

Optimal Caps under Uncertainty

Luca Di Corato

Università Ca' Foscari Venezia

Yishay D. Maoz

The Open University of Israel

February 2020

Abstract

In markets where production has adverse externalities, policy makers may wish to increase welfare by imposing a cap on market entries. In this study, we examine the implications that the cap has on the firms' entry equilibrium policy and on social welfare in the presence of market uncertainty. In contrast with previous literature, we explicitly consider the presence of an externality which is a convex function of aggregate quantity in the market and then let the social planner choose the cap level maximizing welfare. We find that the planner's optimal policy is either to ban any further market entries, or allow more entries until market quantity reaches a certain cap. The likelihood that the cap option will be preferred, and the size of the optimal cap, increase in profit uncertainty, decrease in the speed by which the external cost grows faster than the private cost, and decrease in the market quantity already installed when the decision on the cap is taken.

Key words: Investment, Uncertainty, Caps, Competition, Externalities, Welfare.

JEL codes: C61, D41, D62

Corresponding author: Yishay D. Maoz, Department of Management and Economics, The Open University of Israel, 1 University Road, Raanana, Israel. Email address: yishayma@openu.ac.il.

Luca Di Corato, Dipartimento di Economia, Università Ca' Foscari Venezia, Fondamenta San Giobbe 873, 30121, Venezia, Italy. Phone: +390412347668. Email address: luca.dicorato@unive.it.

1. Introduction

The introduction of a cap on market entries is a quite common policy option for regulators dealing with issues arising in the context of industry policy, international trade and foreign investment, environmental damages and resources control.

One of the main reason why caps are used is that whenever production costs are not fully internalized by firms, a cap on the aggregate industry output may be welfare-increasing. This may be the case when, for instance, excessive entries cause welfare losses due to the presence of fixed setup costs and of a “*business stealing effect*” (Mankiw and Whinston, 1986). Further, the penetration of foreign firms negatively affects the performance of a domestic industry¹ (Bartolini, 1995) or when a negative environmental externality, such as global warming, pollution, or exhaustion of natural resources, is associated with production (Katsoulacos and Xepapadeas, 1995; Fujiwara, 2009).

Bartolini (1993) was the first study investigating how the presence of cap affects the industry equilibrium within a dynamic decentralized setting characterized by uncertain market conditions and irreversible investment prior to production. He shows that the introduction of the cap leads to investment dynamics that differ from those portrayed by the literature on irreversible investment under uncertainty. This literature, summarized by Dixit and Pindyck (1994), has shown that, when no cap is present, firms invest (in order to enter the market) sequentially. Further, due to the uncertainty characterizing future profits and the irreversibility of the initial investment, firms invest only when the output price is sufficiently above the marginal cost of production.

¹ See also Chao and Eden (2003) on the adoption of foreign investment quotas and Calzolari and Lambertini (2007) on the use of voluntary export restraints.

Differently, Bartolini (1993) shows that when a cap on the aggregate industry output is present, firms enter the market sequentially only up to a certain aggregate quantity, then a “*competitive run*” starts and exhausts at once the residual entry slots. During the run, output is sold at a price below its marginal cost of production. This is because firms fear that they may lose their entry option while postponing investment in order to wait for higher prices. The run reduces welfare for two reasons; i) as it brings on the market additional output at once and not sequentially and ii) as the market good is sold at a price below its marginal cost of production.

In Bartolini (1993, 1995), the introduction of the cap is justified by the welfare gains that could be achieved by restricting private economic actions. However, the welfare analysis in these papers is not conclusive since the level of the cap is taken as exogenous and no external effects associated with firms’ investment and production are explicitly considered. This gap has been recently filled by Di Corato and Maoz (2019) where adverse production externalities are included in the welfare objective and the cap is set endogenously. They show that a welfare-maximizing regulator has only two alternative policies when it comes to set the cap: i) immediately banning further market entries by setting the cap at the currently supplied market quantity, or ii) have no cap at all. The choice depends on the level of market uncertainty characterizing the firm’s profits. In fact, when the output price triggering firms’ entry is, due to the consideration of the market uncertainty, above the social marginal cost of production, no cap should be introduced in that further entries pay welfare gains. In contrast, when the “*uncertainty premium*” does not suffice to counterbalance the externality, banning further market entries is preferable in that these entries would occur at a price below the social marginal cost.

A key assumption in Di Corato and Maoz (2019) is that the external cost and private cost grow at sufficiently close rates as the industry output grows. This enabled the possibility that the uncertainty premium dominates the externalities regardless of industry output. As far as, up to empirical evidence, this may not always apply, it becomes of interest to study the complementary case where the external cost rises faster than the private cost as the industry output grows.

Thus, in this paper, we change perspective and aim to complete their analysis by i) investigating the circumstances under which having a finite cap is optimal and ii) investigating the properties of the cap level and in particular how it is affected by market uncertainty and by the extra-speed by which the external cost grows faster than the private cost.

In our model, mostly following Di Corato and Maoz (2019), we consider both a scenario where firms may competitively enter the market and a scenario where entries are rationed by distributing licenses when the cap is announced. In addition, we introduce an external cost function quadratic in the aggregate industry output alongside a linear private cost function.

We then determine the optimal entry policy set by private firms acting in a decentralized setting and the optimal level of the cap to be set by a welfare-maximizing regulator.

Our main findings are as follows. We identify the circumstances in which a finite cap is optimal. In this respect, irrespective of the scenario considered, the key element is the gap between the marginal external cost and the marginal surplus gains. If at the current market quantity, in response to an additional entry, the external cost would grow

more than the surplus, a ban on further market entries should be imposed. Otherwise, a finite cap above the current market quantity should be set.

In addition, we show under both scenarios that the cap is increasing in the level of uncertainty. This confirms, the counterbalancing effect that, in the presence of an externality, the uncertainty premium may have.

We also find, when comparing the two scenarios, that when uncertainty goes to infinity, having no cap at all is a policy applying only when entries are rationed. The insight behind this result is subtle. Rationing the right to enter is crucial, since, otherwise, we would observe massive runs producing huge losses as firms enters the market at a price level below the social marginal cost of production. This happens in that the increase in the scarcity rents (due to the higher market uncertainty) dominates the effect of relaxing the entry constraint by setting a higher cap. Therefore, as the higher scarcity rents makes entry more attractive, the number of firms involved in the run increases.

The paper remainder is as follows. In Section 2, we present our model set-up. In Section 3, we determine the industry equilibrium under competitive entry. We present the optimal entry policy, characterize and discuss the emergence of a competitive run and determine the welfare-maximizing cap. Section 4 considers the scenario where the right to enter is rationed. We determine and discuss the optimal entry policy and the welfare-maximizing cap. Section 7 concludes.

2. The basic model

Within a continuous time setting, we consider a competitive industry comprised of a large number of firms producing a certain good. The following assumptions

characterize the industry and the formation of the industry equilibrium:

Ass 1. The market demand at each point in time is given by:²

$$(1) \quad P_t = X_t / Q_t,$$

where P_t and Q_t denote the price and quantity of the good at time t while X_t represents the state of demand.

Ass 2. The state of demand, X_t , evolves randomly over time according to the following Geometric Brownian Motion:

$$(2) \quad dX_t = \mu \cdot X_t \cdot dt + \sigma \cdot X_t \cdot dZ_t,$$

where μ and σ are constant parameters which measure the drift and the volatility of X_t , respectively, and dZ_t is the increment of a standard Wiener process with $E(dZ_t) = 0$ and $E[(dZ_t)^2] = dt$.

Ass 3. All firms are identical and their production capacity, denoted ΔQ , is infinitesimally small with respect to the market.³ Due to their infinitesimally small size, all firms are price-takers.

Ass 4. Each firm rationally forecasts the future evolution of the whole market.

² The use of a hyperbolic demand function is common in consumer theory and solidly supported by empirical evidence (see for instance Lambertini, 2010).

³ The assumption of infinitesimally small agents is pretty standard in infinite horizon models investigating the competitive industry equilibrium in a dynamic setting. See for instance Jovanovic (1982), Hopenhayn (1992), Lambson (1992), Leahy (1993), Dixit and Pindyck (1994, Ch. 8), Bartolini (1993, 1995) and Moretto (2008).

Ass 5. Market entry is free and an idle firm may decide to enter the market at any time t .

Ass 6. Each firm, when entering the market, commits to offer permanently one unit of A.

Ass 7. All firms face the same cost structure in which, at each time t , producing one unit of the good entails an instantaneous cost equal to $M > 0$.

Ass 8. At each time t , production has an external cost for society that firms do not incur equal to:

$$(3) \quad E(Q) = e \cdot Q^2 / 2.$$

Ass 9. Due to the welfare-harming externality, the government sets a cap, \bar{Q} , on the market size.

Ass 10. Firms are risk-neutral and r denotes the rate at which they discount future payoffs.⁴ As standard in the literature, we assume that $r > \mu$ to secure the convergence of the firm's value.⁵

⁴ Note that introducing risk aversion would not change our results. Its introduction would simply require the development of the analysis under a risk-neutral probability measure. See Cox and Ross (1976) for further details.

⁵ See Appendix A where the value of an active firm is determined.

Last, the following three remarks are in order:

- i. by Assumptions 6 and 7, at society level, the instantaneous total cost associated with the production of a generic quantity Q_t is:

$$(4) \quad S(Q) = M \cdot Q + E(Q),$$

that is, aggregate private production cost plus external cost.

- ii. by Itô's lemma, when Q_t is unchanged, the evolution of P_t is governed by the following Geometric Brownian Motion:

$$(5) \quad dP_t = \mu \cdot P_t \cdot dt + \sigma \cdot P_t \cdot dZ_t;$$

the present value of the flow of private production costs, M/r , and the present value of the flow of social production costs, $S(Q)/r$, may equivalently be viewed as the irreversible investment cost to be paid when entering the market at firm and at society level, respectively. This is to say that our framework may be used to study both the case of a costs associated with periodic production and the case of a lump-sum costs arising when setting up the firm.

3. Industry equilibrium and optimal cap under free entry

Under the setup just introduced, our model of competitive industry equilibrium is a specific case of the more general one analyzed in Bartolini (1993, 1995).

At the generic time t , each idle firm has to decide whether to enter the market in order to produce and supply an additional of the good or not. Given the current Q_t , entry will

take place only when the price P_t is high enough to pay, once accounted for the private production cost, an expected net present value not lower than zero.⁶ Further, the presence of a cap generates, by limiting market entries, scarcity rents that firms wish to cash. Cashing these rents is of course conditional on entering the market before the cap becomes binding. Therefore, as one may intuitively anticipate, the entry policy set at firm level must keep into account the risk of being left out.

