ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

COMPETITION FOR SCARCE RESOURCES

Peter Eso, Volker Nocke and Lucy White

Number 365
October 2007

Manor Road Building, Oxford OX1 3UQ

Competition for Scarce Resources*

Péter Eső
Kellogg School of Management, Northwestern University

Volker Nocke
University of Oxford and CEPR

Lucy White
Harvard Business School, HEC Lausanne and CEPR

This version: July 14, 2007

Abstract

We show that the efficient allocation of production capacity can turn a competitive industry and downstream market into an imperfectly competitive one. Even though downstream firms have symmetric production technologies, the downstream industry structure will be symmetric only if capacity is sufficiently scarce. Otherwise it will be asymmetric, with one large "fat" capacity-hoarding firm and a fringe of smaller "lean and fit" firms, so that Tobin's Q varies inversely with firm size. This is so even if the number of firms is infinitely large. As demand or input quantity varies, the industry may switch between symmetric and asymmetric phases, generating predictions for firm size and costs across the business cycle. Surprisingly, an increase in available capacity resulting in such a switch can cause a reduction in total output and consumer surplus.

*We thank Liam Brunt, Luis Cabral, Zhiqi Chen, Jacques Crémer, Joe Harrington, Massimo Motta, Rob Porter, Bill Rogerson, Xavier Vives, and seminar participants at the Northwestern-Toulouse IO Workshop (2005), the IIOC conference in Boston (2006), the University of Bern, Bocconi University, the CEPR Applied Industrial Organization workshop in Madeira (2006), the ESEM in Vienna (2006), INFORMS in Pittsburgh (2007) and ESSET in Gerzensee (2007), for valuable comments.

1 Introduction

Standard models of industrial organization treat inputs as being in perfectly elastic supply and their trade disconnected from the downstream market. However, in many real-world industries the firms that compete downstream also face each other in the input market where supply is inelastic. For example, jewelry makers that vie for the same customers also compete for precious stones whose supply is limited.[1] Competing airlines divide a fixed number of landing slots at a given airport; software companies that produce competing operating systems draw from the same pool of highly specialized programmers; and so on. Retailers of gas (petrol) also have a common input that is in scarce supply.

In this paper we investigate the *interaction* between efficient input markets and competitive downstream industries, and find some unexpected results. We study a model where firms with the same decreasing-returns technology compete first for scarce production "capacity" in an input market. We model the input market as a Vickrey auction; but the same capacity allocation would be reached in a uniform-price share auction or via Coasian barganing as well. Firms then compete downstream à la Cournot subject to their resulting capacity constraints.

We show that strategic behavior in the input market can transform an otherwise symmetric downstream industry into a natural oligopoly with endogenous asymmetries. When the amount of total capacity is sufficiently large, the only equilibrium of the game is asymmetric: one firm is large and seemingly inefficient, whilst the other firms are small, efficient, and capacity-constrained. This result can explain the "size-discount" puzzle that there is a negative relationship between Tobin's Q and firm size in the data.[2] In contrast, when the total capacity is lower than a certain threshold, all firms are of equal size. We show that this conditional industry structure persists even as the number of firms approaches infinity. Perhaps the most surprising result is that as the total available capacity *increases* and crosses the threshold, total downstream production *decreases* and so do consumer and social welfare. Our model also suggests that the size distribution of firms should be more asymmetric in recessions than in booms.

When the total available capacity is large, the asymmetric size distribution arises because one firm finds it optimal to hoard capacity. This behavior keeps up the market price of output for two reasons. First, it limits the other firms' production by

[1] See Spar (2006) for an account of the diamond market.

