Optimal capacity allocation in a vertical industry

Antelo, Manel and Bru, Lluís

January 2022

Optimal capacity allocation in a vertical industry

Manel Antelo[a] and Lluís Bru[b]

Abstract

We examine how a social planner should allocate productive capacity in a downstream industry when, upstream, there is an efficient supplier and a set of less efficient suppliers of an essential input. We show that optimal allocation consists of setting a large quota and small quotas for the remaining capacity. This allows the planner, without necessarily harming consumers, to reap licensing rents above those that would be obtained in a competitive downstream market or under public management of capacity. We also discuss circumstances under which a use-or-lose requirement for the large quota is welfare enhancing or welfare reducing, and under which banning price discrimination in the intermediate market may be socially optimal.

JEL classification: D43, F13, L13

Keywords: Capacity allocation, dominant firm, use-or-lose requirement, price discrimination, quota licence, soft-budget constraint

[a] Universidade de Santiago de Compostela (Spain) and ECOBAS. e-mail: manel.antelo@usc.es.
[b] Universitat de les Illes Balears (Spain). e-mail: lluis.bru@uib.es

1. Introduction

Some features of certain industries may reduce the rate at which production flows and, as result, how the market performs. One such feature is physical bottlenecks as occur in airport slots (Katsigiannis and Zografos, 2021), gas pipelines (Petrovich et al., 2017) and electricity transmission links (Cao et al., 2018). Another feature is direct quantity control by a social planner as happens with import quotas (Cadot et al. 2012) and taxi services allocated through licences (Frazzani et al., 2016). In turn, capacity "created" by a social planner may be publicly owned and managed or, alternatively, publicly owned, but privately managed (Walley, 2012), with the social planner sometimes setting rules for private firms based on use-or-lose (UOL) clauses (Dosi and Moretto, 2010; Gale and O'Brien, 2010).[1] On the other hand, it is not uncommon in a (vertical) industry that a main supplier of a basic input enjoys a competitive advantage over rival suppliers. This occurs, for example, with workers organized in a labour union, and also the supply of natural gas in Europe, where, depending on the country and due to the paucity of alternatives, either Algeria, Norway, or Russia is the dominant gas supplier.

A number of key questions can be posed regarding the allocation of production capacity in the downstream segment of an industry in which, upstream, there is a dominant supplier and a set of competitive suppliers of an essential input used by downstream firms. First, how can capacity be allocated if a public monopoly is less efficient than private firms? Or, put differently, how can production capacity be distributed by the planner to guarantee a certain level of wellbeing for consumers and simultaneously capture the maximum income from licensing capacity? Second, what is

[1] See https://www.theguardian.com/world/live/2020/jul/22 for this issue in the European airline industry.

the socially optimal way to manage production capacity, i.e., how does the public or private nature of firms awarded capacity affect the achievement of both objectives? Third, when capacity is allocated, do UOL requirements play a role and what could be the rationale behind such UOL clauses? Fourth, what is the welfare impact of prohibiting price discrimination in intermediate markets in this context? In this paper, we offer an answer to those questions in a vertical industry, where downstream capacity is allocated among firms through a competitive bidding; next, upstream and downstream firms with a given amount of capacity set contracts in the intermediate market to transfer an essential input to produce the final good, and, finally, downstream firms interact in the product market.

First, we demonstrate that the capacity allocation that simultaneously maximizes consumer surplus and licensing revenue consists of granting most (but not necessarily all) capacity to a single firm (which becomes dominant) and the remaining capacity to small firms (the competitive fringe). The dominant firm will use its market power in the final market to obtain better deals in the intermediate market: if the input supplier demands a high wholesale price, the downstream firm can move to (worse) alternative suppliers, and simultaneously partially pass the extra cost through to the final price, thereby reducing use of the input. Due to this possible reaction by the dominant downstream firm, in equilibrium, this firm will pay a lower wholesale price than the fringe firms. This buying clout of the downstream dominant firm translates into higher profits, and therefore, if capacity has been previously allocated in a competitive auction, it will also translate into higher bids for capacity and, as a consequence, into higher revenues for the social planner (the capacity owner). This allows the planner to reap part of the industry rents without necessarily hurting consumers, and provided capacity is auctioned in lots of an adequate size. Below we show how to determine the correct lot size and also show how sometimes it can be socially profitable to sacrifice some consumer surplus to increased revenues from the auctioning of capacity.

Second, we find that auctioning rather than retaining public management of capacity can lead to a better social outcome. The rationale is as follows: if the firm with the greatest capacity operates as a public firm – and so is concerned not just with maximizing profits, but also with consumer wellbeing – it will not pass through any increase in input cost into the final price. This reduces its buying power against the main supplier, who will charge a high input cost. Our analysis thus shows that the privatization of capacity may improve the observed costs of using capacity, in accordance with the literature suggesting that private firms tend to be more efficient than public firms (Thi Minh Phi et al., 2019). However, while the usual rationale (Maskin and Qian, 1999; Bertero and Rondi, 2000) is the soft-budget constraint of public firms, i.e., the possibility that any loss incurred will be covered by public funds, we offer a different but complementary rationale.

Third, our analysis also provides a rationale for the UOL requirements frequently implemented in allocating and managing capacity. We analyse the practicality of imposing UOL clauses and the importance of licence non-transferability. While Gale and O'Brien (2013) show that UOL clauses may or may not improve consumer surplus, depending on whether the dominant firm is more or less efficient than fringe firms, in our analysis we show that UOL clauses can sometimes be combined with capacity allocation to the dominant firm to improve welfare. In relation to licence transferability, Lott (1987) argues in its favour that the licence always goes to the most efficient holder; however, we argue that transferability may increase buying power in the final market at the expense of consumers.

Finally, and given that price discrimination in intermediate markets is an issue in network industries such as telecommunications, gas, electricity, and rail transport, where pricing rules for access to essential facilities are typically set by regulatory bodies (Dertwinkel-Kalt et al., 2013),[2] we examine the impact of banning price discrimination. Our model suggests that, depending on the amount of downstream capacity, banning price discrimination in the intermediate market (i.e., the obligation of imposing a uniform price) may increase licensing revenues, but at the cost of sacrificing some consumer surplus. Thus, legal price discrimination bans adopted by many countries, aimed at avoiding dominant firms charging different prices to different buyers for the same product, may not always have the benefits indicated by Katz (1987), DeGraba (1990), and Yoshida (2000).

Maybe the paper most closely related with ours is that by Ikonnikova and Zwart (2014), regarding downstream buyers negotiating contracts with upstream sellers. They analyse whether intermediate market trade quotas that limit the market share of sellers in the downstream market improve the bargaining power of buyers. In their analysis, the market share of downstream firms is given (they are local monopolies); in contrast, in our analysis, we focus on the role played by capacity allocation in the downstream industry. More generally, our analysis is closely related to Galbraith's countervailing power theory (Galbraith, 1952), which shows how large buyers, relative to small buyers, obtain price concessions from sellers – an argument that has been formalized in several studies (see Snyder, 2008, for a review). In our model, the existence of a dominant supplier upstream leads the social planner, when allocating capacity, to take into account the potential countervailing power of a dominant firm downstream.

---

[2] These access regulations almost always prescribe that access prices to upstream facilities have to be non-discriminatory (see, e.g., Vickers, 1995).

Our model is also related to the literature on import quotas (as in Krishna, 1990 and 1993), whereby a social planner may reap rents by auctioning licensed quotas. However, to the best of our knowledge, previous results show that this can only be achieved at the expense of the consumer surplus, while we show that this is not necessarily the case in our setup.[3]

The rest of the paper is organized as follows. In Section 2 we outline the model, in Section 3 we solve the game, establishing vertical contracts to transfer inputs for production in the final market, in Section 4 we discuss optimal allocation of capacity, and in Section 5 we describe a number of design constraints. Section 6 concludes with some final comments.

2. The model

Consider a vertical industry composed of upstream suppliers of an essential input and downstream firms that purchase the input to manufacture a final good. Upstream, a super-competitive or efficient supplier U, capable of producing the input at a (low) constant marginal cost $c_L \geq 0$, coexists with a fringe of competitive suppliers that produce the input at a higher marginal cost c_H, $c_H > c_L$. The input is transformed into a final product on a one-to-one basis and with no additional cost beyond the price paid in the wholesale market. Downstream production capacity, which is given, amounts to

[3] Krishna (1993) and Campos and Santos (1996) discuss several ways to sell total capacity (sequentially or simultaneously selling in lots of equal or different sizes). In both papers, a dominant firm exists downstream, i.e., a firm already possesses capacity, and one firm has market power, while the other firms are price-takers. The issue they analyse is whether the incumbent outbids potential entrants for capacity.

$X > 0$. The final product is homogeneous and inverse market demand for that product is $P(Q)$, satisfying $P'(Q) < 0$.

Below we analyse industry performance in the following framework. Capacity, which is publicly owned, is divided into quotas summing up to a total amount X, and the quotas are allocated among downstream firms through a competitive bidding process. Once downstream capacity has been allocated, and before production takes place, upstream and downstream firms set contracts in the intermediate market that establish the terms under which the input is transferred. Finally, downstream firms interact in the final (product) market.