3.1 Optimal entry policy

Let $V(Q, X)$ be the value of a firm active in the market. By a standard no-arbitrage analysis,⁷ this value is equal to:

$$(6) \quad V(Q, X) = Y(Q) \cdot X^\beta + \frac{X}{Q \cdot (r - \mu)} - \frac{M}{r}$$

where $\beta > 1$ is the positive root of the quadratic equation:

$$(6.1) \quad (1/2) \cdot \sigma^2 \cdot x \cdot (x-1) + \mu \cdot x - r = 0.$$

Note that the second and third terms of (6) represent the expected present value of the flow of the firm's future profits conditional on Q remaining forever at its current level, i.e.:

$$E \left[\int_{t=0}^{\infty} (P - M) \cdot e^{-rt} \cdot dt \right] = \frac{X}{Q \cdot (r - \mu)} - \frac{M}{r}$$

⁶ In the following, we will drop the time subscript for notational convenience.

⁷ See Appendix A.

Therefore, the first term in (6) accounts for how market entries will affect, by increasing the supplied market quantity Q , the value of the firm.

Now, let us consider an idle firm contemplating entry and denote the entry threshold function by $X^*(Q)$. As well known, due to free entry, the value of its option to wait is null. In fact, as entry is an attractive opportunity for other firms as for itself, the firm may lose it by postponing its entry (see Dixit and Pindyck, 1994, Ch. 8, pp. 256-258).

Following Bartolini (1993, 1995), the industry's equilibrium can be determined by using the following *Value Matching Condition*:

$$(7) \quad v[Q, X^*(Q)] = 0,$$

which, as one can immediately see, represents a zero-profit condition at the entry, and the following boundary condition:⁸

$$(8) \quad v_Q[Q, X^*(Q)] = 0.$$

Note that (7) and (8) are not optimality conditions. They should hold for any $X^*(Q)$ since they simply require that, given a certain threshold, the no-arbitrage condition set on the value of the firm holds. This in turn implies that, once taken the derivative with respect to Q on both sides of (7),

⁸ Bartolini (1993) proves the existence of Condition (8) in *Proposition 1* of his article. See Di Corato and Maoz (2019) for further explanations.

$$(9) \quad \frac{dV[Q, X^*(Q)]}{dQ} = V_Q[Q, X^*(Q)] + V_X[Q, X^*(Q)] \cdot \frac{dX^*(Q)}{dQ} = 0.$$

Substituting (8) into (9) yields the following condition:

$$(10) \quad V_X[Q, X^*(Q)] \cdot \frac{dX^*(Q)}{dQ} = 0$$

For (10) to hold, one should have either $V_X[Q, X^*(Q)] = 0$ or $dX^*(Q)/dQ = 0$. In the first case, that is, when the so-called *Smooth Pasting Condition*, i.e., $V_X[Q, X^*(Q)] = 0$ holds, the investment threshold function is:

$$(11) \quad X^*(Q) = \bar{\beta} \cdot (r - \mu) \cdot \frac{M}{r} \cdot Q,$$

where $\bar{\beta} \equiv 1 + \frac{1}{\beta-1} > 1$. The second case applies when the *Smooth Pasting Condition* does not hold. To understand how this case should be handled, we recall that the term $Y(Q)$ represents how future entries affect the value of the firm. Hence, when Q is already at the cap, as no further entries can occur

$$(12) \quad Y(\bar{Q}) = 0.$$

By (12), (6) and (7), when $Q = \bar{Q}$, the investment threshold is:

$$(13) \quad \bar{X} = (r - \mu) \cdot \frac{M}{r} \cdot \bar{Q}.$$

As it can be easily verified, the *Smooth Pasting Condition* does not hold in \bar{Q} and, by continuity, it also does not hold within the interval $[\tilde{Q}, \bar{Q})$. Thus, as $dX^*(Q)/dQ = 0$, \tilde{Q} is the solution of the following equation:

$$(14) \quad X^*(\tilde{Q}) = \bar{X},$$

that is:

$$(15) \quad \tilde{Q} = \bar{Q} / \bar{\beta}.$$

Summing up, the optimal entry policy at firm level is:

- when $Q < \tilde{Q}$, entry occurs whenever the process $\{X_t\}$ reaches the threshold $X^*(Q)$. In *Figure 1* below this is described by the move from point E to point F; This threshold is an increasing function of Q since the larger the market quantity, the higher competition and, ceteris paribus, the higher the expected profitability required for triggering further entries. Note that, by the Marshallian rule, a firm should enter as long as $X > (r - \mu) \cdot \frac{M}{r} \cdot Q$. Hence, the term $\frac{1}{\beta-1}$ in $\bar{\beta}$ is the correction needed in order to take into account the presence of uncertainty (see e.g. Dixit and Pindyck, 1994, Ch. 5, Section 2). Further,
- the higher the volatility of the state of demand, the higher the threshold triggering firm's entry since $dX^*(Q)/d\sigma > 0$, which in expected terms implies that new entries are delayed;

- the higher the trend of the state of demand, the higher the threshold triggering firm's entry since $dX^*(Q)/d\alpha > 0$, which in expected terms implies that new entries are delayed;
 - the higher the entry cost, i.e. M/r , the higher the threshold triggering firm's entry since $dX^*(Q)/d(M/r) > 0$, which in expected terms implies that new entries are delayed.
- if $Q = \tilde{Q}$, when X hits the threshold $X^*(\tilde{Q}) = \bar{X}$ a new entry occurs but, in contrast with what happens in the interval $Q < \tilde{Q}$, the threshold does not increase in response to the higher Q as $dX^*(Q)/dQ = 0$ within the interval $[\tilde{Q}, \bar{Q}]$. Thus, firms keep entering the market until the cap is reached. In *Figure 1* below these dynamics are described by the move from point G to point H. It is worth stressing that, in this case contrast with what happens within the interval $Q < \tilde{Q}$, the entry threshold \bar{X} does not take into account for the presence of uncertainty. This is because its effect on the timing of entry is counterbalanced by the threat of being left out. Thus, firms prefer entering earlier setting a lower threshold. In fact, note that $\bar{X} \leq X^*(Q)$ in the interval $[\tilde{Q}, \bar{Q}]$. Last, note that $\lim_{\bar{Q} \rightarrow \infty} \tilde{Q} = \infty$. This means that in the absence of a cap, i.e. $\bar{Q} \rightarrow \infty$, no run occurs and entries proceed sequentially following the entry policy illustrated by (11).

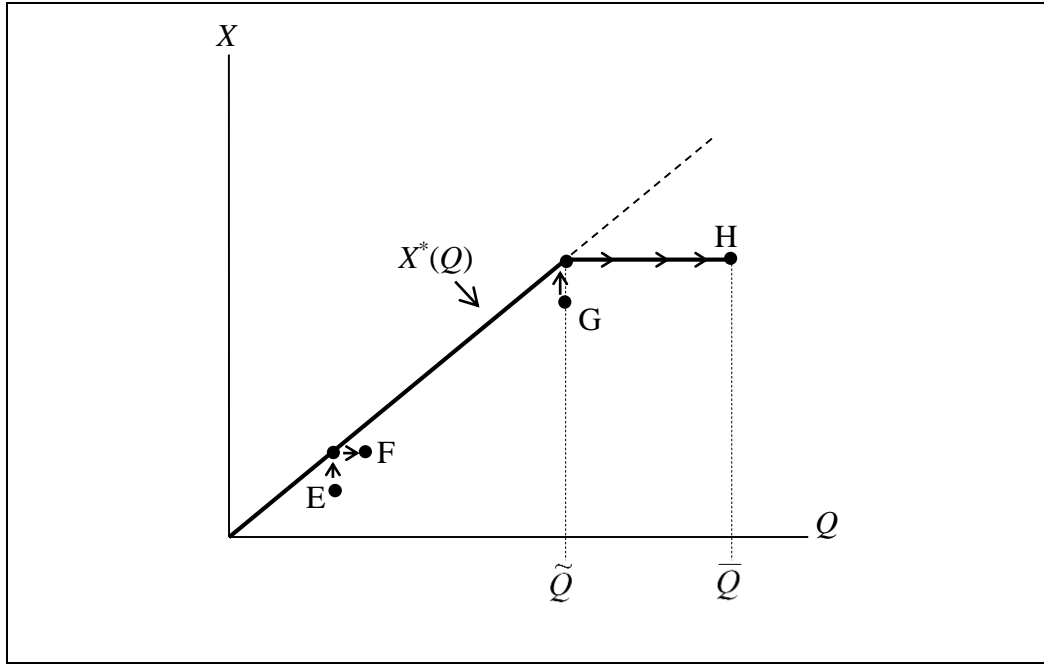


Figure 1: Investment dynamics with a cap and free entry.

3.2 Welfare and optimal cap under free entry

The entry policy set at firm level does not account for the negative externality that each firm's production will generate. Hence, aiming to trade it off with the net welfare gains associated with a higher quantity of the good on the market, the government has set a cap, \bar{Q} , on the market size.

The expected discounted social welfare, given the current levels of X and Q , is:⁹

$$(16) \quad w(Q, X, \bar{Q}) = c(Q, \bar{Q}) \cdot X^\beta + \frac{X \cdot \ln(Q)}{r - \mu} - \frac{M \cdot Q + e \cdot \frac{Q^2}{2}}{r},$$

In (16), the second and the third term represents the expected present value of the net surplus flow associated with the current market quantity Q , that is, the surplus resulting

⁹ This function can be derived by following the same procedure illustrated in detail in Appendix A (for the value of an active firm).

from the supply of those units minus the social cost associated with their production while the first term stands for the expected present value of the net surplus flow associated with future market entries.

As follows from (16), maximizing the welfare function $W(Q, X, \bar{Q})$ with respect to \bar{Q} is equivalent, as the other elements of $W(Q, X, \bar{Q})$ do not contain \bar{Q} , to maximizing $C(Q, \bar{Q})$ with respect to the same variable. In order to find the optimal level of \bar{Q} , two different ranges of values must be considered, i.e.

- if \bar{Q} is set sufficiently close to the current level of Q , specifically – within the range $Q \leq \bar{Q} \leq \bar{\beta} \cdot Q$, then $\tilde{Q} \leq Q$ and as soon as the process $\{X_t\}$ reaches the threshold \bar{X} a run up to the cap will be ignited.
- If \bar{Q} is set sufficiently far from the current level of Q , specifically – in the range $\bar{Q} > \bar{\beta} \cdot Q$, then $Q < \tilde{Q}$ and the next increments in Q (as long as $Q < \tilde{Q}$) are according to the incremental and gradual entry policy indicated by (11).