[2] See, for example, Lang and Stulz (1994), Eeckhout and Jovanovic (2002).

making them capacity-constrained. Second, as the large firm acquires more capacity, it is more willing to leave a larger fraction of it unused. The large firm thus appears to be a "fat cat", purchasing too much capacity and using it very inefficiently; whereas the other firms appear "lean and mean", making high rates of profit despite their low capacity purchases. Nevertheless, the small, productive firms do not expand to steal the downstream market from the large, inefficient firm.[3] An econometrician or policy-maker observing such a situation might suspect that some unobserved regulation, illegal anti-competitive behavior or political influence protect the large firm from its more efficient rivals. But our model helps us to understand that this is not necessarily the case: the asymmetric outcome may simply be the result of standard non-cooperative behavior.

Our results can help to explain the size distribution of firms in industries where inputs are scarce, which may be very asymmetric even though it is far from clear that the largest firms enjoy any cost advantage over the smaller firms. For example, De Beers, the firm that dominates the diamond market, has followed a strategy of buying up uncut diamonds and hoarding them in order to maintain the price of cut diamonds (Spar (2006)). Similarly, in the UK it was documented that the large petrol companies were buying up petrol-retailing forecourts (gas stations) and removing this essential input from the market by filling the underground tanks with concrete (MMC (1990)). Another example may be Microsoft, the dominant firm in the market for PC operating systems, which employs more software engineers and yet has a slower update cycle than some of its rivals (e.g., Apple).[4] Famously, Salomon Brothers engineered a "short squeeze" in the market for Treasury Bonds, submitting bids for up to 94% of the Treasury bonds available at auction in order to monopolize the secondary market in these securities (Jegadesh (1993) shows that after-market prices were significantly higher as a result). Finally, the prevailing market structure at many large airports is that one airline hoards most of the landing slots (see Borenstein (1989) and (1991)). Of course, in many of these industries, capacity or input is not literally sold in an "efficient auction". However, we show that the same allocation can be generated if

[3]In fact, the small firms may benefit from the presence of the large firm, making more profits than they would if the input were distributed symmetrically by fiat (thus resulting in a symmetric Cournot outcome). Indeed, depending upon parameters, the large firm may be providing a public good in buying up the excess input to reduce the supply of output.

[4]Walter Mossberg, an influential technology writer notes (WSJ, 2004 September 14): "[T]he Mac operating system is more capable, more modern and more attractive than Windows XP, and just as stable." The number of programmers working on Microsoft's OS is estimated to be an order of magnitude greater than the number of Apple's engineers.

inputs are sold in a uniform-price share auction, or indeed, are simply allocated by Coasian bargaining among the firms themselves.

The threshold level of capacity at which production switches from being symmetric to being asymmetric varies according to the level of demand downstream. Starting at the threshold with symmetric firms (which is the socially most efficient point), an expansion in demand generates a rise in the price of output, but the industry remains symmetric. By contrast, a contraction in demand causes all firms but one to shrink their output, while the remaining firm absorbs the excess capacity. Thus the efficient capacity auction exacerbates the procyclicality of output across the cycle because there is input hoarding during recessions.

Our results have important policy implications for industries with scarce inputs. While a government may be concerned about the emergence of a dominant firm in such markets, we show that encouraging downstream entry will not help much. Even as the number of downstream firms tends to infinity, the equilibrium when available capacity is large remains asymmetric and uncompetitive, resembling the textbook model of a dominant firm constrained by a competitive fringe. But in contrast to that model, downstream output does not converge to competitive levels in our model and the one-firm concentration ratio remains bounded away from zero.[5]

More surprisingly, encouraging upstream entry as a response to input scarcity might even make things worse. Perhaps the most unexpected result of our model is that an *increase* in the quantity of input (capacity) available can result in a *reduction* in the total quantity of output. For low capacity levels, all firms are symmetric and capacity-constrained. A well-intentioned government might try to encourage additional provision of the scarce input so as to increase output, benefitting consumers. But we show that near the capacity threshold such an attempt would be misguided, because increasing capacity beyond the threshold actually results in a discrete reduction in output and consumer surplus. This reduction in output results from the switch to the asymmetric allocation of input, which, as noted above, is productively inefficient. The total profit of downstream firms is continuous at the capacity-threshold (they are indifferent between the symmetric and asymmetric allocations) but total output and consumer surplus fall discontinuously due to the introduction of production inefficiencies.