In this setting, we focus our analysis of the following key issues. How does the vertical industry perform when the social planner "creates" a dominant downstream firm with capacity Z, $0 \leq Z \leq X$, and a fringe of price-taking firms with remaining capacity $X - Z$? How does a social planner, concerned about both consumer welfare and capacity licensing revenues, allocate capacity X to downstream firms? Should firms in the downstream industry be publicly or privately managed? If the regulator could regulate the practice of price discrimination in the intermediate market, under what conditions would this be allowed or prohibited? Is there a rationale for imposing UOL requirements?

To answer these questions, let us define $Q^m(c_i)$ as the optimal production of a downstream monopolist without capacity constraints at marginal cost c_i, $i = L, H$. Throughout the paper we make use of the following assumptions:

Assumption 1. $c_H < P(X)$.

Assumption 2. $Q^m(c_H) < X$.

Assumption 3. $P''(Q)Q + 2P'(Q) < 0$, for all $Q \geq 0$.

Assumption 1 guarantees that all downstream firms are active, and that, when they do not have market power (to be defined more precisely below), they use all their capacity, even at marginal production cost c_H. By virtue of Assumption 2 we obviate the trivial case in which total capacity X is so small that a monopolist would never restrict the use of capacity, even at marginal production cost c_H. Finally, Assumption 3 is standard in oligopoly theory (see, e.g., Vives, 1999) to ensure that, for unconstrained firms, we may look at first-order conditions to find their optimal production level.

In the intermediate market we assume that the efficient supplier U offers take-it-or-leave-it input supply contracts to downstream firms, and can practice price discrimination, by offering a two-part tariff contract. If a downstream firm rejects the offer, it purchases the input from competitive suppliers at unit price $w = c_H$. Although the alternative input supply source is never used in equilibrium, its existence will affect how market profits are shared between the efficient supplier U and downstream firms.

Finally, when allocating downstream capacity, we assume that the bidding process is competitive and as a consequence, licensing revenues for any quota will be equal to the downstream profits it generates. We consider the creation of a single "large" quota (defined below) that we denote Z; and "small" quotas that sum to the remaining capacity $X - Z$. When deciding the size of the large quota, we analyse two possible objective functions for the planner. In the first objective function, capacity is allocated to maximize licensing revenues, subject to the restriction that market equilibrium leads to full use of capacity. Since we assume competitive bidding, this is equivalent to solving:

$$\max_Z L(Z) = \sum_i \pi_i^D, \quad \text{s.t.:} \quad \sum_i q_i = X \tag{1}$$

In the second objective function of the social planner, the sum of consumer surplus and licensing revenues is maximized:

$$\max_Z W(Z) = CS(\sum_i q_i) + \sum_i \pi_i^D \tag{2}$$

3. The market game equilibrium

In order to analyse the optimal allocation of capacity from the social planner's point of view, we first explore the outcome of market interaction.

3.1. *Final market outcome*

First, we analyse the optimal behaviour of a downstream firm *D*, assigned capacity Z, $0 < Z \leq X$, with marginal production cost c, and with remaining capacity $X - Z$ fully used. Thus, D's profit (gross of any fixed fee paid to the input supplier) amounts to:

$$\pi^D(Z, c) = \max_q \ (P(X - Z + q) - c)q, \ \text{s.t: } 0 \leq q \leq Z \tag{3}$$

We denote the output that solves Eq. (3) as $q(Z, c)$ and total output as $Q(Z, c) = q(Z, c) + X - Z$. Hence, the final price is $P(q(Z, c) + X - Z)$ or, equivalently, $P\big(Q(Z, c)\big)$. When does a downstream firm have an incentive to restrict its output? For this incentive to exist, the firm must possess a sufficiently high level of capacity, as stated in the following lemma.

Lemma 1. *Under Assumption 3, the following hold in the final market:*

(i) *If $X < Q^m(c)$, then downstream firm D uses all its capacity Z. Formally, $q^D(Z, c) = Z$ or, equivalently, $Q(Z, c) = X$ for all $Z \in [0, X]$, whenever $X < Q^m(c)$.*

(ii) *If $Q^m(c) < X < P^{-1}(c)$, then downstream firm D only uses all its capacity Z if lower than a cut-off level $Z(c)$. Formally, $Q^m(c) < X < P^{-1}(c)$ implies that there exists $Z(c) \in (0, X)$ for which $Q(Z, c) = X$ iff $0 < Z \leq Z(c)$ and $Q(Z, c) < X$ iff $Z(c) < Z \leq X$. Furthermore, the cut-off value $Z(c)$ is decreasing in c, and when $Z > Z(c)$, total output $Q(Z, c)$ is strictly decreasing in both Z and c.*

Proof. See the Appendix.

According to Lemma 1, when its rivals use all their aggregate capacity $X - Z$, the behaviour of a downstream firm with capacity Z depends simultaneously on its own capacity and the marginal wholesale price it must pay (either w or c_H). Supposing that the marginal cost is w, with $w < c_H$, then there is a capacity threshold $Z(w)$ that satisfies $Z(w) > Z(c_H)$, such that this firm would restrict production only if Z satisfies $Z > Z(w)$, and would otherwise produce at full capacity. Moreover, the capacity threshold $Z(w)$ increases if the upstream supplier reduces the marginal wholesale price w. However, a downstream firm with capacity $Z < Z(c_H)$ would never restrict production, even if its marginal production cost was c_L. The content of Lemma 1 can be easily illustrated in the following two examples.

Linear demand. If demand for the final product is linear, $P(Q) = a - Q$, with $a > 0$, and firms have marginal production cost c, then a downstream firm would only have idle capacity when Z is above $Z(c) \equiv a - X - c$. Note that $\frac{\partial Z(c)}{\partial X} < 0$, i.e., the threshold level $Z(c)$ is decreasing in total capacity X.

Constant elasticity demand. If demand for the final product is isoelastic, $P(Q) = a/Q^s$, with $0 < s < 1$, and firms have marginal cost c, then a downstream firm only restricts production when its capacity Z is above $Z(c) \equiv \left(1 - \frac{c}{P(X)}\right)\frac{X}{s}$.

3.2. *Intermediate market outcome*

We now turn to the analysis of the input market. Downstream firms' production costs are determined by $T(q) = F + wq$, the contract offered by the efficient supplier U. For the sake of simplicity, we assume that U has all the bargaining power in negotiating the transfer of the input, and hence, downstream firms face a take-it-or-leave offer.[4] Thus, a downstream firm will accept the contract whenever net profit is at least $\pi^D(Z, c_H)$, the profit obtained with the alternative source of the input. Hence, U sets the fee F equal to $\pi^D(Z, w) - \pi^D(Z, c_H)$, which is the increased profit of the downstream firm. The downstream firm has profit $\pi^D(Z, c_H)$, i.e., the value of its outside option of not accepting U's offer, and U's contract features a marginal wholesale price w that maximizes $F + (w - c_L)q(Z, w)$, or $\left(P\left(X - Z + q(Z, w)\right) - c_L\right)q(Z, w) - \pi^D(Z, c_H)$.

[4] We could allow, without changing the qualitative results, some sharing of rents between firms, such that the profits of a downstream firm would be given by $\beta\pi^D(Z, c_L) + (1 - \beta)\pi^D(Z, c_H)$. with $0 < \beta < 1$.

Below, Lemma 2 summarizes the equilibrium contracts in the intermediate market, as a function of the capacity Z of the downstream firm.

Lemma 2. *The following hold in the intermediate market:*

(i) *A downstream firm with capacity $Z < Z(c_L)$ is offered an input contract that leads it to produce at full capacity, $q(Z,w) = Z$, while a downstream firm with capacity $Z > Z(c_L)$ is offered an input contract that features $w = c_L$ and leads it to produce $q(Z, c_L)$.*

(ii) *A downstream firm with capacity $Z < Z(c_H)$ pays $w = c_H$ per unit of input, whereas a downstream firm with capacity $Z > Z(c_H)$ pays $w < c_H$.*

According to part (i) of Lemma 1, downstream firm D is offered a contract such that it produces $q(Z, c_L)$, the level of production that maximizes joint profits, and it obtains market profit $\pi^D(Z, c_L)$. This firm also has the option of acquiring the input from less efficient (competitive) suppliers at wholesale price $w = c_H$, in which case its profit is $\pi^D(Z, c_H)$. Therefore, the efficient supplier U can appropriate the increase in D's profit by imposing the fixed fee $F = \pi^D(Z, c_L) - \pi^D(Z, c_H)$ in the input contract, according to which D's net profit is in fact $\pi^D(Z, c_H)$. This is the profit firm D considers when we analyse how candidates for licences bid for capacity.