In the next sections, first, we determine the two branches of the function $C(Q, \bar{Q})$, i.e.

$$C(Q, \bar{Q}) = \begin{cases} C^L(Q, \bar{Q}) & \text{if } Q \leq \bar{Q} \leq \bar{\beta} \cdot Q \\ C^R(Q, \bar{Q}) & \text{if } \bar{Q} > \bar{\beta} \cdot Q \end{cases}$$

and then, we maximize $C(Q, \bar{Q})$ with respect to \bar{Q} , to determine the optimal cap.

3.2.1 $C(Q, \bar{Q})$ when $Q \leq \bar{Q} \leq \bar{\beta} \cdot Q$

In the first range, $Q \leq \bar{Q} \leq \bar{\beta} \cdot Q$, the function $C(Q, \bar{Q})$ is found by the condition:

$$(17) \quad w(Q, \bar{X}, \bar{Q}) = \frac{\bar{X} \cdot \ln(\bar{Q})}{r - \mu} - \frac{M \cdot \bar{Q} + \frac{e \cdot \bar{Q}^2}{2}}{r},$$

which states that in that range Q is above \tilde{Q} and if the threshold \bar{X} is hit then a run brings the quantity to the cap at once, no more changes in Q shall take place, and welfare is therefore the present value of the welfare flow based on \bar{Q} .

Evaluating (16) at \bar{X} , comparing it to (17), applying (13) and simplifying, yields that in the range $Q \leq \bar{Q} \leq \bar{\beta} \cdot Q$ the function $C(Q, \bar{Q})$ is given by:

$$(18) \quad C^L(Q, \bar{Q}) = \bar{\beta}^\beta \cdot K \cdot \frac{M \cdot \bar{Q} \cdot \ln\left(\frac{\bar{Q}}{Q}\right) - M \cdot (\bar{Q} - Q) - \frac{e}{2} \cdot (\bar{Q}^2 - Q^2)}{\bar{Q}^\beta},$$

where:

$$(18.1) \quad K \equiv \frac{r^{\beta-1}}{M^\beta \cdot \bar{\beta}^\beta \cdot (r - \mu)^\beta} > 0.$$

Note from (18) that if the cap is set at its lowest possible level, i.e., at the current level of Q , then $C^L(Q, \bar{Q}) = 0$.

Analyzing $C(Q, \bar{Q})$ as captured by (18), reveals that there are many different patterns it can take within its definition range, $Q \leq \bar{Q} \leq \bar{\beta} \cdot Q$. Specifically, depending on parameter values and the value of Q , it can be monotonically decreasing, have a u-

shape, or even have a more intricate pattern with a local minimum point with a local maximum point to its right. These patterns will be shown in the following sections.

3.2.2 $C(Q, \bar{Q})$ when $\bar{Q} > \bar{\beta} \cdot Q$

 We now turn to the case where \bar{Q} is set in the range $\bar{Q} > \bar{\beta} \cdot Q$. The point where the two ranges connect is at $Q = \tilde{Q} = \frac{1}{\bar{\beta}} \cdot \bar{Q}$, and applying this value of Q in (18) yields:

$$(19) \quad C^R(\tilde{Q}, \bar{Q}) = \bar{\beta}^\beta \cdot K \cdot \frac{\beta \cdot \ln(\bar{\beta}) \cdot M - M - e^{\frac{2 \cdot \beta - 1}{2 \cdot \beta} \cdot \bar{Q}}}{\beta \cdot \bar{Q}^{\beta - 1}}.$$

In the considered range $\bar{Q} > \bar{\beta} \cdot Q$, as $Q < \tilde{Q}$, the changes in Q are incremental additions occurring whenever the process $\{X_t\}$ hits the threshold function $X^*(Q)$. Therefore, at a generic $Q < \tilde{Q}$, the expected discounted social welfare, given the current level of X , is:

$$(20) \quad W(Q, \bar{Q} | X) = C^R(Q, \bar{Q}) \cdot X^\beta + \frac{X \cdot \ln(Q)}{r - \mu} - \frac{M \cdot Q + e \cdot \frac{Q^2}{2}}{r}.$$

In order to determine the term $C^R(Q, \bar{Q})$, we impose that, at the threshold $X^*(Q)$, the following *Value Matching Condition* holds:

$$(21) \quad W_Q[Q, \bar{Q}, X^*(Q)] = 0$$

Substituting (20) in (21), partially differentiating with respect to Q , using (11), and rearranging terms, yields:

$$(22) \quad C_Q^R(Q, \bar{Q}) = K \cdot \left[\frac{e}{Q^{\beta-1}} - \frac{M}{(\beta-1) \cdot Q^\beta} \right],$$

Integrating on both sides of (22) up to \tilde{Q} and using (19) yields:

$$(23) \quad C^R(Q, \bar{Q}) = K \cdot \left[\frac{M}{(\beta-1)^2 \cdot Q^{\beta-1}} - \frac{e}{(\beta-2) \cdot Q^{\beta-2}} \right] + G(\bar{Q}),$$

where:

$$(23.1) \quad G(\bar{Q}) = \bar{\beta}^\beta \cdot K \cdot \left[\frac{e}{\bar{Q}^{\beta-2}} \cdot \frac{1}{2 \cdot \beta \cdot (\beta-2)} - \frac{M}{\bar{Q}^{\beta-1}} \cdot \frac{g(\beta)}{\beta-1} \right],$$

with:

$$(23.2) \quad g(\beta) \equiv 1 - (\beta-1) \cdot \ln(\beta).$$

In appendix D of Di-Corato and Maoz (2019) it is shown that $0 < g(\beta) < 1$ throughout the definition range of β , namely, $\beta > 1$.

Differentiating (23) with respect to \bar{Q} yields:

$$(24) \quad C_{\bar{Q}}^R(Q, \bar{Q}) = G'(\bar{Q}) = \bar{\beta}^\beta \cdot K \cdot \left[\frac{M}{\bar{Q}^\beta} \cdot g(\beta) - \frac{e}{2 \cdot \beta \cdot \bar{Q}^{\beta-1}} \right],$$

implying that $C^R(Q, \bar{Q})$ is an inverse u-shaped function of \bar{Q} which is maximized when the cap is set at:

$$(25) \quad \bar{Q} = \frac{2 \cdot \beta \cdot M \cdot g(\beta)}{e} \equiv \bar{Q}^{opt}.$$

Note that if the current level of Q is sufficiently large then \bar{Q}^{opt} is not within the definition range of $C^R(Q, \bar{Q})$, i.e., within the range $\bar{Q} > \bar{\beta} \cdot Q$. In that case, $C^R(Q, \bar{Q})$ is decreasing in \bar{Q} throughout its definition range, implying that the optimal choice of \bar{Q} is not within this range, but within the range $Q \leq \bar{Q} \leq \bar{\beta} \cdot Q$. Specifically, by (25) this occurs when:

$$(25.1) \quad Q > \frac{1}{\bar{\beta}} \cdot \bar{Q}^{opt} = \frac{2 \cdot (\beta - 1) \cdot M \cdot g(\beta)}{e} \equiv Q_1.$$

Thus, if $Q < Q_1$ then \bar{Q}^{opt} is within the definition range of $C^R(Q, \bar{Q})$ and brings it to a local maximum within that range. *Proposition 1* presents the conditions for $C^R(Q, \bar{Q})$ to have a positive value in that point.

Proposition 1: *If $Q < Q_1$ then within its definition range $C^R(Q, \bar{Q})$ is an inverse u-shape function, maximized by at \bar{Q}^{opt} and satisfying $C^R(Q, \bar{Q}^{opt}) > 0$ if and only if one of the following conditions hold:*

$$(a) \quad \beta < \beta_1 \approx 1.3$$

$$(b) \quad \beta \geq \beta_1 \text{ and } Q < Q^* < Q_1 \text{ where, } Q^* \text{ is the unique root that the equation } C^R(Q, \bar{Q}^{opt}) = 0 \text{ has within the range } Q < Q_1.$$

Proof: See Appendix B.

3.2.3 $C(Q, \bar{Q})$ throughout its entire definition range

As described above, $C^R(Q, \bar{Q})$ and $C^L(Q, \bar{Q})$ meet at the point where $\bar{Q} = \bar{\beta} \cdot Q$. The contact between them at that point is a "super contact" in the sense that their derivatives with respect to \bar{Q} are also equal to one another. This can be seen from differentiating $C^L(Q, \bar{Q})$, as captured by (18), with respect to \bar{Q} , applying $\bar{Q} = \bar{\beta} \cdot Q$ and comparing the result to (24) evaluated at $\bar{Q} = \bar{\beta} \cdot Q$ too.

Also note that at this meeting point the value of these derivatives is:

$$(26) \quad C^L_{\bar{Q}}(Q, \bar{\beta} \cdot Q) = C^R_{\bar{Q}}(Q, \bar{\beta} \cdot Q) = K \cdot \frac{M \cdot g(\beta) - \frac{e}{2 \cdot (\beta - 1)} \cdot Q}{Q^\beta}.$$

From (26) it follows that the two functions meet at a point where they both rise if and only if, Q is sufficiently small.

As the following proposition shows, the following two possibilities emerge for each combination of the model parameters:

- If Q is sufficiently small then the optimal cap is the constant \bar{Q}^{opt} , captured by (25), and some incremental entry dynamics take place before the run.
- If Q is sufficiently large then the optimal cap is at the current level of Q , and no further entry takes place.

Alongside these possibilities there is a third one that occurs if and only if $\beta < \beta_1 \approx 1.3$ and Q is not too small or too large but between two specific values specified later. This is the possibility that the optimal cap is within the range where $C^L(Q, \bar{Q})$ represents

$C(Q, \bar{Q})$ and the entry dynamics only contain a run but no incremental entry before it.

We denote the optimal cap in that case by $\bar{Q}^L(Q)$ and show that it is a decreasing function of Q satisfying $Q < \bar{Q}^L(Q) < \bar{Q}^{opt}$.

Proposition 2:

(a) If $\beta \geq \beta_1 \approx 1.3$ then there exist a certain $Q^* < \frac{1}{\beta} \cdot \bar{Q}^{opt}$ such that:

(i) If $Q < Q^*$ then the optimal cap is \bar{Q}^{opt}

(ii) If $Q \geq Q^*$ then it is optimal to set the cap at Q ,

where Q^* is the unique root that $C^R(Q, \bar{Q}^{opt}) = 0$ has within the range $0 < Q < \frac{1}{\beta} \cdot \bar{Q}^{opt}$

(b) If $\beta < \beta_3$ then there exists a certain $\frac{1}{\beta} \cdot \bar{Q}^{opt} < Q^{**} < \bar{Q}^{opt}$, such that:

(i) If $Q < Q_1$ then the optimal cap is \bar{Q}^{opt}

(ii) If $Q_1 \leq Q < Q^{**}$ then the optimal cap is $\bar{Q}^L(Q)$, where $\bar{Q}^L(Q)$ is a decreasing function of Q satisfying $0 < \bar{Q}^L(Q) < \bar{Q}^{opt}$

(iii) If $Q \geq Q^{**}$ then it is optimal to set the cap at Q ,

where Q^{**} is the unique level of Q for which $C^L(Q, \bar{Q}^L) = 0$.

