u\right)}{\left(p_u - \frac{\partial C}{\partial q_u} + \frac{\partial P_u}{\partial q_u} q_u\right)}$, but operation always calls for $\frac{dq_u}{dq_r} < 0$[11]. On the other hand, the mechanism works differently from the second-best policy of Section 3. It does not establish a leader-follower relationship between regulator and firm. This hinges on the nature of the constraint, which does not impose any specific obligation on the provision of $q_r$ to be internalized in the decisions about $q_u$. The producer adjusts the two quantities so that the largest possible profit is achieved and the regulatory requirement met. As a result, the Lerner index for good $r$ is again a weighed sum of standard and "modified" mark-up (of the kind presented in (19)) with the property that, once $p_r$ approaches $p_r^{t-1}$, it recovers the familiar structure

$$\left(\frac{p_r - \frac{\partial C}{\partial q_r}}{p_r}\right) = \left(\frac{1}{1+\lambda^t}\right)\eta_{rr} + \eta_{ru}. \tag{25}$$

In (25) only the own elasticity of the regulated good is corrected by the term expressing the shadow cost of the constraint, the cross elasticity appearing with unitary weight. On the other hand, for good $u$ we are faced with

$$\left(\frac{p_u - \frac{\partial C}{\partial q_u}}{p_u}\right) = \left(\frac{1}{1+\lambda^t}\right)\eta_{ur} + \eta_{uu}, \tag{26}$$

where $\lambda^t$ enters multiplicatively through the elasticity of $p_u$ to the regulated output.

(25) is very simple in structure. The same could not be said for (9), which included several terms deriving from the role of follower played by the enterprise *vis à vis* the regulator. The similarity between constrained socially optimal mark-up and Lerner index under the mechanism is no longer evident. Formulas suggest that the mark-up charged over the uncontrolled product is restricted by the obligation imposed on the second commodity because the price of the former depends on the regulated output. Therefore, the shadow cost operates through the sensitivity of $p_u$ to variations in the supply of good $r$.

Notice that the left-hand side ratio in (26) decreases in the value of the multiplier; however, as $q_r$ enlarges under the regulatory constraint, also $\eta_{ur}$ may increase, provided that so does the ratio $\frac{q_r}{p_u}$[12]. Even if the constraint imposes a significant penalty, calling for $\lambda^t$ large, the monopoly may be able to keep $\left(\frac{p_u - \frac{\partial C}{\partial q_u}}{p_u}\right)$ quite

[11] The optimality condition for $q_u$ may be written as $(1+\lambda^t)\left(\frac{\partial P_u}{\partial q_u}q_u + p_u - \frac{\partial C}{\partial q_u}\right) = -\frac{\partial P_r}{\partial q_u}q_r$. Provided that $\lambda^t \geq 0$ and that the right-hand side of the equality is positive, it must be the case that $\left(\frac{\partial P_u}{\partial q_u}q_u + p_u - \frac{\partial C}{\partial q_u}\right) > 0$. Moreover, under convergence of $p_r$ to $p_r^{t-1}$, we analogously have $(1+\lambda^t)\left(p_r^{t-1} + \frac{\partial P_u}{\partial q_r}q_u - \frac{\partial C}{\partial q_r}\right) = -\frac{\partial P_r}{\partial q_r}q_r$. For similar reasons, $\left(p_r^{t-1} - \frac{\partial C}{\partial q_r} + \frac{\partial P_u}{\partial q_r}q_u\right) > 0$.

[12] As an example, one may again consider the linear demand function $P_u(q_r, q_u) = \alpha - \beta q_u - \gamma q_r$. Then we have $\eta_{ur} = \gamma\left(\frac{q_r}{\alpha - \beta q_u - \gamma q_r}\right)$, which is increasing in $q_r$ as the derivative $\frac{\partial \eta_{ur}}{\partial q_r} = \frac{\gamma^2 q_r}{(\alpha - \beta q_u - \gamma q_r)^2} + \frac{\gamma}{(\alpha - \beta q_u - \gamma q_r)}$ is evidently a positive quantity.

high and simultaneously satisfy its obligations, to the extent that the value of $\eta_{ur}$ overwhelms that of the multiplier.