Given that we model fringe firms as total capacity $X - Z$, and given that Assumption 1 implies that fringe firms always produce at full capacity, we need not be very specific about how fringe firms purchase the input. This is certainly the case for any firm with capacity below $Z(c_H)$. In any case, the efficient supplier will find it optimal to sell the input to fringe firms at a wholesale price slightly below c_H, and the fringe profit per unit of capacity will be $V^f = P(X) - c_H$.[5]

Part (ii) of Lemma 2 explicitly accounts for buying power as a function of capacity *Z*. A "large" downstream firm, i.e., a firm with capacity level $Z > Z(c_H)$, pays a lower price for input in the intermediate market than any fringe firm. According to Snyder (1996), it is buyer size in the intermediate market (through the power to break collusion among input suppliers) that awards buying power. In our model, in contrast, it is buyer size in the final market (the power to impact the final good price) that awards buying power in the intermediate market.[6]

Before we examine capacity bidding in the next section, note that aggregate profits of downstream firms may increase with consolidation (D's size) through two different channels: an increase in the buying power of D against efficient supplier U, and an increase in the level of collusion in the final market.

If we assume that all downstream firms have capacity $Z < Z(c_H)$, the downstream industry is competitive and aggregate profits amount to $V^f X = (P(X) - c_H)X$. The first channel emerges when there is a downstream firm D with capacity $Z(c_H) < Z \leq Z(c_L)$, and a set of fringe firms with remaining capacity $X - Z$. In equilibrium, the dominant firm uses all its capacity. Hence, final prices, fringe profit

[5] Throughout the paper V^f denotes the expected profit per unit of capacity for a fringe firm and for a given market structure.

[6] Note, however, that the average input price does not necessarily decrease with buyer size. In the linear demand case, for instance, a downstream firm with capacity *Z* above $Z(c_L) = A - X - c_L$ is asked to pay a fixed fee $F = \pi^D(Z, c_L) - \pi^D(Z, c_H)$ for the input and, in equilibrium, it produces $q(Z, c_L)$. Thus, the average wholesale price actually paid amounts to $\frac{T(q(Z,c_L))}{q(Z,c_L)} = \frac{\pi^D(Z,c_L) - \pi^D(Z,c_H)}{q(Z,c_L)} + c_L = \frac{1}{2}\frac{(A-(X-Z)-c_L)^2-(A-(X-Z)-c_H)^2}{A-(X-Z)-c_L} + c_L$. It follows that, for $Z > Z(c_L)$, the average input price increases, $\frac{\partial}{\partial Z}\left\{\frac{T(q(Z,c_L))}{q(Z,c_L)}\right\} = \frac{1}{2}\frac{(c_H-c_L)^2}{(A-(X-Z)-c_L)^2} > 0$.

(per unit of capacity), and consumer surplus do not change with changes in D's capacity. Aggregate downstream profits increase in Z solely because downstream firm D increases its buying power against U. Formally:

$$\frac{\partial[\pi^D(Z,c_H)+V^f(X-Z)]}{\partial Z} = P\big(X-Z+q(Z,c_H)\big) - P\big(X-Z+q(Z,c_L)\big) > 0 \quad (4)$$

The second channel through which aggregate profits of downstream firms may increase with the level of consolidation emerges if the dominant firm D has capacity $Z > Z(c_L)$. In equilibrium, D now restricts output, $Q(Z,c_L) < X$, and hence, final prices are above $P(X)$. Thus, a very large D results in fringe firms achieving higher profits and consumers obtaining a smaller surplus. Fringe firm profit per unit of capacity amounts to $V^f = P\big(X-Z+q(Z,c_L)\big) - c_H$ and the fact that:

$$\frac{\partial V^f(Z)}{\partial Z} = -P'\big(X-Z+q(Z,c_L)\big)\left(1-\frac{\partial q(Z,c_L)}{\partial Z}\right) = -\frac{(P')^2}{2P'+P''q} > 0 \quad (5)$$

means that profit increases with D's size. Total output also decreases with Z since:

$$\frac{\partial\big(X-Z+q(Z,c_L)\big)}{\partial Z} = -\left(1-\frac{\partial q(Z,c_L)}{\partial Z}\right) = -\frac{(P')^2}{2P'+P''q} < 0 \quad (6)$$

and therefore, consumer surplus decreases with D's size. Finally, aggregate downstream firm profits increase in Z since:

$$\frac{\partial\big[\pi^D(Z,c_H)+V^f(Z)(X-Z)\big]}{\partial Z}$$

$$= -P'\big(X-Z+q(Z,c_H)\big)q(Z,c_H) - V^f(Z) + \frac{\partial V^f(Z)}{\partial Z}(X-Z)$$

$$= P\big(X-Z+q(Z,c_H)\big) - P\big(X-Z+q(Z,c_L)\big) - \frac{\left(P'\right)^2}{2P'+P''q}(X-Z) > 0 \quad (7)$$

Hence, if downstream firm D has capacity $Z(c_H) < Z \leq Z(c_L)$, there is a redistribution of profits from the efficient supplier U to downstream firms, while consumers remain unaffected. However, if D's capacity is larger, $Z > Z(c_L)$, there is not only a redistribution of profits from the efficient supplier U to downstream firms, but also an increase in the price of final good and hence a more collusive outcome in that market.

4. Bidding for capacity and optimal capacity allocation

Below we analyse the optimal allocation of downstream capacity by the social planner. Consider first a social planner concerned with maximizing licensing revenues from the allocation of capacity X, but also wanting to guarantee that, in equilibrium, all capacity is used. We consider the creation of at most one large downstream firm,[7] and we assume there are enough bidders for competitive bidding to be the reference for evaluating licensing revenues. The planner licenses one possibly large quota Z and many small quotas (each one smaller than $Z(c_H)$) summing to capacity $X - Z$.

According to Lemma 1, if the social planner wants to guarantee that, in equilibrium, all capacity will be used, then the large quota to be granted, Z, must satisfy $0 < Z < \min\{Z(c_L), X\}$. Under competitive bidding, the bid for the large quota is $\pi^D(Z, c_H)$, and those from the small quotas sum to $(P(X) - c_H)(X - Z)$. Thus, the planner chooses the quota Z that solves:

[7] We want to simplify the discussion of market interaction in the final market. Further, the idea of the paper is that the available capacity is limited, as otherwise, competition in the final market would be strong.

$$\max_{0<Z<\min\{Z(c_L),X\}} L(Z) = \pi^D(Z, c_H) + (P(X) - c_H)(X - Z) \tag{8}$$

which affords the following result.

Proposition 1. *The size of the large quota that maximizes licensing revenues is* $Z^* = min\{Z(c_L), X\}$.

According to Krishna (1990), the only way the social planner can raise revenues is by reducing the consumer surplus. We argue, however, that this is not necessarily the case: by creating a downstream firm with buying power, it is possible to raise revenues without harming consumers. The planner can increase the buying power of capacity users without harming consumers, because, if quotas are chosen adequately, rents are merely redistributed within the (home) industry, leaving the final price unaffected. The novel result of our model holds from the fact that we depart from Krishna's (1990) setup – the foreign firm is a monopolist – in that we assume the existence of an alternative (less efficient) source for the input.

4.1. *Privatizing capacity management and the soft-budget-constraint*

According to Proposition 1, and under Assumption 1 that total capacity exceeds $Q^m(c_H)$, licensing revenues can be increased without reducing consumer surplus if the social planner auctions off a large quota of size Z^*. The (dominant) firm with this licence has buying power in the intermediate market, and thus, additional profit per unit of capacity is generated, translating into a larger licensing payment for the large quota.

If instead of privatizing all downstream capacity X through a bidding process, what if the social planner assigns capacity $Z \leq X$ to a publicly managed firm P and the remaining capacity $X - Z$ to a set of privately-managed fringe firms? And to what extent is privatization of the downstream dominant firm socially efficient? To investigate this issue, note that, from Assumption 2, and assuming that social interests predominate, public firm P will use all its capacity under any input source, that is, $q(c_H) = q(c_L) = Z$. However, this leads it not to have buying power against the efficient supplier U since its reservation value amounts to:

$$\pi^P(Z, c_H) = (P(X) - c_H)Z, \quad (9)$$

which is the same buying power per unit of capacity as that of any fringe firm.

Proposition 1 shows that if downstream firms are given the opportunity to bid for capacity Z satisfying $Z(c_H) \leq Z \leq Z(c_L)$, they are willing to pay more than the profit of a public firm. Hence, privatization of the dominant firm allows the planner to raise more revenues than would be obtained through public management of that firm, while the final price, at $P(X)$, remains unaffected and likewise consumer surplus. Note further that if the capacity of the public firm is sold in a fragmented way, thus creating a competitive final market, fringe firms do not perform any better than the public firm, and so the revenues raised by the capacity seller obtains would equal the profit of the public firm.

The above analysis accords with the fact that public firms tend to be less efficient than private firms, which are typically tougher in negotiations with suppliers, unions, etc., and so tend to produce at a lower marginal cost. A frequently suggested rationale for the lower efficiency of public firms is their soft-budget constraint, since any loss suffered will be accounted for by public funds (Maskin, 1996; Maskin and Qian, 1999). Our paper restates the soft-budget constraint theory: compared to a private firm, a well-intentioned public firm has less buyer power against upstream suppliers, because it cannot credibly commit to price increases in a final market not served by the most efficient supplier.