Proof: See Appendix C.

The remaining of this section is based on a numerical example with the following parameter values: $K = 1$, $\beta = 1.8$ (implying $\bar{\beta} = 2.25$), $M = 100$, $e = 10$. Figures 2, 3, and 4 show $C^R(Q, \bar{Q})$, $C^L(Q, \bar{Q})$ and the resulting $C(Q, \bar{Q})$ for different levels of the initial quantity, Q . In each of these three figures:

- The blue dashed line shows $C^L(Q, \bar{Q})$ for each positive \bar{Q} and not merely within its definition range.
- The blue dashed line shows $C^R(Q, \bar{Q})$ for each positive \bar{Q} and not merely within its definition range.
- The straight black line shows the resulting $C(Q, \bar{Q})$

We start with Figure 1 which refers to the case where $Q = 10$. For this relatively high level of Q , the function $C^L(Q, \bar{Q})$ decrease for each \bar{Q} , and in particular within its definition range which lies between $\bar{Q} = Q = 10$ to $\bar{Q} = \bar{\beta} \cdot Q = 22.5$. This implies that, due to the super contact property, $C^L(Q, \bar{Q})$ meets $C^R(Q, \bar{Q})$ when the latter is decreasing in \bar{Q} , i.e., to the left of its peak. Consequently $C(Q, \bar{Q})$ is decreasing for all $\bar{Q} > Q$ and maximized at $\bar{Q} = Q$.

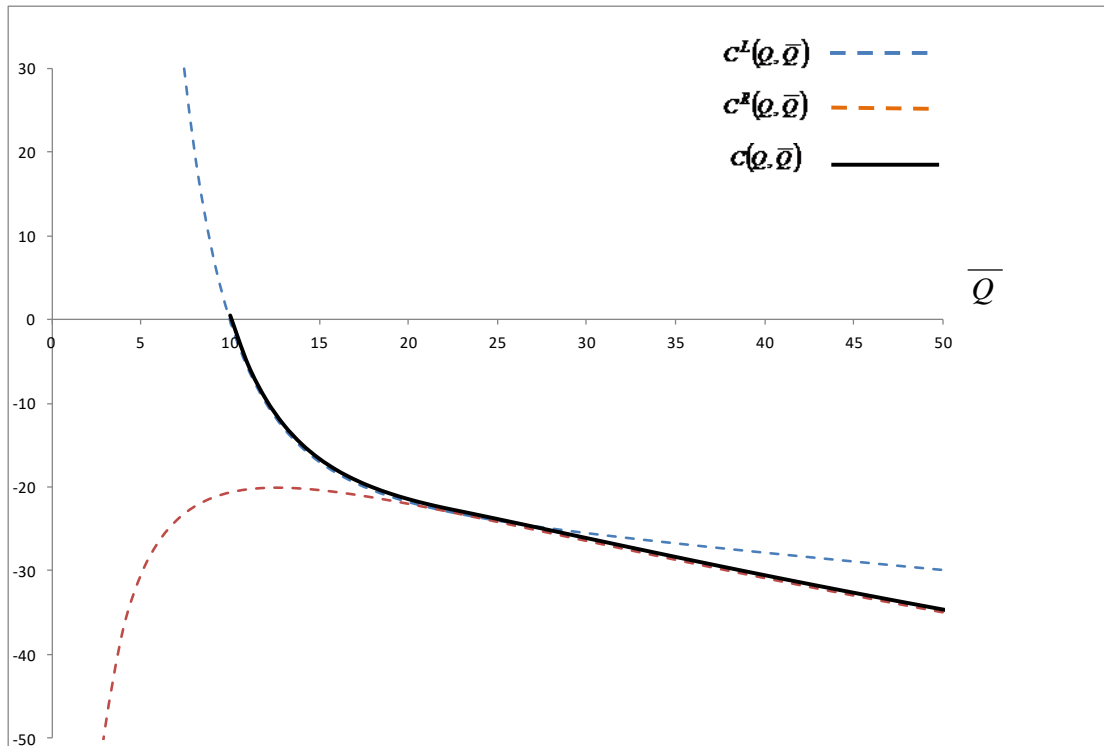


Figure 2. When $Q = 10$ $C(Q, \bar{Q})$ is decreasing for all $\bar{Q} > Q$ and maximized at Q .

With a smaller level of Q , namely $Q = 4$, the function $C^L(Q, \bar{Q})$ is not decreasing for each level of \bar{Q} , as in the case when $Q = 10$. Instead it decreases only in relatively low

levels of \bar{Q} , but beyond a certain point it starts to rise towards a local maximum point, and then falls again in \bar{Q} .

In this case $C^L(Q, \bar{Q})$ represents $C(Q, \bar{Q})$ within the range $4 < \bar{Q} < 9$. Its minimum point is within that range, but its maximum point is not.

This level of Q is sufficiently small for $C^L(Q, \bar{Q})$ and $C^R(Q, \bar{Q})$ to meet when they are both rising in \bar{Q} and therefore the resulting $C(Q, \bar{Q})$ has a local maximum at the point where $C^R(Q, \bar{Q})$ peaks. Yet, with $Q = 4$, this peak is with a negative value, so the global maximum of $C(Q, \bar{Q})$ is once again at $\bar{Q} = Q$.

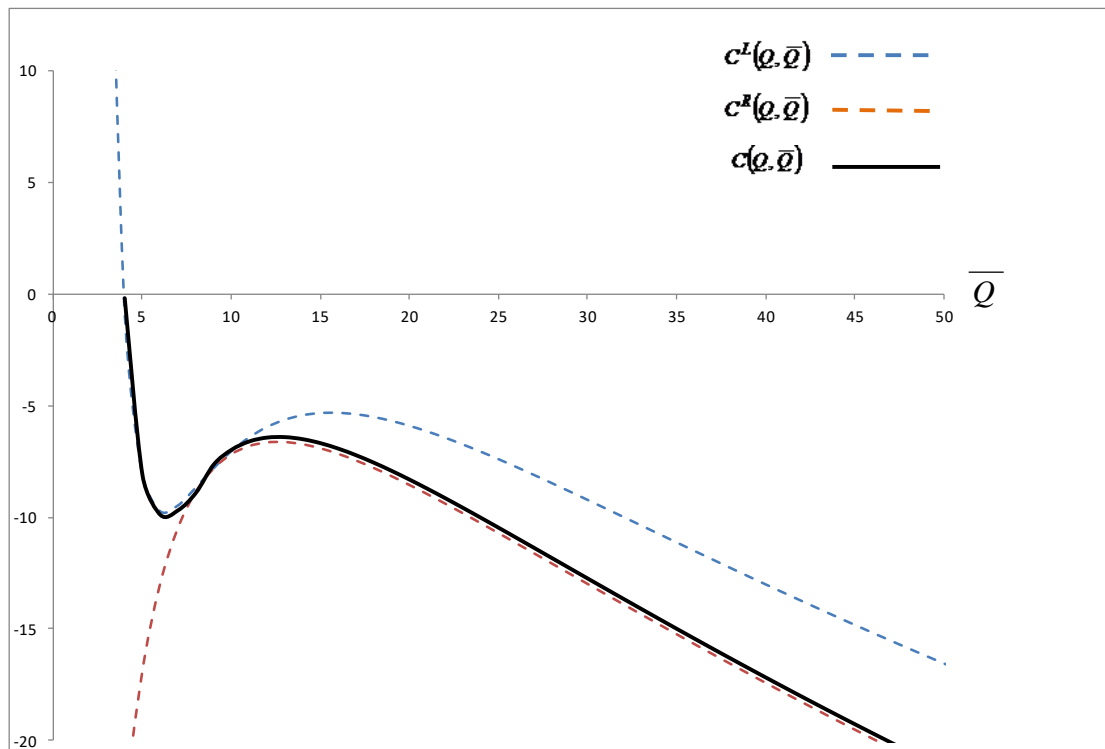


Figure 3. When $Q = 4$, $C(Q, \bar{Q})$ has a local maximum at $\bar{Q} = \bar{Q}^{opt}$, but its global maximum is at $\bar{Q} = Q$.

We now turn to look at a case where Q is even lower than in the previous one, namely $Q = 3$. As in the previous case, here too $C^L(Q, \bar{Q})$ as a function of \bar{Q} is initially decreasing, then rising towards a local maximum, and then falls again, and consequently

this is also the shape of $C(Q, \bar{Q})$. However, due to the smaller Q in this case, the local peak of $C(Q, \bar{Q})$ is now positive, implying that it is also a global maximum and therefore that it is optimal to set the cap at \bar{Q}^{opt} , which given the above parameter values, and based on (23.2) and (25), equals $\bar{Q}^{opt} = 12.64$. The market quantity is going therefore to gradually grow from the initial quantity of $Q = 3$ until it reaches $\tilde{Q} = 5.62$. Then, when the entry threshold is hit again – a run takes Q immediately to 12.64.