It might be the case that the production of good $u$ reduces with respect to the unregulated quantity; hence, it is *a fortiori* small if compared to the optimal size. This does not mean that social welfare decreases by executing production under the regulatory regime, in which case implementation would be senseless. On the opposite, the scheme allows for an overall improvement through its impact on $q_r$. To see this, let us compare (25) to (26). In (25) $\eta_{ru}$ is likely to be small to the extent that the regulation induces small values for $q_u$ and $p_r$; on the other hand, $\eta_{rr}$ is probably large but divided by $\left(1+\lambda^t\right)$ which may also be large. Everything considered, the mark-up on good $r$ should significantly decrease with respect to its original size. In (26) the shadow cost of the constraint multiplies the cross rather than the own elasticity of $p_u$; provided that prices are relatively more sensitive to variations in own quantities, the negative impact caused on $q_u$ is likely to be less important than the positive incentive on $q_r$. Therefore, it is reasonable to expect the constraint to improve welfare as long as it is not too stringent.

For illustrative purpose, we will rely on the linear cost and demand functions already adopted. Then the contour of the allowed region identifies as

$$\begin{aligned} q_r \;&=\; -\frac{\beta}{\gamma}q_u + \frac{1}{\gamma}\left[\alpha - c_u - \frac{\beta}{\gamma}\left(p_r^{t-1}-c_r\right)\right] + \\ &+\left\{\frac{\left(\frac{p_r^{t-1}-c_r}{\gamma}\right)\left[\alpha - c_u - \frac{\beta}{\gamma}\left(p_r^{t-1}-c_r\right)\right] - \left[\pi^{t-1}(1+\delta)+F\right]}{\gamma q_u - \left(p_r^{t-1}-c_r\right)}\right\}. \end{aligned} \tag{27}$$

The only relevant arm of (27) is the one located below the horizontal and above the oblique asymptote in the $(q_u, q_r)-$ plane. As the price of the regulated good reduces, it tends to become narrower all along the evolution of the process[13]. Moreover the point at which the curve crosses the horizontal Cartesian axis, that is $\frac{\pi^{t-1}(1+\delta)+F}{p_r^{t-1}-c_r}$, progressively moves rightward (and so diverges from the origin) at the speed impulsed by $\delta$. The consequences imposed on the production region provide a reason for smoothing the obligation by setting $\delta$ not excessively large. Indeed, if $\gamma q_u - \left(p_r^{t-1}-c_r\right) > 0$, then legitimate allocations are those which lie to the left of the curve; whenever $\gamma q_u - \left(p_r^{t-1}-c_r\right) < 0$, the operative area moves to the right. The latter is the more desirable event since, for the former to occur, $p_r^{t-1}$ must be so low that one has $\gamma q_u > p_r^{t-1} - c_r$, despite the partial scheme impulses a progressive reduction in $q_u$. However, at this point, even the unconstrained monopoly solution may belong to the authorized set, meaning that the regulatory regime fails.

[13] This is so because the horizontal asymptote moves downward and gets closer to the horizontal Cartesian axis. On the other hand, the oblique asymptote shifts parallelly upward as the intercept of the line decreases in $p_r^{t-1}$.

In other words, the regulator should preserve the price of the controlled commodity sufficiently large, and so contain the growth of $q_r$, in order to ensure the correct functioning of the mechanism. As a desirable by-product, this requirement contributes to limit the decrease in the supply of commodity $u$ through good substitutability.