4.2. *Is it socially optimal to use all existing capacity?*

Proposition 1 referred to capacity allocations in which all capacity is used in equilibrium. We now consider a social planner that, as stated in Eq. (2), wishes to maximize the sum of consumer surplus and licensing revenues, $W(Z) = CS\big(X - Z + q(Z, c_L)\big) + L(Z)$, and investigate how capacity is now allocated among downstream firms, and the circumstances that lead to full or partial use of capacity. We can state the following result.

Proposition 2. *Let* $E(c) \equiv \frac{P''(X-Z+q(Z,c))}{P'(X-Z+q(Z,c))} q(Z,c)$*. If* $E'(c) \geq 0$*, a social planner seeking to maximize social welfare, defined as* $W(Z) = CS\big(X - Z + q(Z, c_L)\big) + L(Z)$*, chooses* $Z^* = min\,\{Z(c_L), X\}$.

Proof. To abbreviate notation, we write q_i for $q(Z, c_i)$ and Q_i for $X - Z + q_i$. Social welfare is defined as:

$$W(Z) = CS\big(X - Z + q(Z, c_L)\big) + L(Z)$$

$$= \left(\int_0^{Q_L} P(s)ds - P(Q_L)Q_L \right) + \pi^D(Z, c_H) + V^f(X - Z)$$

where $V^f = P(Q_L) - c_H$. If $Z(c_H) < Z \leq Z(c_L)$, we know from the previous analysis that $Q_L = X$ and that:

$$\frac{\partial W}{\partial Z} = -P'(Q_H)q_H - (P(X) - c_H) = P(Q_H) - P(X) > 0$$

If, on the other hand, $Z(c_L) < Z \leq X$, then:

$$\frac{\partial W}{\partial Z} = P'(Q_L)q_L \left(1 - \frac{\partial q_L}{\partial Z}\right) - P'(Q_H)q_H - (P(Q_L) - c_H)$$

$$= -(P(Q_L) - c_L)\left(1 - \frac{\partial q_L}{\partial Z}\right) + \left(P(Q_H) - P(Q_L)\right) =$$

(write $1 - \frac{\partial q_L}{\partial Z}$ as $\frac{1}{2+E(c_L)}$ and $P(Q_H) - P(Q_L)$ as $\int_{c_L}^{c_H} P'(Q_i)\frac{\partial q_i}{\partial c}dc = \int_{c_L}^{c_H} \frac{1}{2+E(c)}dc$)

$$= -\frac{P(Q_L)-c_L}{2+E(c_L)} + \int_{c_L}^{c_H} \frac{1}{2+E(c)}dc = -\frac{P(Q_L)-c_H}{2+E(c_L)} - \frac{c_H-c_L}{2+E(c_L)} + \int_{c_L}^{c_H} \frac{1}{2+E(c)}dc$$

$$= -\frac{P(Q_L)-c_H}{2+E(c_L)} - \int_{c_L}^{c_H} \left\{\frac{1}{2+E(c_L)} - \frac{1}{2+E(c)}\right\} dc.$$

In the last equation, the first term is negative and the second term is non-positive if $E'(c) \geq 0$. Thus, $\frac{\partial W}{\partial Z} > 0$, which completes the proof of the stated result. ■

It is immediate to see that social welfare $W(Z)$ increases when a larger quota in the interval $Z(c_H) < Z \leq Z(c_L)$ is chosen, since consumer surplus is unaffected (in equilibrium there is full capacity use), while aggregate downstream profits increase (leading to increased licensing revenues), because the dominant firm *D* obtains a larger discount on input prices.

When we analyse the welfare impact of a marginal increase in quota Z above $Z(c_L)$, it is instructive to consider Proposition 2 in relation to the effect of a large quota when $c_H = c_L$. Total welfare is $W(Q) = \int_0^{Q_L} P(s)ds - c_L Q_L$, which is increasing in the level of production, $W'(Q) = P(Q) - c_L > 0$ for $Q \leq X < P^{-1}(c_L)$, which is the standard result – that production increases welfare as long as the marginal utility of an extra unit exceeds the (marginal) cost of producing it. Written in terms of the large quota Z, we have $W(Z) = CS\big(X - Z + q(Z, c_L)\big) + L(Z) = \int_0^{Q_L} P(s)ds - c_L Q_L$, and in terms of Z above $Z(c_L)$, we have $W'(Z) = -\frac{P(Q_L) - c_L}{2 + E(c_L)} < 0$, i.e., the increase in the quota decreases welfare, because $P(X) - c_L > 0$, and the importance of the effect depends on how much production is reduced when we increase Z, $\frac{1}{2+E(c_L)} > 0$.

When $c_H > c_L$, the quota achieves a discount on input prices that is measured by the price difference between the equilibrium and off-equilibrium paths (the process terminates if D rejects U's contract offer), $P(Q_H) - P(Q_L)$. This beneficial effect is compared with the collusive effect of the higher final price that reduces consumer surplus and increases downstream profits. The importance of this second effect depends on the inherent scarcity of capacity and on the reduction in total production when quota Z is increased. While capacity scarcity always implies a negative outcome of an increased quota, measured by $= -\frac{P(Q_L) - c_H}{2 + E(c_L)}$, there is now a second negative part to the collusive effect, measured by $-\frac{c_H - c_L}{2 + E(c_L)}$, i.e., the reduction in production when the marginal cost is not c_H but c_L. This second effect is not higher than the positive effect of the discount on input prices, but only if $E'(c) \geq 0$.

This sufficient condition for the social planner to prefer full capacity use is satisfied, for instance, by linear demand and constant elasticity demand, so it is quite a general condition. However, an exception is demand function $P(Q) = 1 - Q^s$ with $s > 1$, as it does not fulfil that condition.[8] Hence, it can be easily shown that for any capacity only slightly below $P^{-1}(c_H)$, the social planner is willing to sacrifice some consumer surplus, but when capacity is further reduced, capacity scarcity once again becomes a concern, as the social planner seeks full use of capacity.

Remark 1 (the rationale against transferable quotas). From Proposition 2 we can infer that it is crucial to ban the transferability of quotas, as otherwise the large downstream firm has an incentive to acquire capacity from fringe firms and become an even larger firm, with capacity above $Z(c_L)$. This would lead to a reduction in production, $Q < X$, and a reduction in welfare.

Consider that the dominant downstream firm D acquires capacity and ends up with total capacity Z above $Z(c_L)$. To acquire capacity $\Delta \equiv Z - Z(c_L)$ from fringe firms, D must pay their expected per unit profit $b(Z) \equiv P\big(X - Z + q(Z, c_L)\big) - c_H$, which leads D's profits to be:

$$\pi(\Delta) = \pi^D(Z, c_H) - b(Z)(Z - Z(c_L)).$$

Hence, it follows that:

$$\pi'(Z) = \frac{\partial \pi^D(Z, c_H)}{\partial Z} - b(Z) - b'(Z)(Z - Z(c_L))$$

$$= P\big(X - Z + q(Z, c_H)\big) - P\big(X - Z + q(Z, c_L)\big) - \frac{P\big(X - Z + q(Z, c_L)\big) - c_L}{2 + E(c_L)} \frac{Z - Z(c_L)}{q_L} \quad (10)$$

[8] For this demand function, it follows that $E(c) \equiv (s-1)\frac{q(Z,c)}{X - Z + q(Z,c)}$. Hence, $E(c)$ is increasing in $q(Z, c)$, and since the dominant firm reduces its production if *c* increases, we conclude that $E'(c) < 0$.

The dominant downstream firm obtains a discount on the input price, reflected in the first term of Eq. (10), $P(X-Z+q(Z,c_H)) - P(X-Z+q(Z,c_L)) > 0$. Against this, downstream small, fringe firms increase their profits at the expense of consumers, as now, the larger the dominant firm becomes, the more it restricts output in equilibrium, $Q < X$, yielding a more collusive final market. However, collusion is more profitable for fringe firms operating at full capacity. This is reflected in the second term of Eq. (10), $b'(Z - Z(c_L)) = \frac{P(X-Z+q(Z,c_L))-c_L}{[2+E(c_L)]q_L} > 0$. Finally, since $\pi'(Z(c_L)) = P(X - Z + q(Z,c_H)) - P(X-Z+q(Z,c_L)) > 0$, then by continuity it follows that $Z^* > Z(c_L)$.

Linear demand. If market demand is linear, $P(Q) = a - Q$, the dominant downstream firm obtains additional capacity $Z^* - Z(c_L) = \min\{c_H - c_L, X - Z(c_L)\}$:. Hence, if the existing capacity is small, $X < \frac{a+c_H-2c_L}{2}$, transferable quotas can lead to a monopolization of the capacity.