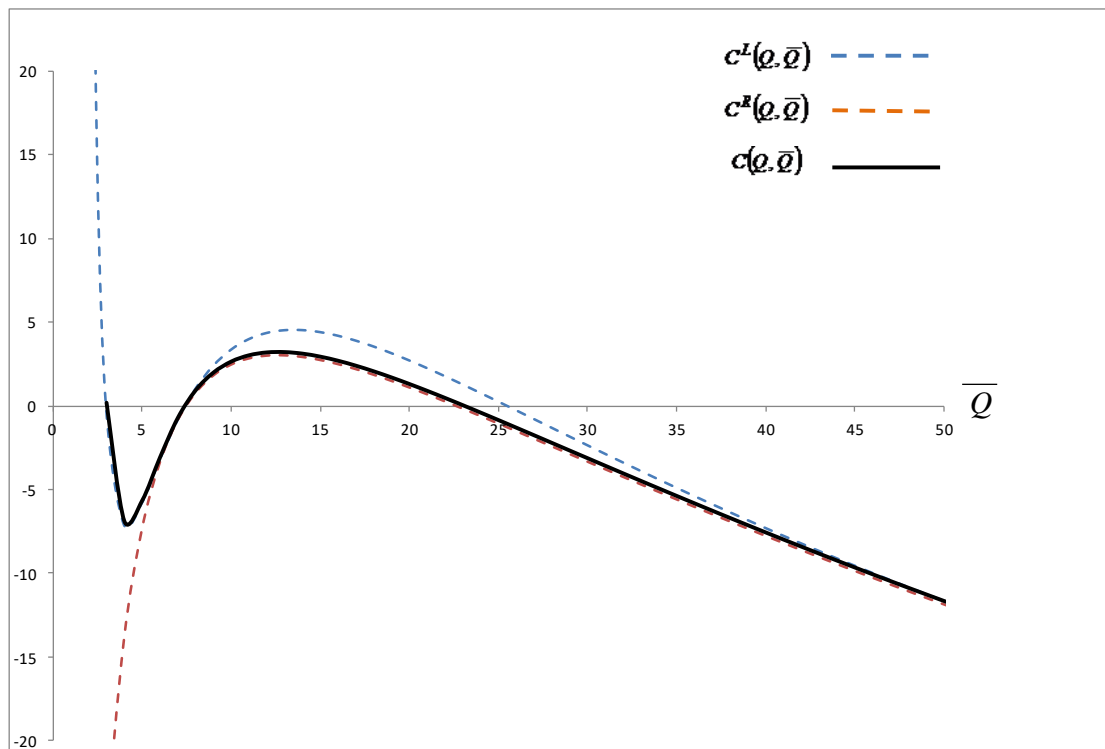


Figure 4. When $Q = 3$. $C(Q, \bar{Q})$ has a global maximum at $\bar{Q} = \bar{Q}^{opt}$.

3.3 Comparative statics

By (25), \bar{Q}^{opt} is decreasing in e/M , which shows the relative speed by which the external cost grows compared to the private costs. This is very intuitive and shows that the faster the welfare loss grows the earlier the cap is needed.

It also follows from (36) that \bar{Q}^{opt} is falling in β implying that it is rising in σ . Thus, the higher the uncertainty the more investment allowed, which captures the diluting effect that the "uncertainty wedge" has on the welfare loss induced by the externality.

A similar conclusion applies not only for the size of the optimal cap but also for the likelihood that it will be imposed at all, rather than stopping entry immediately at the current Q . In particular it is shown by *Proposition 3* which shows that if that $\beta < \beta_1$ then a cap is set even when the current Q exceeds $\frac{1}{\beta} \cdot \bar{Q}^{opt}$, while if that $\beta > \beta_1$ then a cap is set only if Q is below $\frac{1}{\beta} \cdot \bar{Q}^{opt}$. These shows the effect of uncertainty because β is a decreasing function of the uncertainty parameter σ^2 as follows from (6.1).

4. Optimal cap under rationing

In this section, we assume that entry licenses are distributed when the cap is announced. Each license is for an infinitesimally small quantity ΔQ , and their number is equal to difference between the current market quantity, Q , and the cap.¹⁰ We abstract from how the licenses were distributed, whether by auction, lottery, or any other way. We merely assume that their distribution has no other implications except for providing each license owner with a right to enter the market at any time.¹¹

4.1 Optimal investment policy

The analysis of the firm's optimal policy under rationing is technically similar to the analysis in Section 3 for the case of competitive entry. The relevant difference between

¹⁰ Bartolini (1995) introduces this case and analyzes firms' entry policy and social welfare in equilibrium.

¹¹ See Bartolini (1995) for a discussion of how alternative mechanisms impact on surplus extraction.

the two cases is that under licensing the option to enter is an asset having a positive value that the firm gives up by entering the market. Thus, alongside the function $V(Q, X)$ which represents the value of an active firm, we define the function $F(Q, X)$ which stands for the value of the option to enter the market. A standard no-arbitrage analysis yields:¹²

$$(27) \quad F(Q, X) = H(Q) \cdot X^\beta,$$

where $H(Q)$ is to be found by imposing the following boundary conditions:

$$(28) \quad V[Q, X^*(Q)] = F[Q, X^*(Q)],$$

$$(29) \quad V_X[Q, X^*(Q)] = F_X[Q, X^*(Q)].$$

Condition (28) is the so called *Value Matching condition* and states that at the entry time the value of an idle firm is equal to that of an active one. Condition (29) is a *Smooth Pasting condition* requiring that at the entry time the two value functions have the same slope with regard to X .¹³

Substituting (6) and (27) into the system (28-29) and solving for the optimal entry yields:

$$(30) \quad X^*(Q) = \bar{\beta} \cdot (r - \mu) \cdot \frac{M}{r} \cdot Q,$$

¹² Note that also this function can be derived by following the same procedure illustrated in detail in Appendix A (for the value of an active firm).

¹³ Note that, as shown by Dixit (1993), the *Value Matching condition* holds for any threshold and merely reflects a no arbitrage assumption, while the *Smooth Pasting condition* is an optimality condition which holds only at the optimal threshold.

that is, the threshold function illustrating the entry process up to \tilde{Q} in the case of competitive entry. The main difference is that, under rationing, (30) drives the entry process until the cap is reached. This means that, in contrast with the case of competitive entry, no run arises in the equilibrium under rationing. This is because firms holding a license do not fear that they will lose their entry option.

Last, the solution of the system (28-29) yields also an expression for $H(Q) - Y(Q)$ from which $F(Q, X)$ can be fully characterized, once $V(Q, X)$ is found using the boundary condition (12).

4.2 Welfare and the optimal cap

Following the same procedure as that conducted in Section 3, the expected discounted social welfare, given the current levels of X and Q , is:

$$(31) \quad W(Q, X, \bar{Q}) = C(Q, \bar{Q}) \cdot X^\beta + \frac{X \cdot \ln(Q)}{r - \mu} - \frac{M \cdot Q + e \cdot \frac{Q^2}{2}}{r},$$

where $C(Q, \bar{Q})$ is to be determined by boundary conditions. The first such condition is the following *Value Matching Condition* at times of hitting the investment threshold:

$$(32) \quad w_Q[Q, X^*(Q), \bar{Q}] = 0.$$

Applying (31) in (32), partially differentiating with respect to Q , applying (30), and rearranging terms, yields:

$$(33) \quad C_{\bar{Q}}(Q, \bar{Q}) = K \cdot \left[\frac{e}{Q^{\beta-1}} - \frac{M}{(\beta-1) \cdot Q^{\beta}} \right].$$

Integrating (33) yields:

$$(34) \quad C(Q, \bar{Q}) = K \cdot \left[\frac{M}{(\beta-1)^2 \cdot Q^{\beta-1}} - \frac{e}{(\beta-2) \cdot Q^{\beta-2}} \right] + G.$$

The integration constant is found using the following boundary condition:

$$(35) \quad C(\bar{Q}, \bar{Q}) = 0,$$

which states that when Q is at \bar{Q} no more changes in Q are going to take place, and therefore $C(\bar{Q}, \bar{Q}) \cdot X^{\beta}$, which shows the value of such changes within the welfare function (31), should equal zero.

Applying (34) in (35), extracting the integration coefficient G , applying it in (34) and simplifying, yields:

$$(36) \quad C(Q, \bar{Q}) = K \cdot \left[\frac{M}{(\beta-1)^2} \cdot \left(\frac{1}{Q^{\beta-1}} - \frac{1}{\bar{Q}^{\beta-1}} \right) - \frac{e}{\beta-2} \cdot \left(\frac{1}{Q^{\beta-2}} - \frac{1}{\bar{Q}^{\beta-2}} \right) \right].$$

Note from (36) that if the cap is set at its lowest possible level, i.e., at the current level of Q , then $C(Q, \bar{Q}) = 0$. This reflects the fact that the first term in the welfare function, $C(Q, \bar{Q}) \cdot \bar{X}^{\beta}$ shows the welfare value of future changes in Q , and no such changes will take place if the cap is imposed at the current level of Q .

Differentiating (36) with respect to \bar{Q} yields that $C(Q, \bar{Q})$, and therefore welfare, is an inverse u-shape function of \bar{Q} maximized at:

$$(37) \quad \bar{Q} = \frac{M}{(\beta-1) \cdot e} \equiv \bar{Q}^*$$

Thus, if the current level of Q is above \bar{Q}^* then it is optimal to immediately stop any further investment by setting the cap at the current level of Q . Otherwise, if the current level of Q is still below \bar{Q}^* then additions to quantity, based on firms' optimal behavior, is still beneficial for welfare and a cap should be set at \bar{Q}^* .

Note that in the latter case the cap is a decreasing function of the parameter ratio e/M which captures the speed by which the external cost grows with respect to the private cost. Thus, the faster the external cost grows with respect to the private cost, the lower, *ceteris paribus*, the number of licenses that will be distributed. Note also that, as β is a decreasing function of σ^2 , the larger the volatility of future profits, the higher the level of the cap.

5. Conclusion

In this study, we have presented a model of endogenous market entry under uncertainty, with production externalities regulated introducing a cap on aggregate entry. Most of the previous literature on this topic has assumed that the cap is exogenous. The only exception is Di Corato and Maoz (2019) who explicitly model the externalities and determine endogenously the optimal cap.

A key assumption in Di Corato and Maoz (2019) is that that the external cost and private cost grow at sufficiently close rates to one another as the industry output grows. Here, differently, we explore the case where the external cost rises faster than the private costs as the industry output grows.

While a large mass of empirical evidence shows that external costs, and in particular the environmental ones, tend to be convex in the aggregate industry output, there is very little evidence about whether they grow at a speed close to that of the private cost or whether they grow faster. Thus, it is of interest to study both cases.

The main finding in Di Corato and Maoz (2019) was that due an "*uncertainty premium*" that makes the price which trigger entry higher than marginal private cost – it is possible that optimal entry should take place at prices that exceed the full cost to society, i.e., the sum of the private cost and the external cost. The assumption that private and external costs grow at similar speed made this result independent of industry output. Thus, they find that if market uncertainty is high enough than no cap should be imposed, and otherwise that the cap should be imposed immediately at the current level of quantity, preventing thus any further entry.

Here, due to the assumption that the external cost grows faster than the private costs, the possibility of an optimal finite cap does emerge. We also find that the size of the cap is increasing in the level of uncertainty, where this a result springs from the counterbalancing effect that the uncertainty premium has with regard to how the welfare loss caused by the externalities.

A similar conclusion applies not only for the size of the optimal cap but also for the likelihood that it will be imposed at all, rather than stopping entry at the current

quantity. The higher the uncertainty – the wider the range of the model parameter values for which a finite cap is preferred over stopping entry at the current quantity.