## 6.3 Partial Regulation of Quantities in a Cournot Duopoly

In the asymmetric scenario where duopolist $r$ is subject to

$$p_r^{t-1} q_r - C_r(q_r) \geqslant \pi_r^{t-1}(1+\delta) \tag{28}$$

and, in the long-run, $p_r$ tends to $p_r^{t-1}$, one easily finds the standard Lerner index

$$\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_r} = \left(\frac{1}{1+\lambda^t}\right) \eta_{rr}. \tag{29}$$

As for the completely regulated monopoly, the claim that the scheme closely implements the constrained second-best policy is supported by the similarity between (29) and (13). Once again, $\left(\frac{\lambda}{1+\lambda}\right)$ is replaced by $\left(\frac{1}{1+\lambda^t}\right)$ to account for the per-period multiplier. Under the regime, producer $r$ offers more than $q_r^c$. The competitor simultaneously reduces $q_u$ along the reaction function, clinging to the rule $\left(\frac{p_u - \frac{\partial C}{\partial q_u}}{p_u}\right) = \eta_{uu}$.

The independence of production decisions, proper of the Cournot game, reappears through the characteristics of the set contour. More precisely, the latter is independent of the realization of $q_u$ in period $t$, which shows up through the linearity in the $(q_u, q_r)-$ plane. Hence, some given level of regulated output is identified, whatever the value of $q_u$, a property which is robust to variations in the technology of the regulated producer. For instance, with a linear cost function, the allowed range for the regulated output is given by $\left[\frac{\pi_r^{t-1}(1+\delta)+F}{p_r^{t-1}-c_r}, +\infty\right)$. The ratio $\frac{\pi_r^{t-1}(1+\delta)+F}{p_r^{t-1}-c_r}$ increasing over time, the consented output pairs contain progressively larger values of the regulated quantity. This is so because the denominator reduces relatively more than the numerator as, in the latter, the quantity increase partially countervails the decrease in $p_r^{t-1}$.

Figure 1 depicts the family of firm $r$'s isoprofits in a linear environment with the parameter values already adopted for Table 1. Each vertical line represents the lower extreme of the region allowed by the current constraint; as time passes, the position moves rightward and a larger size progressively entails for $q_r$, as coupled with a relatively small reduction in $q_u$. Hence, Figure 1 shows the evolution of the realized output pairs. In particular, the red curve represents the isoprofit on which the supplier would operate if unregulated. The intersection between black isoprofit and black vertical line determines the allocation for Period 1; the blue crossing point refers to the tenth Period, the green one to the twentieth. The small value chosen for $\delta$ smooths the path. The pink combination of outputs, the one at which the whole rent has been extracted from the regulated operator, preserves thereafter. As we will clarify at later stage, this allocation identifies a fixed point to which, under appropriate conditions, the system stably converges over time.

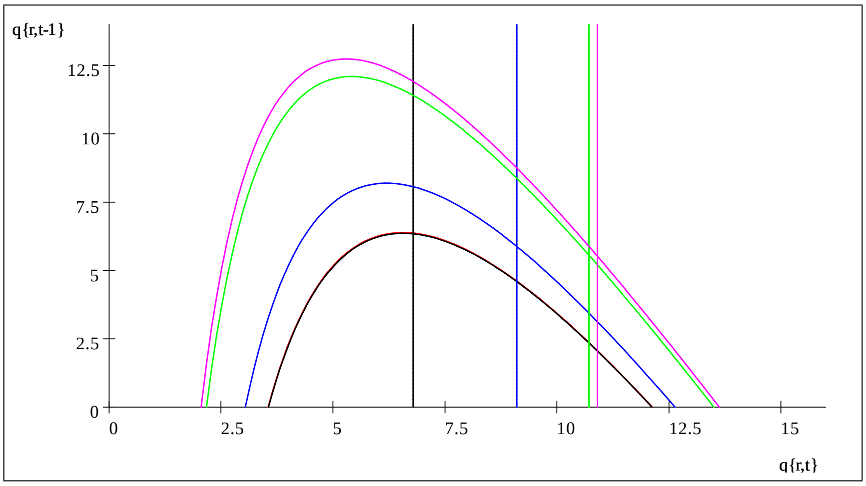


Figure 1. Implementation of the partial scheme in the linear Cournot environment

Interestingly, depending on the value of $q_r$ which identifies the lower extreme of the consented interval, the regulatory mechanism may induce the firm, initially playing *à la* Cournot, to acquire the quantity leadership on the market as well as to preserve it at least in the early implementation stages. This is not a concern as endowing one producer with the first-player advantage already involves an improvement upon the original situation, through an expansion of aggregate output.