Remark 2 ("sleeping quotas"). "Sleeping quotas", i.e., import quotas that are not fully used (Campos and Santos, 1996), are sometimes observed, and their existence as an optimal choice for the social planner is easily explained in our model. For a total import quota X, the social planner must set a large quota of size $Z^* = \min\{Z(c_L), X\}$ to maximize welfare. The social planner may alternatively auction off licences that amount to a total import quota X' that satisfies $X < X' < P^{-1}(c_H)$, setting one large quota of size $Z' = (X' - X) + Z^*$. In terms of both consumer surplus and licensing revenues, in equilibrium, the market outcome is equivalent to selling a total level of licences X and a quota of size Z^*; since the fringe is of the same size, $X' - Z' = X - Z^*$, the dominant firm will produce the same amount $q(Z(c_L), c_i)$ in both cases. A sleeping quota will be observed in the second case, however, since the dominant firm will not use licences $Z' - Z(c_L)$ in equilibrium.

4.3. *Demand uncertainty*

Uncertainty in market demand could be another reason for a social planner needing to sacrifice some consumer surplus if the dominant downstream firm is private. To examine this issue, we consider the example of a linear demand function $P(Q) = a_i - Q$, with two demand states: a low-demand state a_1 occurring with $\text{Prob}(a_1) = \mu$, $0 < \mu < 1$, and a high-demand state a_2, where $0 < a_1 < a_2$. We further normalize $c_L = 0$ and write $c_H = c > 0$. We make use of Assumptions 1 and 2, translated to the uncertainty setup; thus, $\frac{a_2-c}{2} < X < a_1 - c$, and we also assume that $a_1 \leq a_2 - c$. As before, Assumption 1 guarantees that fringe firms have strictly positive profits for any realization of demand and any cost level, while Assumption 2 guarantees that there is enough capacity to create a dominant firm that obtains discounts in the high-demand state. Condition $a_1 \leq a_2 - c$ simplifies the analysis, since when a firm has a quota of size $Z \leq Z(a_1, 0) = a_1 - X$, it has buying power only in the low-demand state.[9] Finally, we also assume that the demand state is revealed only after capacity has been allocated, but before input contracts are signed.[10]

We look for the optimal choice of a social planner that seeks to maximize the expected welfare defined in Eq. (2), i.e., the sum of consumer surplus and licensing

[9] When there is more than one state of demand, we write the Lemma 1 capacity threshold $Z(a_i, c_i)$ at which a downstream firm restricts production as a function of the intercept of demand a_i and the marginal cost c_i.
[10] Similar results can be derived if the demand state is revealed later.

revenues. According to Lemma 1, for a given demand state a_i and for marginal production cost c_i, the dominant downstream firm D restricts production whenever its capacity satisfies $Z > Z(a_i, c_i) \equiv a_i - X - c_i$. Only two quota size scenarios are optimal:[11] (a) firm D never restricts output, i.e., it has capacity $Z_1 = Z(a_1, 0) = a_1 - X$; or (b) firm D does not restrict output in the high-demand state, i.e., it has capacity $Z_2 = \min\{Z(a_2, 0), X\} = \min\{a_2 - X, X\}$.

When firm D has capacity Z_1 as defined above, downstream capacity is fully used in equilibrium, and hence the consumer surplus amounts to $CS = \frac{X^2}{2}$ for any demand state. But downstream firm D obtains discounts in the intermediate market only in the low-demand state. Thus, downstream profits are $\left(\frac{a_1-(X-Z_1)-c}{2}\right)^2 + (a_1 - X - c)(X - Z_1)$ in the low-demand state and $(a_2 - X - c)X$ in the high-demand state. The sum of consumer surplus and downstream profits at $Z = Z_1$ yields:

$$W(Z_1, a_1) = \frac{X^2}{2} + \left(\frac{a_1-(X-Z_1)-c}{2}\right)^2 + (a_1 - X - c)(X - Z_1)$$

for low demand, and

$$W(Z_1, a_2) = \frac{X^2}{2} + (a_2 - X - c)X,$$

[11] See the proof of Proposition 3 for details.

for high demand, while expected welfare amounts to:

$$EW(Z_1) = \mu\, W(Z_1, a_1) + (1 - \mu) W(Z_1, a_2) \tag{11}$$

When downstream firm *D* has capacity Z_2, it restricts production in equilibrium when demand is low, i.e., $Q(Z_2, a_1) < X$, and consumer surplus is therefore negatively affected, although in exchange, firm *D* has buying power in both states of demand. Hence, we can write the sum of consumer surplus and downstream profits as:

$$W(Z_2, a_1) = \begin{cases} \frac{{a_1}^2}{8} + \left(\frac{a_1 - c}{2}\right)^2, & \text{if } \ X < \frac{a_2}{2} \\ \frac{(a_1 + X - Z_2)^2}{8} + \left(\frac{a_1 - (X - Z_2) - c}{2}\right)^2 + \left(\frac{a_1 - (X - Z_2)}{2} - c\right)(X - Z_2), & \text{if } \ \frac{a_2}{2} < X \end{cases}$$

in the low-demand state and

$$W(Z_2, a_2) = \begin{cases} \frac{X^2}{2} + \left(\frac{a_2 - c}{2}\right)^2, & \text{if } \ X < \frac{a_2}{2} \\ \frac{X^2}{2} + \left(\frac{a_2 - (X - Z_2) - c}{2}\right)^2 + (a_2 - X - c)(X - Z_2), & \text{if } \ \frac{a_2}{2} < X \end{cases}$$

in the high-demand state. Thus, expected welfare amounts to:

$$EW(Z_2) = \mu S(Z_2, a_1) + (1 - \mu) S(Z_2, a_1) \quad (12)$$

In the low-demand state, total welfare is higher if the larger quota is Z_1, since this is the largest quota that will be fully used, and thus, $W(Z_1, a_1) > W(Z_2, a_1)$. In the high-demand state, total welfare is higher if the larger quota is Z_2, since it is fully used, and firm *D* pays a lower price for the input, and hence, $W(Z_2, a_2) > W(Z_1, a_2)$. Direct comparison of expected welfare achieved under both quotas, $EW(Z_1)$ in Eq. (11) and $EW(Z_2)$ in Eq. (12), allows us to obtain the following result.

Proposition 3. *The optimal quota is either* $Z_1 = a_1 - X$ *or* $Z_2 = min\{a_2 - X, X\}$. *It is optimal to create a downstream firm with a quota* Z_2 *that restricts production in the low-demand state whenever the probability* μ *of a low-demand state is infrequent in the sense of* $0 \leq \mu \leq \bar{\mu}$, *where* $0 < \bar{\mu} \equiv \frac{W(Z_2,a_2) - W(Z_1,a_2)}{(W(Z_1,a_1) - W(Z_2,a_1)) + (W(Z_2,a_2) - W(Z_1,a_2))} < 1$ *takes the value:*

$$\bar{\mu} = \begin{cases} \dfrac{8\left[\left(\frac{a_2 - c}{2}\right)^2 - (a_2 - X - c)X\right]}{8\left[\left(\frac{a_2 - c}{2}\right)^2 - (a_2 - X - c)X\right] + 4\left[\left(\frac{a_1}{2} - c\right)^2 - (a_1 - X - c)^2\right]}, & if \quad \dfrac{a_2 - c}{2} < X < \dfrac{a_2}{2} \\ \dfrac{2c^2}{4(a_2 - a_1)(a_1 - X - c) + (a_2 - a_1)^2 + 2c^2}, & if \quad \dfrac{a_2}{2} < X < 2a_1 - a_2 \end{cases}$$

Proof. First, we check that we can only have two local maxima, Z_1 and Z_2. For a quota $Z \leq a_1 - X - c$, capacity is fully used and all downstream firms pay *c* for the input. Hence, expected welfare is constant for these values of *Z* at:

$$EW(Z) = \mu\left(\frac{X^2}{2} + (a_1 - X - c)X\right) + (1-\mu)\left(\frac{X^2}{2} + (a_2 - X - c)X\right)$$

For a quota $a_1 - X - c < Z \leq a_1 - X$, expected welfare amounts to:

$$EW(Z) = \mu\left(\frac{X^2}{2} + \left(\frac{a_1-(X-Z)-c}{2}\right)^2 + (a_1 - X - c)(X - Z)\ \right) + (1-\mu)\left(\frac{X^2}{2} + (a_2 - X - c)X\right)$$

and it increases in Z because capacity is fully used, and firm *D* obtains increasing discounts in the low-demand state, since $EW'(Z) = \ \mu\frac{Z-(a_1-X-c)}{2} > 0$.

For a quota $a_1 - X < Z \leq a_2 - X - c$, expected welfare

$$EW(Z) = \mu\left(\frac{(a_1+X-Z)^2}{8} + \left(\frac{a_1-(X-Z)-c}{2}\right)^2 + \left(\frac{a_1-(X-Z)}{2} - c\right)(X - Z)\ \right) + (1-\mu)\left(\frac{X^2}{2} + (a_2 - X - c)X\right)$$

is decreasing in Z. In the low-demand state, production decreases, which, from Proposition 2, leads to a welfare decrease, while in the high-demand state, Z is not large enough for *D* to obtain discounts, since $EW'(Z) = \ -\mu\frac{a_1-X-2c+Z}{4} < -\mu\frac{a_1-X-c}{2} < 0$. Hence, we have a local maximum at $Z = Z_1$.