We also find the rather intuitive result that the size of the cap and the likelihood of its optimality are decreasing in the speed by which the external cost rise faster than the private costs as the quantity grows.

Alongside caps, taxes are often considered and used as means for improving welfare in the presence of externalities. In this study we have ignored altogether the issue of whether taxes can be better than caps tool for that end, mostly for the sake of tractability of analysis of the model. Conducting an analysis that contains both taxes and a cap in a model of investment under uncertainty with production uncertainties is therefore a task for future research in this field.

Appendix A - The value of an active firm

In this Appendix we show that (6) represents the general form of the function $V(Q, X)$. For that, we use the standard no-arbitrage analysis of the literature on investment under uncertainty (see e.g. Dixit 1989). We start this analysis with the no-arbitrage condition:

$$(A.1) \quad r \cdot V(Q, X) \cdot dt = \frac{X}{Q} - M + E[dV(Q, X)],$$

states that the instantaneous profit, $\frac{X}{Q} - M$, along with the expected instantaneous capital gain, $E[dV(Q, X)]$, from a change in X , must equal the instantaneous normal return, $r \cdot V(Q, X) \cdot dt$.

By Ito's lemma:

$$(A.2) \quad \frac{E[dV(Q, X)]}{dt} = \frac{1}{2} \cdot \sigma^2 \cdot X^2 \cdot V_{XX}(Q, X) + \mu \cdot X \cdot V_X(Q, X).$$

Substituting (A.2) in (A.1) yields:

$$(A.3) \quad \frac{1}{2} \cdot \sigma^2 \cdot X^2 \cdot V_{XX}(Q, X) + \mu \cdot X \cdot V_X(Q, X) - r \cdot V(Q, X) = -\left(\frac{X}{Q} - M\right).$$

Trying a solution of the type X^b for the homogenous part of (A.3) and a linear form as a particular solution to the entire equation, yields:

$$(A.4) \quad V(Q, X) = Z(Q) \cdot X^\alpha + Y(Q) \cdot X^\beta + \frac{X}{Q \cdot (r - \mu)} - \frac{M}{r},$$

where $\alpha < 0$ and $\beta > 1$ solve the quadratic:

$$(A.5) \quad \frac{1}{2} \cdot \sigma^2 \cdot x \cdot (x - 1) + \mu \cdot x - r = 0.$$

Note that the term $\frac{X}{Q(r-\mu)} - \frac{M}{r}$ represents the expected value of the flow of profits if Q remains forever at its current level. The two other elements of the RHS of (A.4) represent therefore how expected future changes in Q are affect the value of the firm.

By the properties of the Geometric Brownian Motion, when X goes to 0 the probability of ever hitting $X^*(Q)$, and, consequently, Q increasing, tends to 0. This implies:

$$(A.6) \quad \lim_{X \rightarrow 0^+} [Z(Q) \cdot X^\alpha + Y(Q) \cdot X^\beta] = 0.$$

Note that as $\alpha < 0$, (A.6) holds only if $Z(Q) = 0$. Finally, substituting $Z(Q) = 0$ into (A.4) gives (6).

Appendix B – Proof of Proposition 1

Applying (25) and (23.1) in (23) and simplifying, yields:

$$(B.1) \quad C^R(Q, \bar{Q}^{opt}) = \frac{K}{Q^{\beta-1}} \cdot \left[\frac{M}{(\beta-1)^2} - \frac{e \cdot Q}{\beta-2} \right] + \frac{K \cdot \bar{\beta}^\beta \cdot M \cdot g(\beta)}{(\beta-1) \cdot (\beta-2) \cdot \bar{Q}^{\beta-1}}.$$

The relevant range of Q in this proposition is $0 < Q < Q_1$, where Q_1 is given by (25.1).

From (B.1) it follows that at the right end of this range:

$$(B.2) \quad \lim_{Q \rightarrow 0} C^R(Q, \bar{Q}^{opt}) = \infty.$$

Evaluating (B.1) at $Q = Q_1$ and simplifying, shows that at the left end of the relevant range:

$$(B.3) \quad C^R(Q_1, \bar{Q}^{opt}) = \frac{\bar{\beta}^{\beta-1} \cdot K \cdot M}{(\beta-1)^2 \cdot (\beta-2) \cdot \bar{Q}^{\beta-1}} \cdot h(\beta),$$

where

$$(B.4) \quad h(\beta) \equiv \beta - 2 - 2 \cdot (\beta - 1)^3 \cdot g(\beta) + \beta \cdot g(\beta).$$

$h(\beta)$ is a single-variable function of β . Numerically calculating its values shows that it is of the following shape:

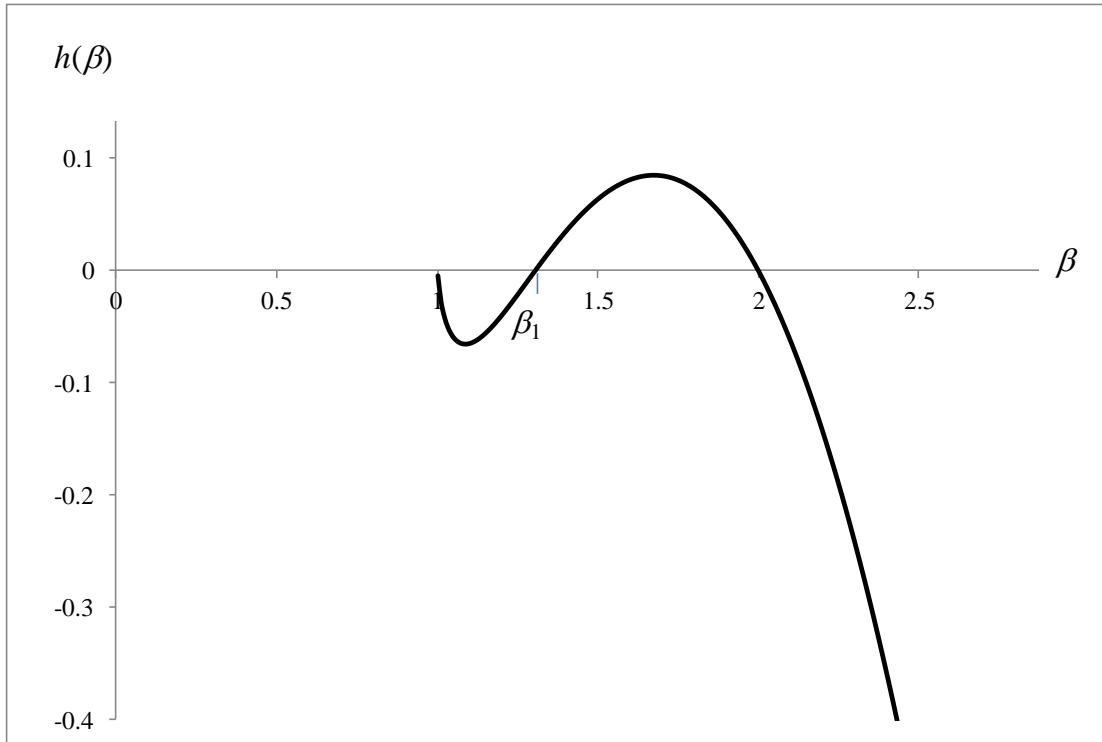


Figure B.1: The function $h(\beta)$.

From (B.4) it immediately follows that that $h(\beta) = 0$ when $\beta = 1$ and when $\beta = 2$. A numerical analysis of (B.4) also yields that the other root of this equation, denoted β_1 , approximately equals 1.3.

Thus, by (B.3) and the properties of $h(\beta)$, at the left end of the relevant range $C^R(Q, \bar{Q}^{opt}) > 0$ if and only if $\beta < \beta_1$.

Within the relevant range $C^R(Q, \bar{Q}^{opt})$ decreases in Q , as follows from:

$$\begin{aligned}
 \text{(B.3)} \quad C^R_Q(Q, \bar{Q}) &= \frac{K}{Q^\beta} \cdot \left(e \cdot Q - \frac{M}{\beta-1} \right) < \frac{K}{Q^\beta} \cdot \left(e \cdot Q_1 - \frac{M}{\beta-1} \right) \\
 &= \frac{K \cdot M}{(\beta-1) \cdot Q^\beta} \cdot \left[2 \cdot (\beta-1)^2 \cdot g(\beta) - 1 \right] < 0,
 \end{aligned}$$

The first equality follows from straightforward differentiation of (B.1) with respect to Q and simplifying. The first inequality follows from $Q < Q_1$ which is one of the conditions of the proposition, the second equality follows from applying (25.1) and simplifying. The second inequality follows from numerically analyzing the expression in the squared brackets which is a single-variable function of β which is negative throughout the range $1 < \beta < \beta_1 \approx 1.3$ assumed in the conditions of this proposition.

Thus, if $\beta < \beta_1 \approx 1.3$ then $C^R(Q, \bar{Q}^{opt})$ is positive at both ends of the relevant range, and monotonic between them implying that it is positive throughout that range. This proves part (a) of the proposition.

Similarly, if $\beta \geq \beta_1 \approx 1.3$ then $C^R(Q, \bar{Q}^{opt})$ is positive at the left end of the relevant range, negative at its right end, and monotonically decreasing throughout this range. This implies the existence of a unique root for the equation $C^R(Q, \bar{Q}^{opt}) = 0$ within that range. This proves part (b) of the Proposition.