In recent years, governments have employed economists to help them (re-)design

[5] There are also some technical differences. In the dominant firm model, unlike in ours, the dominant firm sets price, taking as given the supply curve of the competitive fringe.

markets for the allocation of scarce inputs.[6] The typical prescription has been that the old "beauty contests" (in the case of spectrum) or rigid structures of bilateral contracts and vertical integration (in the case of electricity and gas) should be replaced by centralized auction markets to place the input in the hands of those who value it the most. Our results suggest that this prescription is misplaced in a context where the purchasing firms compete downstream. It is not entirely surprising that an efficient auction, since it maximizes the bidders' surplus, may allocate the input in a way that results in a lack of downstream competition.[7] Perhaps more surprising is that an "efficient" auction will result in production inefficiencies in the presence of diseconomies of scale or complementarities between inputs. Our model suggests that allocating input by some more decentralized means and restricting resale amongst firms might actually be better for consumers, contrary to intuition.

Related literature. One of the first papers to study the interaction between upstream and downstream markets is that of Stahl (1988). He studies a model where middlemen bid first for a homogenous good to be later resold to consumers via price competition. He finds that, as long as the sales-revenue maximizing price in the second stage is lower than the Walrasian price that would arise in the absence of middlemen, the same (Walrasian) price prevails in their presence. This outcome roughly corresponds to the special case of our model where the firms' marginal cost is constant and available capacity is less than the amount a downstream cartel would need to produce its optimal output. Our focus is different, though, in that we are interested in the asymmetric firm size distribution that occurs when total capacity is large, and the surprising effects that arise when switching between the symmetric and asymmetric regimes.[8]

Yanelle (1997) studies a variant of the Stahl model applied to the banking industry.

---

[6] Examples include the sale of electicity by generating companies to retailers; the sale of licences to mobile phone operators; the sale of oil tracts to production companies; the sale of forestry tracts to logging companies; the sale of Treasury Bills at auction.

[7] This idea originates in Borenstein (1988). However, the result is not completely evident. For example, McAfee (1999) argues that, when large and small incumbents compete in an auction to purchase an additional unit of capacity, a small (constrained) firm will win the auction if there are at least two large (unconstrained) firms. McAfee does not consider the full dynamic game in which capacity is acquired over time. We show that, when inputs are allocated simultaneously, the symmetric outcome that he identifies is no longer an equilibrium.

[8] Stahl (1988) also studies the case where the Walrasian price is lower than the sales-revenue maximizing one. Assuming a random tie-breaking rule in the first stage, he finds that the middlemen bid for the right to be a monopoly in the second stage, where the winner charges a downstream price between the monopoly price and the sales-revenue maximizing one, and all middlemen make zero total profits.

In her model, banks compete with entrepreneurs to acquire funds from small investors, and also with other banks to finance loans to entrepreneurs. Unlike Stahl, she allows banks to ration potential depositors, and shows that, depending on parameter values, the equilibrium may be in pure strategies (as in Stahl) or in mixed strategies. In addition, financial disintermediation can arise, whereby banks obtain no funds from depositors, and firms (inefficiently) obtain direct finance from investors.