## 6.4 Partial Regulation of Quantities in a Stackelberg Duopoly

We will hereafter conclude our investigation by looking at the regulatory scheme as imposed to the Stackelberg leader. Since $R^t$ writes as in a Cournot duopoly, once again the constraint is not a function of the unregulated supply and exhibits the same linear contour in the quantity-plane, whatever the technology used by enterprise $r$. On the other hand, the regulated Lerner index, in the limit, becomes

$$\frac{p_r - \frac{\partial C_r}{\partial q_r}}{p_r} = \left(\frac{1}{1+\lambda^t}\right)\left(\eta_{rr} + \eta_{ru}\frac{dq_u}{dq_r}\frac{q_r}{q_u}\right). \tag{30}$$

It is interesting to parallel (30) to (17). The term $\left(\eta_{rr} + \eta_{ru}\frac{dq_u}{dq_r}\frac{q_r}{q_u}\right)$ involving the elasticities of $p_r$ appears in both formulas, the difference residing in that the multiplicative ratio translates from $\left(\frac{\lambda}{1+\lambda}\right)$ into $\left(\frac{1}{1+\lambda^t}\right)$. The term containing the own elasticity of $p_u$, which adds up to the previous one in (17), is missing in (30).

This is the case because $P_u(q_r, q_u)$ appears neither in the profit function nor in the constraint of the regulated enterprise, implying that $\eta_{uu}$ has no bite in the solution. It is only relevant in the condition $\left(\frac{p_u - \frac{\partial C}{\partial q_u}}{p_u}\right) = \eta_{uu}$, still codifying the follower's behavioural rule. Intuition suggests that the leader under regulation is prevented from fully internalizing the competitor's strategy by the need to meet obligations. It does not become as myopic as a Cournot duopolist because its offer still accounts for the change induced in the price of its product by the strategic variation caused in the rival's supply. What is no longer considered is the impact of the change in $q_u$ on $p_u$. In other words, the scheme makes the Stackelberg leader one-eyed so that it behaves as a slightly more sophisticated Cournot player. The ensuing long-run supply of good $r$ should finally exceed $q_r^{SB}$, as deducible from the comparison between mark-up formulas. However, to the extent that the growth in regulated output is just partially compensated by the reduction in unregulated offer, both aggregate production and social welfare increase, as in the Cournot duopoly.

Our analysis of the second-best policy has highlighted that the perfectly informed social planner disposes of more instruments if faced with market structures, such as the Stackelberg duopoly, where the operator subject to partial regulation somehow internalizes the choice of the second offered output. Implementing the scheme here proposed in a Stackelberg game, the uninformed regulator gives up the possibility of exerting an indirect control on the unregulated supply through the regulated quantity, an authority loss which shows up in the leader's Lerner index. Nevertheless, as the mechanism makes the trajectories followed by the regulated production very similar in the simultaneous and sequential games, the properties characterizing either of them remain (nearly) valid for the second scenario as well. One such property is illustrated in the following Subsection.

## 6.5 Magnitude of $\delta$ and Convergence of the Process

In the analysis performed so far, we have repeatedly derived one crucial instruction for the regulatory authority: whatever the market structure, the parameter $\delta$ should be chosen of small magnitude. Due to the complexity of the formulas, in most of the cases we have not been analytically rigorous at showing the advantages of such a criterion, though we have been arguing in its favour by looking at specific scenarios. We will hereafter investigate one final case for which a fully analytical conclusion can be derived, by considering a duopoly *à la* Cournot where the usual linearity assumption is adopted. This effort of stylization will return a tractable problem to go through, the main result of which has already been announced in Subsection 6.3; it will help clarify how important the magnitude of $\delta$ is as to the convergence of the process. Furthermore, concentrating the investigation on (and providing an analytical solution for) a duopolistic context should not be viewed as a limitation. Recalling that we mainly refer to newly competitive sectors that used to be monopolies, we are confident that the analysis is well suited.