Appendix C – Proof of Proposition 2

By (18),

$$(C.1) \quad C^L_{\bar{Q}}(Q, \bar{Q}) = \frac{\bar{\beta}^\beta \cdot K \cdot f(\bar{Q})}{\bar{Q}^\beta},$$

where:

$$(C.2) \quad f(\bar{Q}) \equiv -(\beta - 1) \cdot M \cdot \bar{Q} \cdot \ln\left(\frac{\bar{Q}}{Q}\right) + \beta \cdot M \cdot (\bar{Q} - Q) + \frac{e}{2} \cdot [(\beta - 2) \cdot \bar{Q}^2 - \beta \cdot Q^2]$$

From (C.1) it follows that $f(\bar{Q})$ determines the sign of $C^{L\bar{Q}}(Q, \bar{Q})$. Therefore, much of the analysis shall focus on properties of $f(\bar{Q})$. From (C.2) it follows that:

$$(C.3) \quad \lim_{\bar{Q} \rightarrow Q} f(\bar{Q}) = -e \cdot Q^2$$

$$(C.4) \quad f(\bar{\beta} \cdot Q) = \bar{\beta} \cdot Q \cdot \left[g(\beta) \cdot M - \frac{e}{2 \cdot (\beta - 1)} \cdot Q \right],$$

where $g(\beta)$ is defined by (23.2) and satisfies $0 < g(\beta) < 1$. From $g(\beta) > 0$, $\beta > 1$ and (C.4) it follows that $f(\bar{\beta} \cdot Q) > 0$ if and only if:

$$(C.5) \quad Q < \frac{2 \cdot (\beta - 1) \cdot g(\beta) \cdot M}{e} \equiv Q_1.$$

Straightforward differentiation of (C.2) yields:

$$(C.6) \quad f'(\bar{Q}) = -(\beta - 1) \cdot M \cdot \ln\left(\frac{\bar{Q}}{Q}\right) + M + e \cdot (\beta - 2) \cdot \bar{Q}$$

$$(C.7) \quad \lim_{\bar{Q} \rightarrow Q} f'(\bar{Q}) = M + e \cdot (\beta - 2) \cdot Q$$

$$(C.8) \quad f'(\bar{\beta} \cdot Q) = M \cdot g(\beta) + e \cdot (\beta - 2) \cdot \bar{\beta} \cdot Q$$

Straightforward differentiation of (C.6) yields:

$$(C.9) \quad f''(\bar{Q}) = -(\beta - 1) \cdot M \cdot \frac{1}{\bar{Q}} + e \cdot (\beta - 2)$$

Lemma C.1: Throughout the range $Q < \bar{Q} < \bar{\beta} \cdot Q$,

- (a) If $\beta > 2$ then $f'(\bar{Q}) > 0$
- (b) If $\beta < 2$ then $f''(\bar{Q}) < 0$

Proof: if $\beta > 2$ then (C.6), taken together with $\bar{Q} < \bar{\beta} \cdot Q$, leads to:

$$(C.10) \quad f'(\bar{Q}) > -(\beta - 1) \cdot M \cdot \ln\left(\frac{\bar{\beta} \cdot Q}{\bar{Q}}\right) + M + e \cdot (\beta - 2) \cdot \bar{Q}$$

$$= M \cdot g(\beta) + e \cdot (\beta - 2) \cdot \bar{Q} > 0,$$

where the equality follows from (23.2) and the second inequality follows from $g(\beta) > 0$ and from $\beta > 2$. This proves part (a). Part (b) follows immediately from (C.9).

Lemma C.2: When $\beta < 2$:

- (a) $\lim_{\bar{Q} \rightarrow Q} f'(\bar{Q}) > 0$ if and only if $Q < \frac{M}{e \cdot (2 - \beta)} \equiv Q_2$,
- (b) $f'(\bar{\beta} \cdot Q) > 0$, if and only if: $Q < \frac{M \cdot g(\beta)}{e \cdot (2 - \beta) \cdot \bar{\beta}} \equiv Q_3$.
- (c) $Q_1 < Q_2$
- (d) $Q_3 < Q_2$
- (e) $Q_1 > Q_3$ if and only if $\beta < \beta_2 = 1 + \sqrt{0.5} \cong 1.7$

Proof: follows directly from (C.7) and (C.8) □

C.1 If $Q < Q_1$

In this case $f(\bar{\beta} \cdot Q) > 0$, as follows from (C.5). Thus, $f(\bar{Q})$ is a function which is negative at its left end, due to (C.3), and positive at its right end. *Lemma C.1* assures that it crosses the $f(\bar{Q})=0$ line just once.

Thus, within the relevant range $C^L(Q, \bar{Q})$ is an inverse u-shape function of \bar{Q} . The super contact property, presented in sub-section 3.2.3, ensures that in this case $C^L(Q, \bar{Q})$ and $C^R(Q, \bar{Q})$ meet at a point where their slopes are positive – implying that $C(Q, \bar{Q})$ has a local maximum within the range where it is represented by $C^R(Q, \bar{Q})$. By *Proposition 1* this local maximum point is at \bar{Q}^{opt} , captured by (25).

Due to the inverse u-shape, $C(Q, \bar{Q})$ has another local maximum point at its left end, i.e., when \bar{Q} equals the current Q . By (18), at that local maximum $C(Q, \bar{Q})$ equals 0. Therefore, the global maximum of $C(Q, \bar{Q})$ is either at $\bar{Q} = Q$ or at $\bar{Q} = \bar{Q}^{opt}$, depending on whether the local maximum at \bar{Q}^{opt} is positive or not. From *Proposition 1* it follows therefore that:

- if $\beta \leq \beta_1$ then the optimal cap equals \bar{Q}^{opt} for any $Q \leq Q_1$
- if $\beta > \beta_1$ then there exists a certain $Q^* < Q_1$ such that:

(i) If $Q < Q^*$ then the optimal cap equals \bar{Q}^{opt}

(ii) if $Q \geq Q^*$ then the optimal cap equals Q ,

where, Q^* is the unique root that $C^R(Q, \bar{Q}^{opt})=0$ has within the range $Q < Q_1$. Figures 2 and 3 exemplify the two possibilities that exist in the case of $\beta > \beta_1$.

This proves part (a.i) and part (b.i) of *Proposition 3*. It also proves part (a.2) for the range $Q^* < Q \leq Q_1$.

C.2 If $Q > Q_1$

In this case $f(\bar{Q})$ is negative in both its ends, due to (C.3) and (C.5). Therefore, by

Lemma C.1, the following four cases are possible:

- (1) $f(\bar{Q})$ is monotonically decreasing in \bar{Q} throughout the relevant range
- (2) $f(\bar{Q})$ is monotonically increasing in \bar{Q} throughout the relevant range
- (3) $f(\bar{Q})$ is a an inverse-shape function of \bar{Q} throughout the relevant range, with a negative peak
- (4) $f(\bar{Q})$ is a an inverse-shape function of \bar{Q} throughout the relevant range, with a positive peak

In case (1), (2) and (3) the function $f(\bar{Q})$ is negative throughout the relevant range and therefore $C^L(Q, \bar{Q})$ is strictly decreasing in \bar{Q} throughout that range.

Due to the super contact property presented in sub-section 3.2.3, in these three cases where $C^L(Q, \bar{Q})$ is strictly decreasing it meets $C^R(Q, \bar{Q})$ at a point where the slopes of $C^R(Q, \bar{Q})$ is negative too. This implies that at the meeting point $C^R(Q, \bar{Q})$ is beyond its peak and already falling in \bar{Q} within the range in which it represents $C(Q, \bar{Q})$. Thus, in these cases $C(Q, \bar{Q})$ is maximized at its left end where \bar{Q} equals Q .

Case (4) is the only one where $f(\bar{Q})$ has a positive part within the relevant range. As in the other three cases, here too $f(\bar{Q})$ is negative at both ends of that range. Yet, in this case it also concavely reaches a positive maximum within this range. This implies that there are two points in which $f(\bar{Q}) = 0$, where the one with the smaller value of \bar{Q} is a local minimum of $C^L(Q, \bar{Q})$ and the other is a local maximum, which we denote by \bar{Q}^L . Figure X schematically shows the resulting shape of $C^L(Q, \bar{Q})$ in this case:

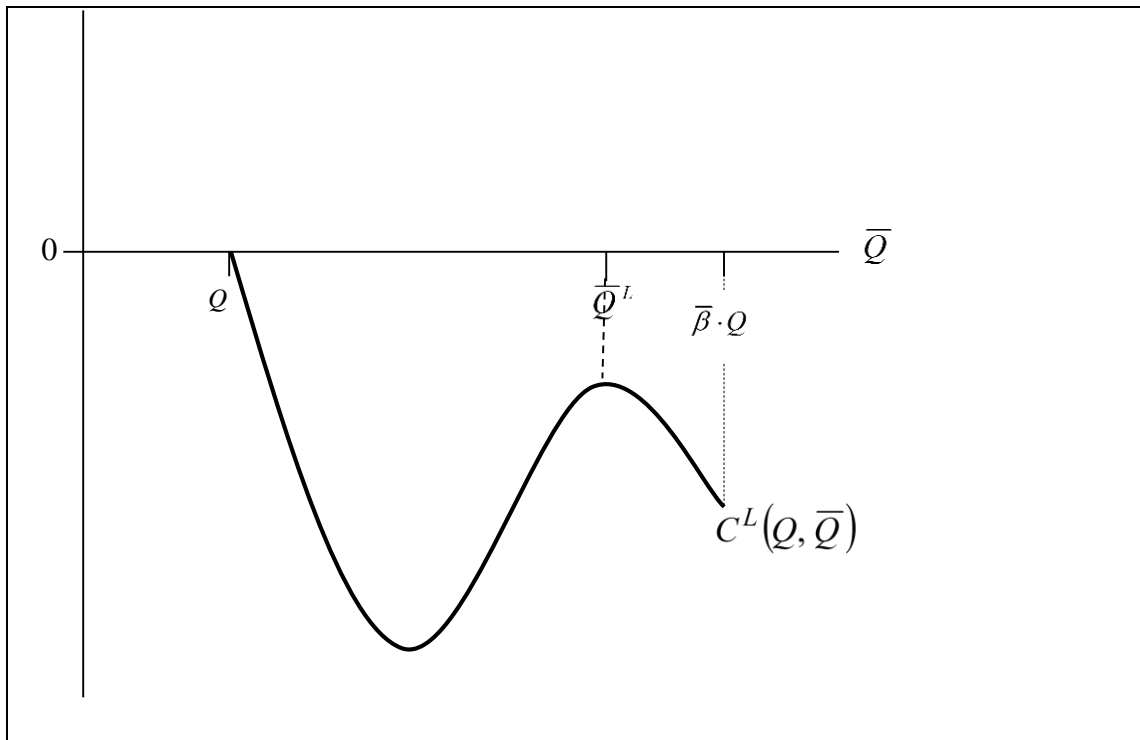


Figure C.1: $C^L(Q, \bar{Q})$ under Case 4.

The local maximum at \bar{Q}^L can be either positive or negative. Due to the super contact property, at this case where $Q > Q_1$ the functions $C^L(Q, \bar{Q})$ and $C^R(Q, \bar{Q})$ meet at a point where both their slopes are negative, implying that $C^R(Q, \bar{Q})$ is already beyond its maximum and in the range where it represent $C(Q, \bar{Q})$ it is decreasing in \bar{Q} . Thus,

in this case the global maximum of $C(Q, \bar{Q})$ can be either at $\bar{Q} = Q$ or at $\bar{Q} = \bar{Q}^L$ depending on whether the local maximum at \bar{Q}^L is positive or not.

A necessary condition for Case 4 to hold is that Q exceeds Q_1 and Q_3 . It ensures, by (C.5) and *Lemma C.2* that $f(\bar{Q}) < 0$ and $f'(\bar{Q}) < 0$ at the right end of the relevant range.