These models are close to ours in that they model the presence of "middlemen" competing both in input and output markets. Less closely related is the large literature on vertical relations (see Rey and Tirole (2007) for a recent survey). Typically these models differ from ours in that they assume that input is sold to downstream firms through bilateral contracting rather than through a centralized market. An exception is Salinger (1988). He sets out a model of "successive oligopoly" where Cournot firms sell intermediate goods to downstream firms which also compete à la Cournot to sell to consumers. Allocation in the input market is always symmetric and is not efficient. Downstream firms act as price-takers in the input market despite their strategic interdependence.[9]

In explaining the asymmetric size distribution of firms, our paper contributes to a sizeable literature. Ghemawat (1990) studies a duopoly model where the initially larger (but not more efficient) competitor ends up absorbing all investment opportunities in order to keep product prices high. His model involves price competition subject to capacity constraints. Besanko and Doraszelski (2004) set up a general dynamic investment game and show that that when firms compete in prices, an asymmetric market structure arises; but that the outcome is symmetric under Cournot competition. In contrast to these two papers, in our model, we have Cournot (quantity) competition in the downstream market, yet we end up with an asymmetric allocation when the total available capacity is large, and not when it is small. Our work is also related to Riordan (1998), who shows that a dominant firm facing a competitive fringe can benefit from raising its rivals' costs by acquiring upstream capacity which is in imperfectly elastic supply; but he assumes rather than derives the asymmetric market structure which arises endogenously in our model.

One interpretation of the equilibrium capacity allocation in our model is that it is the optimal allocation of a cartel in which firms collude on capacity but not on

[9] Another difference between this "vertical relations literature" and our work is that typically this literature assumes that upstream firms make take-it-or-leave-it offers to downstream firms and downstream firms make take-it-or-leave it offers to consumers, so that no firm has price-setting power in both markets. In this respect, our work has more in common with the the literature on "middlemen" mentioned above.

quantity (or price) setting. Our paper is therefore related to the industrial organization literature on 'semi-collusion'. In this literature, it is typically assumed that firms collude on the transitory strategic variable (price or quantity) but not on investment levels; see, for instance, Benoît and Krishna (1987) and Davidson and Deneckere (1990).[10]

Another body of work to which we contribute is the literature on auctions with externalities. In our paper, downstream competition among bidders imposes a particular structure on the externalities between them, which allows us to derive more specific results than has generally been possible in that literature (see also Katz and Shapiro (1986) and Jehiel and Moldovanu (2000), who consider the case when rival firms bid for a patent). Closer to our paper in spirit is Hoppe, Jehiel and Moldovanu (2006). They consider an industry (e.g., mobile telephony) where additional licenses to operate are allocated among incumbents and potential entrants. Incumbents can deter entry by acquiring licenses, but entry deterrence is a public good. Hoppe, Jehiel and Moldovanu show that if the number of licenses per incumbent is not an integer then increasing (or reducing) this number to an integer may help deter entry. The argument relies on the assumption that firms can coordinate on buying only an *equal* number of licenses each. If the number of licenses per incumbent is not an integer, it is assumed that the incumbents play the symmetric mixed strategy equilibrium (rather than an efficient, asymmetric equilibrium), which sometimes results in inefficient miscoordination and entry. Our paper also has the feature that increasing the scarce resource can reduce output, but we always select efficient equilibria, and indeed the structure of our model is completely different.

Plan of the paper. The paper is organized as follows. In Section 2, we outline the model and derive preliminary results. In Section 3, we derive the unique equilibrium of our game—which is Cournot competition following the efficient allocation of capacities—and characterize several of its properties. We analyze the limiting case, in which the number of firms grows infinitely large, and we discuss welfare. In Section 4, we describe some testable predictions of our model, such as the relationship between Tobin's Q and firm size. In Section 5, we show that our main results continue to hold when competition between firms is differentiated-goods Bertrand rather than homogeneous-goods Cournot. We discuss the robustness of our results in in Section 6. Section 7 concludes. All omitted proofs are collected in an Appendix.

[10] Exceptions include d'Aspremont and Jacquemin (1988), where firms may form an R&D cartel but otherwise do not collude, and Nocke (2007), who analyzes a dynamic game where firms may collude on investment in quality.