By convergence we mean that the mechanism iteration should finally lead to some allocation exhibiting both efficiency and stability properties. Firstly, aggregate output should be closer to the socially optimal quantity than the original market supply used to be. Secondly, the achieved allocation should constitute a dynamically stable equilibrium in the sense that the system gets there and no subsequent perturbation impulses it away. In the language of our study, this amounts to having at least one stable fixed point in the function $q_r^t\left(q_r^{t-1}\right)$, whereas $\delta$ is to be viewed as a (constant) adjustment coefficient.

Supposing that the regulatory constraint is binding, we firstly express the quantity to be chosen in period $t$ in terms of the output offered in the previous period

$$q_r^t\left(q_r^{t-1}\right)=(1+\delta)\,q_r^{t-1}-\left[\frac{F\delta}{\alpha-c_r-\frac{\gamma}{2\beta}\left(\alpha-c_u\right)-\left(\beta-\frac{\gamma^2}{2\beta}\right)q_r^{t-1}}\right]. \tag{31}$$

As (31) exhibits a discontinuity in the neighborhoods of the vertical asymptote, the assumptions of Brower Fixed Point Theorem are not met, except if attention is restricted to a range over which $q_r^t$ is continuous. Since a fixed point can exclusively exist on the left arm, we concentrate on this part of the curve, which surely identifies a suitable support. Figure 2 shows several graphs of $q_r^t\left(q_r^{t-1}\right)$ as drawn for the same values of demand and cost parameters we adopted for Table 1.

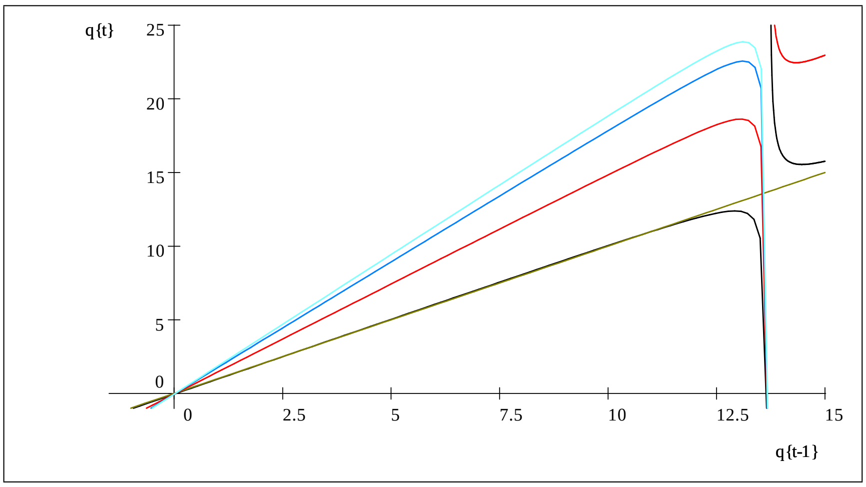


Figure 2. Graph of $q_r^t(q_r^{t-1})$ as $\delta$ varies

In the present case, we let $\delta$ vary and associate one curve to each value it takes, namely 0.02, 0.05, 0.5, 0.8 and 0.9, in order to illustrate the impact of the regulatory choice. The continuous interval we refer to is the one included between the crossing points with the horizontal axis, both belonging to the positive quarter. As $\delta$ grows, the support slightly shrinks and the curve becomes higher and higher since the oblique asymptote gets progressively steeper. On each plotted curve we identify two fixed points, namely $q_{r,f1}^{t-1}$ and $q_{r,f2}^{t-1}$. As one can check, their coordinates are independent of $\delta$: the system seems to encompasse a pair of allocations it is naturally driven to under the regulatory correction. In this sense, the fact that the magnitude of $\delta$ contributes to determine the size of the relevant support remains innocuous.