For a quota $a_2 - X - c < Z \leq Z_2$, expected welfare is a convex function in Z, and increasing at Z_2 for a sufficiently low value of μ. In the low-demand state, production decreases, which, from Proposition 2, decreases welfare; however, in the high-demand state, a bigger Z leads to a larger discount while full capacity use is maintained, since:

$$EW'(Z) = \ -\mu\frac{a_1-X-2c+Z}{4} + (1-\mu)\frac{Z-(a_2-X-c)}{2}$$

and

$$EW''(Z) = \ \frac{2-\mu}{4} > 0.$$

Finally, for a quota above $Z > Z_2$, expected welfare is decreasing in Z according to Proposition 2. Hence, there is another local maximum at Z_2 if μ is small enough. Comparison of $EW(Z_1)$ and $EW(Z_2)$ leads to Z_2 being the global maximum, whenever $0 \leq \mu \leq \bar{\mu}$. ■

4.4. *UOL clause*

For the dominant downstream firm to have buying power in the intermediate market, the ability to leave some capacity unused is crucial. It is a usual practice in the public allocation of capacity that firms are subject to minimum use of their capacity – e.g., UOL requirements in co-tenancies (Gale, 1994) and landing slots (Gale and O'Brien, 2013) – or banning sleeping quotas for import quotas (Campos and Santos, 1996), where firms may lose their quota rights if they leave capacity idle.

In our model, we can interpret a UOL requirement as a minimum level of production $\underline{q}$ that the dominant downstream firm must satisfy, $\underline{q} \leq q$.[12] In the absence of demand uncertainty, our analysis suggests that a requirement $\underline{q}$ below $q(Z(c_L), c_H)$ is innocuous, whereas a requirement $\underline{q}$ satisfying $q(Z(c_L), c_H) < \underline{q} \leq q(Z(c_L), c_L)$ is counterproductive, because the value of the outside option is reduced:

$$(P(X - Z(c_L) + \underline{q}) - c_H)\underline{q} < \pi^D(Z(c_L), c_H) \qquad (13)$$

which reduces the dominant firm's buying power in negotiation with the supplier. As a consequence, the input price discount is lower, and the dominant downstream firm uses all its capacity anyway in equilibrium. Therefore, sleeping quota bans (in the case of

[12] We are not concerned with fringe firms, since they always work at full capacity.

import licences) and UOL requirements restrict the rents that can be accrued from auctioning off licences.

However, demand uncertainty offers a positive role for a UOL requirement.[13] Although it reduces the dominant firm's buying power if $\underline{q}$ is above the intended level of production, in case of disagreement, it can moderate the negative impact of large quotas if firm D restricts production in the low-demand state.

Consider thus that there is demand uncertainty (as in Section 4.3), the dominant firm D receives a quota Z that satisfies[14] $a_2 - X - c \leq Z \leq \min\{a_2 - X, X\}$ and faces a UOL requirement, $\underline{q} \leq q$. The welfare impact of the UOL clause depends on the level $\underline{q}$ at which the clause is established, as follows:

If $\underline{q}$ is below the level of production in the case of disagreement when demand is low, i.e., $\underline{q} \leq \frac{a_1-(X-Z)-c}{2}$, then firm D's production is never affected. Hence, the UOL requirement has no impact on expected welfare.

If $\underline{q}$ satisfies $\frac{a_1-(X-Z)-c}{2} < \underline{q} \leq \frac{a_1-(X-Z)}{2}$, expected welfare decreases in $\underline{q}$, since, when demand is low, D's buying power decreases and its production does not change in equilibrium.

If $\underline{q}$ satisfies $\frac{a_1-(X-Z)}{2} < \underline{q} \leq \frac{a_1-(X-Z)-c}{2}$, then, when $\underline{q}$ increases and demand is low, firm D's buying power decreases, but the consumer surplus:

---

[13] In 2020, the EU authorities reacted to the coronavirus epidemic by suspending the UOL rules on European airport landing slots, freeing airlines to halt "ghost flights" in which planes have been taking off without any passengers. See The Guardian 2020.07.22 (https://www.theguardian.com/world/live/2020/jul/22).

[14] It is immediate to see that a UOL requirement cannot improve expected welfare when the quota is $Z < a_2 - X - c$.

$$W\left(Z,\underline{q};a_1\right)=\frac{\left(X-Z+\underline{q}\right)^2}{2}+\left(a_1-(X-Z)-\underline{q}-c\right)\left(X-Z+\underline{q}\right)$$

increases, since $\frac{\partial W\left(Z,\underline{q};a_1\right)}{\partial \underline{q}}=a_1-(X-Z)-\underline{q}-c>0$ in this interval.

If $\underline{q}$ is such that $\frac{a_1-(X-Z)-c}{2}<\underline{q}\leq Z$, firm D's buying power is reduced when demand is high, $W\left(Z,\underline{q};a_2\right)=\frac{X^2}{2}+\left(a_2-(X-Z)-\underline{q}-c\right)\underline{q}+(a_2-X-c)(X-Z)$ and $\frac{\partial W\left(Z,\underline{q};a_2\right)}{\partial \underline{q}}=a_2-(X-Z)-2\underline{q}-c<0$. Overall, the effect of $\underline{q}$ on expected welfare is:

$$\mu\frac{\partial W\left(Z,\underline{q};a_1\right)}{\partial \underline{q}}+(1-\mu)\frac{\partial W\left(Z,\underline{q};a_2\right)}{\partial \underline{q}}=$$

$$=\mu\left(a_1-(X-Z)-\underline{q}-c\right)+(1-\mu)\left(a_2-(X-Z)-2\underline{q}-c\right) \quad (14)$$

which leads to:

$$\underline{q}(Z)=\frac{\mu(a_1-X-c)+(1-\mu)(a_2-X-c)+Z}{2-\mu} \quad (15)$$

as a local maximum.

If an active UOL requirement is implemented, i.e., one that affects firm D's behaviour, the best UOL is the clause stated in Eq. (15), which changes both buying power and production level in both demand states. We can then check what the optimal quota Z is and what the optimal UOL requirement is when an active UOL clause is implemented. Since expected welfare is:

$$\frac{\partial EW\left(Z,\underline{q}(Z)\right)}{\partial Z} = \mu\frac{\partial W\left(Z,\underline{q}(Z);a_1\right)}{\partial Z} + (1-\mu)\frac{\partial W\left(Z,\underline{q}(Z);a_2\right)}{\partial Z}$$

$$= -\mu(a_1 - X - c) - (1-\mu)(a_2 - X - c) + \underline{q}(Z) - \mu Z \qquad (16)$$

and Eq. (16) is convex in Z, $\frac{\partial^2 EW\left(Z,\underline{q}(Z)\right)}{\partial Z^2} = \frac{(1-\mu)^2}{2-\mu} > 0$, we obtain the following result.

Lemma 3. *If an active UOL requirement is implemented, the best quota is* $Z^* = Z_2 = min\{a_2 - X, X\}$*, and the best UOL requirement is* $\underline{q}(Z_2) = \frac{\mu(a_1-X-c)+(1-\mu)(a_2-X-c)+Z_2}{2-\mu}$.

Note that the optimal active UOL requirement $\underline{q}(Z_2)$ depends not just on the size of the quota Z_2, but also on parameter values. Therefore, our result of Lemma 3 suggests that the usual regulatory practice of setting a UOL clause as a percentage of the quota is poorly adapted to market circumstances. Moreover, to set a UOL clause is not always an optimal policy, as can be seen in the following proposition for given sets of parameter values.

Proposition 4. *Let* $c_q = 2(a_2 - X) - \sqrt{2(a_2 - X)^2 + 2(a_1 - X)^2 + (a_2 - a_1)^2}$, c_e *the value of* c *in the interval* $0 < c < a_2 - a_1$, *for which* $\mu_{12} = \mu_{1q} = \mu_{1q}$, $\mu_{2q} =$

$2\frac{(a_2-a_1)((a_2+3a_1-4X-4c)-4(a_1-X)c+c^2}{(a_2-a_1)((a_2+3a_1-4X-4c)-4(a_1-X)^2}$ *and* $\mu_{1q} = \frac{2(a_1-X)c+c^2-c\sqrt{4(a_1-X)c+c^2}}{2(a_1-X)^2+c^2}$. *Thus, the following hold for certain sets of parameter values:*

(a) *If* $\frac{6}{7}a_2 < a_1 < a_2$ *and* $\frac{a_2}{2} < X < \frac{7a_1-5a_2}{2}$, *the optimal regulatory policy is as follows:*

(i) To set the quota $Z_2 = a_2 - X$ *if* $c_q < c < a_2 - a_1$ *and* $0 < \mu < \mu_{2q}$.

(ii) To set the quota $Z_2 = a_2 - X$ *and the UOL* $\underline{q}(Z_2) \leq q$ *if* $0 < c < a_2 - a_1$ *and* $max\{0, \mu_{2q}\} < \mu < \mu_{1q}$.