2 Model and Preliminary Results

The model is a two-stage game where in the first stage n ex-ante identical firms are allocated *production capacities* so that each unit of capacity ends up with the firm that values it the most. The procedure, which is for now treated as a "black box", may be an efficient auction, or efficient Coasian bargaining among the firms.[11] Then, in the second stage, the same firms compete—à la Cournot and subject to their capacity constraints—in a market for a homogenous good. The firms' production technologies exhibit increasing marginal costs, and the market demand is downward sloping. The participants have no private information, everything is commonly known.

In this section we introduce the notation that formally describes this model, and perform some preliminary analysis. In particular, we characterize certain benchmarks and solve for the unique equilibrium of the second-period subgame (Cournot competition with capacity constraints). This enables us to derive the equilibrium market structure and discuss its properties in Section 3.

2.1 Notation and Assumptions

Denote the total available capacity by K, and the capacities of the firms, determined in the first-period auction (or through efficient Coasian bargaining), by k_i, $i = 1, \ldots, n$, where $\sum_i k_i = K$.

Denote the inverse demand function in the downstream market by $P(Q)$, where Q is the total production. We assume that P is twice differentiable, and that both $P(Q)$ and $P'(Q)Q$ are strictly decreasing for all $Q > 0$. Firm i's cost of producing $q_i \leq k_i$ units is $c(q_i)$, while its cost of producing more than k_i units is infinity. We assume that c is twice differentiable, strictly increasing, and strictly convex. Finally, we assume that producing a limited amount of the good is socially desirable: $P(Q) - c'(Q)$ is positive for $Q = 0$, and negative as $Q \to \infty$.

We can write the profit of firm i in the downstream market for quantity $q_i \leq k_i$ and total output from firms other than i, Q_{-i}, as

$$\pi_i(q_i, Q_{-i}) = P(Q_{-i} + q_i)q_i - c(q_i). \tag{1}$$

The marginal profit of firm i given the other firms' total production is $\partial\pi_i/\partial q_i = P'(Q_{-i} + q_i)q_i + P(Q_{-i} + q_i) - c'(q_i)$.

[11] We further discuss the efficiency criterion in Section 2.3. We discuss auction rules that yield an efficient capacity allocation for the industry in Section 4.1.

The assumptions on the market demand and individual cost functions made above are standard in the literature. They ensure that π_i is concave in q_i, and that the quantities are strategic substitutes,

$$\frac{\partial^2 \pi_i}{\partial q_i \partial Q_{-i}} \equiv P''(Q_{-i}+q_i)q_i + P'(Q_{-i}+q_i) < 0,$$

where the inequality follows from $d\left[P'(Q)Q\right]/dQ = P''(Q)Q + P'(Q) < 0$, $P'(Q) < 0$, and $q_i \in [0, Q_{-i}+q_i]$.

The assumptions are also known to imply that in the Cournot game without capacity constraints, there exists a unique equilibrium. The per-firm output in the unconstrained Cournot equilibrium, denoted by q^*, satisfies

$$\frac{\partial \pi_i(q^*, (n-1)q^*)}{\partial q_i} = P'(nq^*)q^* + P(nq^*) - c'(q^*) = 0. \tag{2}$$

This is just the first-order condition of maximizing π_i in q_i given Q_{-i}, and using $Q_{-i} = (n-1)q^*$. It is easy to see that under our assumptions, there is a unique solution in q^*: Consider the left-hand side of (2). At $q^* = 0$, it is positive, while as $q^* \to \infty$, it becomes negative by assumption. Its derivative in q^* is $nP''(nq^*)q^* + P'(nq^*)(n+1) - c''(q^*) < 0$. Therefore, there exists a unique q^* at which it equals zero, and so (2) is satisfied.