The value of the adjustment parameter is instead crucial at making one of such points stable. Tedious calculations reveal that stability of the larger fixed point calls for

$$|1+\delta(1-h)|<1 \Longleftrightarrow -\frac{2}{\delta}<(1-h)<0, \tag{32}$$

where $h \equiv \frac{\beta F\left(2\beta^2-\gamma^2\right)}{\left[\frac{1}{2\sqrt{2}}s-\frac{1}{2\sqrt{2}}\sqrt{s^2-8\beta F\left(2\beta^2-\gamma^2\right)}\right]^2}$ and $s \equiv [2\beta(\alpha-c_r)-\gamma(\alpha-c_u)]$. The regulator can surely find a suitable value for $\delta$: following the familiar "little magnitude" criterion seems to suffice, despite the informational gap about the cost parameters $h$ depends upon. Observe that $q_{r,f2}^{t-1}$ identifies the pink-colored allocation depicted in Figure 1.

One can similarly prove the instability of $q_{r,f1}^{t-1}$, which identifies a case of regulatory lack of power. This weakness is *de facto* of negligible importance. Firstly, provided that $q_r^t(q_{r,f1}^{t-1})=q_{r,f1}^{t-1}$ is not part of a desirable equilibrium as smaller than $q_r^c$, managing an inefficient regulatory tool is not a concern when the result it is supposed to yield does not improve upon the market outcome. Secondly, if the initial period's constraint embodies the Cournot price and profit, the system is immediately impulsed toward some $q_r^t > q_r^c \equiv q_r^{t-1}$. Furthermore, $q_{r,f2}^{t-1}$ necessarily exceeds $q_r^c$; hence, convergence to $q_{r,f2}^{t-1}$ starts at the very beginning and process reversion is prevented by the stability property.

Overall our analysis reveals that only one of the two intrinsic fixed points may constitute a stationary allocation for the trajectory designed by the quantity under control. *Conditio sine qua non* for this adjustment path to be followed is the appropriate use of the unique regulatory instrument when powerful: the regulated system may be expected to approach the wanted equilibrium conditionally on setting $\delta$ adequately small.

# 7 Conclusion

Our investigation has been inspired by the former monopolies recently opened up to competition, whose dominant firms remain subject to regulatory obligations, whereas one or more entrants operate unregulated. We have observed that, if one of the outputs were supplied by a provider perfectly instructed to maximise social welfare, then the emerging mixed allocations would improve upon the uncontrolled market equilibria.

In the light of this result, we have proposed a multi-period scheme of partial regulation under which the enterprise repeatedly solves its programme subject to a constraint embodying prices and profits of the previous period. As based on the supplier's official book-keeping data, this mechanism is manageable and rules out serious concerns about the informational gap between authority and agent. Moreover, though it does not necessarily replicate the theoretical conditions characterizing the ideal social policy, it generally improves upon the private performance if properly managed. In particular, under the *little magnitude* criterion for the instrument $\delta$, the process converges to some long-run allocation exhibiting both efficiency and stability properties: firstly, it is characterized by larger aggregate supply; secondly, it constitutes a dynamically stable equilibrium.

# References


Biglaiser, G. and C. A. Ma (1995), "Regulating a Dominant Firm: Unknown Demand and Industry Structure", *The Rand Journal of Economics*, 26(1): 1-19

Caillaud, B., R. Guesneri, P. Rey and J. Tirole (1988), "Government Intervention in Production and Incentives Theory: A Review of Recent Contributions", *The Rand Journal of Economics*, 19(1): 1-26

Cremer, H., M. Marchand and J. F. Thiesse (1989), "The Public Firm as an Instrument for Regulating an Oligopolistic Market", *Oxford Economic Papers*, 41: 283-301

Cremer, H., M. Marchand and J. F. Thiesse (1991), "Mixed Oligopoly with Differentiated Products". *International Journal of Industrial Organisation*, 9: 43-53

De Fraja, G. and E. Iossa (1998), "Price caps and Output Floors: A Comparison of Simple Regulatory Rules", *The Economic Journal*, 108: 1404-1421

Finsinger, J. and I. Vogelsang (1979), "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms", *The Bell Journal of Economics*, 10(1): 157-171

Garvie, D. and R. Ware (1996), "Public Firms as Regulatory Instruments with Cost Uncertainty", *The Canadian Journal of Economics*, 29(2): 357-378

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