(iii) To set the quota $Z_1 = a_1 - X$ *if* $\mu_{1q} < \mu < 1$.

(b) *If* $\frac{3}{4}a_2 < a_1 < \frac{6}{7}a_2$ *or* $\frac{6}{7}a_2 < a_1 < a_2$ *and* $\frac{7a_1-5a_2}{2} < X < 2a_1 - a_2$, *the optimal regulatory policy is as follows:*

(i) To set the quota $Z_2 = a_2 - X$ *if* $c_q < c < c_e$ *and* $0 < \mu < \mu_{2q}$ *or* $c_e < c < a_2 - a_1$ *and* $0 < \mu < \mu_{12}$.

(ii) To set the quota $Z_2 = a_2 - X$ *and UOL requirement* $\underline{q}(Z_2) \leq q$ *if* $0 < c < c_e$ *and* $max\{0, \mu_{2q}\} < \mu < \mu_{1q}$.

(iii) To set the quota $Z_1 = a_1 - X$ *if* $0 < c < c_e$ *and* $\mu_{1q} < \mu < 1$ *or* $c_e < c < a_2 - a_1$ *and* $\mu_{12} < \mu < 1$.

Thus, there are always parameter values such that one of the policies we compare is optimal. Whenever low demand is frequent (i.e., the probability μ of low demand is sufficiently high), the optimal regulatory policy is to guarantee that, in equilibrium, all capacity is used, $Z^* = Z_1$. Conversely, setting quota Z_2, either with or without a UOL requirement, is socially optimal only when consumers are infrequently harmed (i.e., the probability μ is low). The UOL requirement is not beneficial when c is high, i.e., when the dominant firm has little buying power. In equilibrium, this firm is expected to produce at marginal cost zero and, hence, it is important to retain as much buying power as possible.

Figures 1 and 2 depict regions in terms of parameter values for the regulatory policies stated in Lemma 3. For the indicated parameter values, a monopoly downstream is never created,[15] $Z_1 < Z_2 < X$. As both figures show, the addition of a UOL requirement enlarges the set of parameter values for which quota Z_2 is chosen (without UOL requirement, the region where Z_1 is chosen is enlarged up to the broken line).

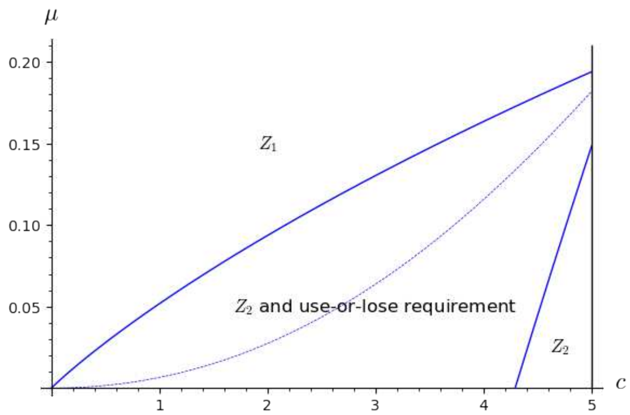

Figure 1. The regulatory policy stated in Proposition 4 (Parameter values: $a_1 = 45, a_2 = 50, X = 30$, and $c \in [0,5]$)

[15] Similar regions emerge if we consider a situation where the existing capacity is low, $X < \frac{a_2}{2}$, and, as a consequence, one of the policies is to create a monopoly downstream, $Z_2 = X$. But because capacity is more scarce, the social planner chooses, for a larger set of the other parameter values, a quota of size Z_1 to guarantee that all existing capacity is always used.

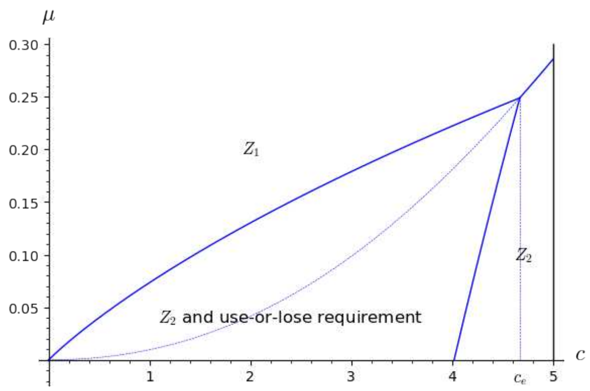

Figure 2. The regulatory policy stated in Proposition 4 (Parameter values: $a_1 = 45, a_2 = 50, X = 35$, and $c \in [0,5]$).

4.5. *The welfare impact of banning price discrimination in the intermediate market*

Below we examine the impact on welfare of a regulatory policy consisting of banning price discrimination in the intermediate market. The sequence of events is as follows. In the first stage (regulatory stage), the social planner announces a ban on price discrimination before the quota bidding stage, so firms evaluate the value of quotas under the expected uniform wholesale price, and licensing revenues in the competitive bidding process will be the aggregate profits of downstream firms. In the second stage (bidding stage), the social planner sets quotas, with a possible large quota *Z*. In the third stage (input contracting stage), the main supplier sets the same wholesale price *w* for all downstream firms. Finally, in the fourth stage (final market stage), downstream firms produce. The welfare measure we consider (the sum of consumer surplus and downstream profits) is the same as before.

In this context, a ban on price discrimination will be welfare improving only if the upstream supplier charges a wholesale price *w* strictly below c_H, which will benefit fringe firms, and, in addition, compensates for any decrease in the profits of the dominant downstream firm if the resulting wholesale price leads to a lower discount than would be obtained from the upstream firm when a two-part tariff contract is allowed.

We limit our analysis to a setting in which demand is linear, $P(Q) = 1 - Q$, and cost values are normalized to $c_L = 0$ and $0 < c_H = c < 1$. Assumptions 1 and 2 are maintained, which implies that feasible values for capacity X are restricted to the interval $\left[\frac{1-c}{2}, 1 - c\right]$.

If the upstream supplier chooses an input price $w \leq c$, then, according to the result obtained in Lemma 1, production of a downstream firm with quota Z is:

$$q(w) = \begin{cases} Z, \text{ if } Z < 1 - X - w \text{ or } w < 1 - X - Z \\ \frac{1-(X-Z)-w}{2}, \quad \text{otherwise} \end{cases}$$

that is, production of a downstream firm with quota $Z < 1 - X - c$ is always at full capacity, whereas production of a downstream firm with a larger quota will depend on the wholesale price set by the upstream supplier. Lemma 4 states when it is feasible to incentivize the upstream supplier to set a wholesale price below c, and indicated the quota Z^l that maximizes social welfare.

Lemma 4. *The upstream supplier sets an input price* $w^l < c$ *only when* $c > \frac{1}{2}$ *, and hence, it is optimal to create a monopoly downstream,* $Z^l = X$*. As a result, the input* *price is* $w^l = \begin{cases} 1-2X, & if \quad 0 < X < \frac{1}{4} \\ \frac{1}{2}, & if \quad \frac{1}{4} < X < \frac{1}{2} \end{cases}$, *total production amounts to* $Q^l = \begin{cases} X, & if \quad 0 < X < \frac{1}{4} \\ \frac{1}{4}, & if \quad \frac{1}{4} < X < \frac{1}{2} \end{cases}$, *and welfare is* $W^l = CS^l + \pi^l = \begin{cases} \frac{3X^2}{2}, & if \quad 0 < X < \frac{1}{4} \\ \frac{3}{32}, & if \quad \frac{1}{4} < X < \frac{1}{2} \end{cases}$.

Proof. If a downstream firm with quota $1 - X - c < Z \leq X$ is created,[16] its profit amounts to $(1 - (X - Z) - w - q)q$, and it is maximized by producing

$$q = \begin{cases} Z, & if \quad 0 < w < 1 - X - Z \\ \frac{1-(X-Z)-w}{2}, & if \quad 1 - X - Z < w < c \end{cases}$$

Consider now the problem of the upstream supplier. For $0 < w < 1 - X - Z$, profit is $\pi^u = Xw$ and, hence, the optimal wholesale price in this interval is $w = 1 - X - Z < c$, while, for $1 - X - Z < w < c$, profit is $\pi^u = \frac{1+(X-Z)-w}{2} w$.

Note that $\left.\frac{\partial \pi^u}{\partial w}\right|_{w=c} = 1 + X - Z - 2c > 0$ for any size of quota Z if $c > \frac{1}{2}$; therefore $w < c$ can only be achieved if $c < \frac{1}{2}$ through a quota of size $Z > 1 + X - 2c$. On the other hand, $\left.\frac{\partial \pi^u}{\partial w}\right|_{w=1-X-Z} = Z - (1 - 3X) > 0$ only if the quota satisfies $1 - 3X < Z \leq X$, which is only feasible if $X > \frac{1}{4}$. Therefore, when $X < \frac{1}{4}$, it follows that $w^l = 1 - 2X < c$, total production is $Q^l = X$, and aggregate welfare is $W^l = \frac{3X^2}{2}$, as stated in Lemma 4.

[16] If $0 < Z \leq 1 - X - c$, the downstream firm produces $q = Z$ for any price in the intermediate market, and the upstream firm chooses $w = c$.