We will use $r_i(Q_{-i})$ to refer to the best response of firm i to the total production of the other firms, Q_{-i}, when firm i does not face a binding capacity constraint. That is, $r_i(Q_{-i}) = \arg\max_{q_i} \pi_i(q_i, Q_{-i})$. Equivalently, $r_i(Q_{-i})$ can be characterized by the first-order condition of this maximization, that is,

$$\frac{\partial \pi_i(r_i(Q_{-i}), Q_{-i})}{\partial q_i} = P'(Q)r_i(Q_{-i}) + P(Q) - c'(r_i(Q_{-i})) \equiv 0, \tag{3}$$

where $Q = Q_{-i} + r_i(Q_{-i})$. By totally differentiating this identity with respect to Q_{-i} and rearranging, we find that

$$r_i'(Q_{-i}) = -\frac{\partial^2 \pi_i / \partial q_i \partial Q_{-i}}{\partial^2 \pi_i / \partial q_i^2} = -\frac{P''(Q)r_i(Q_{-i}) + P'(Q)}{P''(Q)r_i(Q_{-i}) + 2P'(Q) - c''(r_i(Q_{-i}))} \in (-1, 0).$$

The unconstrained Cournot equilibrium satisfies $q^* = r_i((n-1)q^*)$. To ease notation, we drop the reference to firm i's identity when referring to the best response function because the best-response functions are identical across the firms.

The are two other industry-structure benchmarks besides the unconstrained symmetric Cournot outcome (where the production is q^* per firm and nq^* total) that will come up later in the section. The *monopoly production* in the downstream market is denoted by Q^M, where $Q^M = \arg\max_Q P(Q)Q - c(Q)$, that is,

$$P'(Q^M)Q^M + P(Q^M) = c'(Q^M). \tag{4}$$

The other market structure that will turn out to be important for us is that of a perfectly coordinated *symmetric cartel*. By definition, the symmetric cartel output maximizes the firms' joint profits while each firm produces one-n^{th} of the total output. That is, the total output in the cartel, Q^C, maximizies $P(Q)Q - nc(Q/n)$, and so

$$P'(Q^C)Q^C + P(Q^C) = c'(Q^C/n). \tag{5}$$

Note that the monopoly output is *less* than the total production of the symmetric cartel, $Q^M < Q^C$, because the marginal cost function is strictly increasing.

2.2 The Second-Period Cournot Subgame

Let $\Pi_i(k_1, \ldots, k_n)$ denote the (indirect) profit of firm i in the capacity-constrained Cournot game given that the capacity allocation is $(k_1, \ldots, k_n)$. We need to know if Π_i is well-defined, that is, whether there is a unique capacity constrained Cournot equilibrium in the downstream market for any capacity allocation $(k_1, \ldots, k_n)$. Proposition 1 settles this issue.[12] In what follows, without loss of generality and purely for the ease of notation, we relabel the firms in increasing order of capacities, so that $k_1 \leq \ldots \leq k_n$ in any capacity allocation.

Proposition 1 *For any capacity allocation, there is a unique equilibrium in the capacity-constrained Cournot game. The equilibrium is* $q_i = k_i$ *for* $i = 1, \ldots, m$ *and* $q_i = q_m^U$ *for* $i = m+1, \ldots, n$ *for some* $m \in \{0, 1, \ldots, n\}$*, where* q_m^U *solves* $q_m^U = r\left(\sum_{j=1}^m k_j + (n-m-1)q_m^U\right)$.

Denote the capacity-constrained Cournot equilibrium given capacity allocation $(k_1, \ldots, k_n)$ by $(q_i^e(k_1, \ldots, k_n))_{i=1}^n$, and let the indirect profit function of firm i be

$$\Pi_i(k_1, \ldots, k_n) = P\left(\textstyle\sum_i q_i^e(k_1, \ldots, k_n)\right) q_i^e(k_1, \ldots, k_n) - c\left(q_i^e(k_1, \ldots, k_n)\right).$$

[12] Cave and Salant (1995) prove the existence and uniqueness of Cournot equilibrium with capacity constraints under *constant unit costs*.