Another mandatory condition for Case 4 to prevail is that $Q_1 > Q_3$ which implies, by *Lemma C.2*, that $\beta < \beta_2 \approx 1.7$. To see that, note that under the opposite case, if $Q = Q_3$ then by *Lemma C.2*, the function $f(\bar{Q})$ reaches its peak at the left end of the relevant range, and this peak is negative because $Q > Q_1$. Then, for any other $Q > Q_3$, the peak is going to be negative too, as follows from the following *Lemma C.3*.

Lemma C.3: If $f(\bar{Q})$ has a local maximum point within the range $Q < \bar{Q} < \bar{\beta} \cdot Q$, then the values of \bar{Q} and $f(\bar{Q})$ at this peak are both decreasing in Q .

Proof: The maximum condition, $f'(\bar{Q}) = 0$, yields, via implicit differentiation, that at this maximum point:

$$(C.11) \quad \frac{d\bar{Q}}{dQ} = \frac{(\beta-1) \cdot M \cdot \frac{1}{Q}}{(\beta-1) \cdot M \cdot \frac{1}{Q} - e \cdot (\beta-2)} > 0$$

Where the inequality follows from $1 < \beta < 2$. From (C.2) it follows that at this point:

$$(C.12) \quad \frac{df(\bar{Q})}{dQ} = -e \cdot (\beta-2) \cdot \bar{Q} \cdot \frac{d\bar{Q}}{dQ} + (\beta-1) \cdot M \cdot \frac{d\bar{Q}}{dQ} - \beta \cdot M - e \cdot \beta \cdot Q^2$$

$$= (\beta - 1) \cdot M \cdot \frac{\bar{Q}}{Q} - \beta \cdot M - e \cdot \beta \cdot Q^2 < -e \cdot \beta \cdot Q^2 < 0,$$

where the second equality follows from applying (C.11) and simplifying, and the first inequality springs from applying $\bar{Q} < \bar{\beta} \cdot Q$ and simplifying.

Thus, if β is not smaller than β_2 then $f(\bar{Q})$ is negative even at its peak implying that $C^L(Q, \bar{Q})$ is decreasing throughout the relevant range.

While $\beta < \beta_2$ is mandatory for Case 4 to prevail and in particular for $C^L(Q, \bar{Q})$ to have an internal local maximum within the relevant range – it turns out that β should be even smaller for this local maximum to be positive and make \bar{Q}^L the optimal cap. Specifically, the condition for that is that β should be smaller than $\beta_1 \approx 1.3$. To see that, note that if $Q = Q_1$ then $f(\bar{Q}) = 0$ at the left end of the relevant range, as follows from (C.5). Thus, in this case $C^L(Q, \bar{Q})$ reaches its peak exactly at the meeting point with $C^R(Q, \bar{Q})$. By the super contact property this is also the maximum point of $C^L(Q, \bar{Q})$, and as shown in *Proposition 1*, this maximum is positive if and only if $\beta < \beta_1$. If β exceeds β_1 then not only that $C^L(Q, \bar{Q})$ has a negative peak when Q equals Q_1 but it will also have a negative peak for any $Q > Q_1$ as the following *Lemma C.4* shows.

Lemma C.4: If $C^L(Q, \bar{Q})$ has an internal peak within the range $Q < \bar{Q} < \bar{\beta} \cdot Q$, then the values of \bar{Q} and $C^L(Q, \bar{Q})$ at this peak are both decreasing in Q .

Proof: By the first order condition for a maximum, $f(\bar{Q}) = 0$:

$$(C.13) \quad \frac{d\bar{Q}}{dQ} = -\frac{(\beta-1) \cdot M \cdot \bar{Q} \cdot \frac{1}{\bar{Q}} - \beta \cdot M - e \cdot \beta \cdot Q}{f'(\bar{Q})} < \frac{e \cdot \beta \cdot Q}{f'(\bar{Q})} < 0,$$

where the equality follows from implicit differentiation of this condition, and the first inequality follows from $\bar{Q} < \bar{\beta} \cdot Q$ taken together with the second order condition for a maximum which is $f'(\bar{Q}) < 0$. Thus, \bar{Q}^L which is the value of \bar{Q} that brings $C^L(Q, \bar{Q})$ to the maximum in Case 4 is a decreasing function of Q .

Total differentiation of $C^L(Q, \bar{Q})$, as captured by (18), leads to:

$$(C.14) \quad \begin{aligned} \frac{dC^L(Q, \bar{Q})}{dQ} &= C^L_Q(Q, \bar{Q}) + C^L_{\bar{Q}}(Q, \bar{Q}) \cdot \frac{d\bar{Q}}{dQ} = C^L_Q(Q, \bar{Q}) \\ &= \bar{\beta}^\beta \cdot K \cdot \frac{-M \cdot (\bar{Q} - Q) + e \cdot Q^2}{Q \cdot \bar{Q}^\beta}. \end{aligned}$$

The second equality follows from $C^L_{\bar{Q}}(Q, \bar{Q}) = 0$ at the peak. The sign of this derivative is based on the sign of the numerator and the following analysis shows that it is negative. In particular, extracting an expression for $-(\beta-1) \cdot M \cdot \ln\left(\frac{\bar{Q}}{Q}\right)$ from the F.O.C $f(\bar{Q}) = 0$, applying it in the S.O.C $f'(\bar{Q}) = 0$, and simplifying, yields that at the peak:

$$(C.15) \quad -\beta \cdot M \cdot (\bar{Q} - Q) + \frac{e}{2} \cdot \beta \cdot Q^2 + M \cdot \bar{Q} + \frac{e}{2} \cdot (\beta - 2) \cdot \bar{Q}^2 < 0$$

Adding $e \cdot \beta \cdot Q^2$ on both sides and simplifying further yields:

$$(C.16) \quad \beta \cdot \left[-M \cdot (\bar{Q} - Q) + e \cdot Q^2 \right] < -M \cdot \bar{Q} - \frac{e}{2} \cdot [(\beta - 2) \cdot \bar{Q}^2 - \beta \cdot Q^2]$$

Extracting an expression for $-\frac{\epsilon}{2} \cdot [(\beta - 2) \cdot \bar{Q}^2 - \beta \cdot Q^2]$ from the F.O.C $f(\bar{Q}) = 0$, and applying it in the RHS of (C.15) and simplifying yields:

$$\begin{aligned}
\text{(C.17)} \quad & \beta \cdot [-M \cdot (\bar{Q} - Q) + e \cdot Q^2] \\
& < -(\beta - 1) \cdot M \cdot \bar{Q} \cdot \ln\left(\frac{\bar{Q}}{Q}\right) + (\beta - 1) \cdot M \cdot \bar{Q} - \beta \cdot M \cdot Q \\
& < -(\beta - 1) \cdot M \cdot \bar{Q} \cdot \ln\left(\frac{\bar{Q}}{Q}\right) + (\beta - 1) \cdot M \cdot \bar{Q}^{opt} - \beta \cdot M \cdot Q_1 \\
& = -(\beta - 1) \cdot M \cdot \bar{Q} \cdot \ln\left(\frac{\bar{Q}}{Q}\right) < 0,
\end{aligned}$$

where the second inequality follows from \bar{Q}^{opt} being the value of \bar{Q} that maximizes $C^L(Q, \bar{Q})$ too if $Q = Q_1$, taken together with $\frac{d\bar{Q}}{dQ} < 0$ established by (C.13). It also follows from $Q > Q_1$. The third inequality follows from $\bar{Q} > Q$.

To complete the proof of *Proposition 2*, it remains to present the third condition mandatory for Case 4 to emerge, alongside the conditions of $\beta < \beta_1$ and $Q > Q_1$. This condition states that there exists a certain Q^{**} such that Case 4 holds only if $Q < Q^{**}$. To see that note that if Q equals Q_2 then $f'(\bar{Q}) = 0$ and $f(\bar{Q})$ therefore peaks at the right end of the relevant range with a negative value, due to (C.3). Thus, as Q rises, the peak that $f(\bar{Q})$ has falls as *Lemma C.3* shows and turns from positive to negative. Along this way the positive part of $f(\bar{Q})$ shrinks until it vanishes, implying that the peak $C^L(Q, \bar{Q})$ vanishes too.

References

- Bartolini, L., 1993. Competitive runs: The case of a ceiling on aggregate investment. *European Economic Review*, 37(5), 921-948.
- Bartolini, L., 1995. Foreign investment quotas and rent extraction under uncertainty. *Journal of International Economics*, 38, 25-49.
- Calzolari, G., Lambertini, L., 2007. Export restraints in a model of trade with capital accumulation. *Journal of Economic Dynamics and Control*, 31(12), 3822-3842.
- Chao, C. C., Eden, S.H., 2003. Export-performance requirements, foreign investment quotas, and welfare in a small dynamic economy. *Journal of Development Economics*, 72, 387-400.
- Cox, J.C., Ross, S.A., 1976. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3, 145-166.
- Dixit, A.K., 1989. Entry and exit decisions under uncertainty. *Journal of Political Economy*, 97(3), 620-638.
- Dixit, A.K., 1993. *The Art of Smooth Pasting*. Chur, CH: Harwood Academic Publishers.
- Dixit, A.K., Pindyck, R.S., 1994. *Investment and Uncertainty*. Princeton University Press, Princeton.
- Fujiwara, K., 2009. Environmental policies in a differentiated oligopoly revisited. *Resource and Energy Economics*, 31, 239-247.
- Grenadier, S.R., 2002. Option exercise games: an application to the equilibrium investment strategies of firms. *The Review of Financial Studies*, 15, 691-721.
- Hopenhayn, H., 1992. Entry, exit and firm dynamics in long run equilibrium. *Econometrica*, 60, 1127-1150.

- Jovanovic, B., 1982. Selection and the evolution of industry. *Econometrica*, 50, 649-670.
- Katsoulacos, Y., Xepapadeas, A., 1995. Environmental Policy under Oligopoly with Endogenous Market Structure. *The Scandinavian Journal of Economics*, 97, 411-420.
- Lambertini, L., 2010. Oligopoly with Hyperbolic Demand: A Differential Game Approach. *Journal of Optimization Theory and Applications*, 145, 108-119.
- Lambson, V.E., 1992. Competitive profits in the long run. *Review of Economic Studies*, 59, 125-142.
- Leahy, J., 1993. Investment in competitive equilibrium: the optimality of myopic behavior. *Quarterly Journal of Economics*, 108, 1105-1133.
- Mankiw, N.G., Whinston, M.D., 1986. Free Entry and Social Inefficiency. *The RAND Journal of Economics*, 17, 48-58.
- Moretto, M., 2008. Competition and irreversible investment under uncertainty. *Information Economics and Policy*, 20, 75-88.