Meanwhile, if $X > \frac{1}{4}$, depending on quota Z, the upstream firm sets the wholesale price:

$$w = \begin{cases} 1 - X - Z, & if \quad 1 - X - Z < Z < 1 - 3X \\ \frac{1}{2}, & if \quad 1 - 3X < Z < X \end{cases}$$

and, as a consequence, total production amounts to:

$$Q = \begin{cases} X, & if \quad 1 - X - Z < Z < 1 - 3X \\ \frac{1 + X - Z}{2}, & if \quad 1 - 3X < Z < X \end{cases}$$

Hence, the wholesale price decreases as quota Z increases, but there is no full use of capacity, while production decreases as quota Z increases if Z exceeds $1 - 3X$.

In the interval $1 - 3X < Z < X$, total welfare is $W = CS + \pi = \frac{Q^2}{2} + (1 - w - Q)Q = \frac{(3-5X+5Z)(1+X-Z)}{32}$, which is increasing in Z, $\frac{\partial W}{\partial Z} = \frac{1+5X-5Z}{16} > 0$; therefore, it is also optimal to create a monopoly downstream when $X > \frac{1}{4}$. ■

Under a linear price in the intermediate market, the social planner creates a monopoly downstream (i.e., banning price discrimination in the intermediate market becomes equivalent to forcing the upstream supplier to offer the input at a linear price). Strikingly, the social planner's choice means that full capacity is not used when $\frac{1}{4} < X < \frac{1}{2}$.

We now analyse whether welfare increases if price discrimination in the intermediate market is banned. If $c < \frac{1}{2}$, banning price discrimination decreases welfare, because the upstream supplier then sets a wholesale price $w = c$, while price discrimination allows the big downstream firm to obtain a price discount. On the other hand, when $X \leq \frac{1}{4}$, the upstream supplier offers the input at the linear price $w^* = 1 - 2X$ and, as result, there is full use of capacity, $q(w^*) = X$. Hence, the profits of downstream firms are $\pi^l = (1 - X - w^*)X = X^2$, which are larger than the profits under non-linear (two-part tariff) prices in the intermediate market, where the regulator could also choose $Z = X$ and profits would be $\pi^{nl} = \left(\frac{1-c}{2}\right)^2$.

Finally left to compare is the welfare outcome when $\frac{1}{2} < c \leq \frac{3}{4}$ and $\frac{1}{4} < X < 1 - c$. To repeat, under linear wholesale prices the planner creates a monopoly downstream that restricts production, and welfare amounts to $W^l = CS^l + \pi^l = \frac{3}{32}$. The alternative for the planner is to allow the efficient upstream supplier to set a two-part tariff contract, and the optimal quota size is again $Z = X$, i.e., it is likewise optimal to create a monopoly downstream, but now there is full use of capacity in equilibrium. Hence, downstream profits (or licensing revenues) are $\pi^{nl} = \left(\frac{1-c}{2}\right)^2$ and total welfare amounts to $W^{nl} = CS^{nl} + \pi^{nl} = \frac{X^2}{2} + \left(\frac{1-c}{2}\right)^2$, and $W^{nl} > W^l$ requires $X > \overline{X} \equiv \frac{\sqrt{3-8(1-c)^2}}{4}$. To summarize, we obtain the result stated in the following proposition.

Proposition 4. *Prohibiting the sale of input by means of a non-linear (two-part tariff) contract leads to:*

(i) *A welfare increase in the (c, X)-region of parameters where $c \geq \frac{1}{2}$ and $X \in \left[\frac{1-c}{2}, min\{\overline{X}, 1 - c\}\right]$.*

(ii) *An increase in licensing revenues without affecting consumer surplus if* $X \in \left[\frac{1-c}{2}, \frac{1}{4}\right]$.

(iii) *An increase in licensing revenues at the cost of sacrificing consumer surplus if* $X \in \left[\frac{1}{4}, min\{\overline{X}, 1-c\}\right]$.

In Proposition 4 (i), the social planner sets a monopoly downstream, whereas in (iii) there is no full use of capacity. Figure 3 depicts two regions in which the social planner increases licensing revenues in the (*c*,*X*)-space: the region in which the efficient supplier is restricted to sell the input by means of a linear price, due to price discrimination in the intermediate market is banned (the region within the red contour) and the region in which the efficient supplier is allowed to sell the input through a non-linear two-part tariff contract (the region outside the red contour).

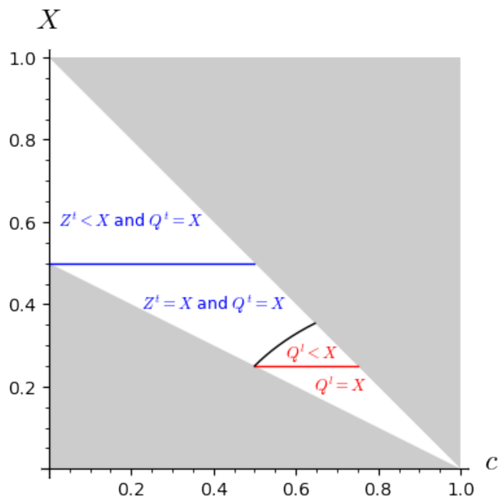

Figure 3. Prohibition of price discrimination in the intermediate market and licensing revenues.

5. Conclusion

We investigated how a social planner should allocate a given amount of productive capacity in the downstream segment of an industry in which an efficient upstream supplier and a set of less efficient upstream suppliers provide an essential input for downstream firms can produce the final good. We show that the optimal allocation consists of allocating capacity by means of a single large quota and a set of small quotas for the remaining capacity. This allows the planner, without necessarily harming consumers, to extract more licensing rents than would be obtained by dividing all capacity into small quotas (a competitive downstream market) or by managing capacity itself (public capacity management). Thus, the creation of a dominant downstream firm is not necessarily socially bad. We also discuss the circumstances under which a use-or-lose requirement for the large quota is either welfare enhancing or welfare reducing, and also the circumstances under which banning price discrimination in the intermediate market may be socially optimal.

In our framework, we assumed just a single large quota. This assumption may not be very limiting, given the thrust of the paper, namely a market where capacity is scarce. Moreover, this simplifies the analysis of intermediate market interactions, as more than one large downstream firm would allow the upstream firm to consider the anticompetitive policies analysed in the foreclosure literature (Rey and Tirole, 1996).

In our view, a more important limitation of our analysis is that we have considered just a single market downstream. An analysis of the consequences of interaction between different markets would be the most important extension of the research reported above. This is left for future research.

References

Bertero, E. and L. Rondi (2000), Financial pressure and the behaviour of public enterprises under soft and hard budget constraints: evidence from Italian panel data, *Journal of Public Economics* 75, 73-98.

Binmore, K. and P. Klemperer (2002), The biggest auction ever: The sale of the British ·G Telecom licenses, *Economic Journal* 112, C74-C96.

Campos, L. and V. Santos (1996), Sleeping quotas, pre-emptive quota bidding and monopoly Power, *Journal of International Economics* 40, 127-48.

Cao, K.-K., J. Metzdorf and S. Birblata (2018), Incorporating Power Transmission Bottlenecks into Aggregated Energy System Models, *Sustainability* 10(6):1916.

Gahvari, F. (1989), Licensing and nontransferable rents: A comment, *American Economic Review* 79, 906-909.

Gale, I. and D.P. O'Brien (2013), The welfare effects of use-or-lose provisions in markets with dominant firms, *American Economic Journal: Microeconomics* 5(1), 175-93.

Ikonnikova, S. and G.T. Zwart (2014), Trade quotas and buyer power, with an application to the EU natural gas market, *Journal of the European Economic Association* 12(1), 177-199.

Krishna, K. (1990), The case of the vanishing revenues: Auction quotas with monopoly, *American Economic Review* 80, 828-836.

Krishna, K. (1993), Theoretical implications of imperfect competition on quota license prices and auctions, *World Bank Economic Review* 7, 113-136.

Lott, J.R. (1987), Licensing and nontransferable rents, *American Economic Review* 77, 453-455.

Maskin, E. and C. Xu (2001), Soft budget constraint theories, *Economics of Transition* 9, 1-27.

Rey, P. and J. Tirole (1996), A Primer on Foreclosure, mimeo, IDEI.

Snyder, C. (2008), Countervailing power, in S.N. Durlauf and L.E. Blume (eds.), *The New Palgrave Dictionary of Economics*, second edition, Palgrave Macmillan.

Thi Minh Phi, N., F. Taghizadeh-Hesary, C. Anh Tu, N. Yoshino, and C. Ju Kim (2019), Performance differential between private and state-owned enterprises: An analysis of profitability and leverage, Asian Development Bank Institute https://www.adb.org/sites/default/files/publication/503476/adbi-wp950.pdf

Vives, X. (1999), *Oligopoly Pricing: Old ideas and New Tools,* The MIT Press, Cambridge, Massachusetts.

Walley, P. (2012), Does the public sector need a more demand-driven approach to capacity management?, *Production Planning and Control* 24, 1-14.