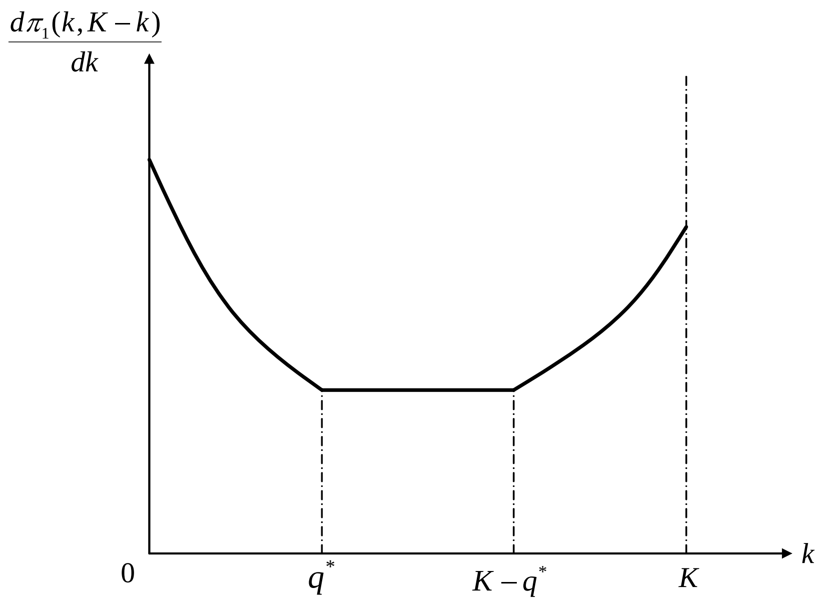

Figure 1: Firm i's marginal value of capacity when $n = 2$ and $K > 2q^*$.

An interesting feature of our model is that the buyers' (firms') marginal valuations for an additional unit of capacity *may not be monotonic* in the amount of capacity that they receive. This can be seen, at the level of intuition, for two firms as follows. When firm 1 is relatively small (has little capacity, which is a binding constraint in the downstream Cournot competition) then the marginal value of an additional unit of capacity is positive but decreasing because expanding the firm's production generates a positive yet decreasing marginal profit in the downstream market. However, if the firm is relatively large, so much so that its capacity constraint is slack while its opponent's constraint is binding in the downstream Cournot game, then the marginal value of additional capacity is increasing. This is so because by buying more capacity the firm tightens the other firm's capacity constraint, and the returns on this activity are increasing for our firm.[13] Therefore, the marginal value of capacity for firm i is U-shaped in the capacity of the firm, as shown in Figure 1.

2.3 Efficient Capacity Allocation

We assume that the first-period capacity allocation, $(k_1, \dots k_n)$ with $\sum_i k_i = K$, is *efficient for the industry* (or *efficient*, for short), that is, it maximizes $\sum_i \Pi_i(k_1, \dots, k_n)$.

[13] These verbal statements can be verified by direct calculation in the case of two firms.

This means that each unit of capacity is allocated to the firm that values it the most. This definition of efficiency ignores the consumer surplus in the downstream market. Therefore, the resulting allocation is not necessarily *socially efficient*. We also assume that the entire stock of capacity is allocated among the firms. This is a reasonable assumption in many of the applications discussed in the Introduction (e.g., airport landing slots, spectrum, etc.), and is further discussed in Section 6. As we will see in Section 3, restricting the sale of capacity may increase the input sellers' revenue, the downstream output, and social welfare.

Our motivation for studying capacity allocations that maximize the producers' surplus is not normative. Instead, the assumption reflects the fact that in practice, when inputs are allocated among producers (either via an auction or bargaining among the firms) the consumers of the final good are not present, and their interests are not represented.[14] Such "efficient" allocation *is* the outcome if the sale of the K units of capacity is organized in an *efficient auction* (where the participants are the firms that compete in the downstream industry), or if the allocation is determined via *Coasian bargaining* among the firms.

It turns out that our main results in Section 3 do not depend on the details of the mechanism yielding the efficient capacity allocation. The exact payment rule will be important only when we compare net profits (which include the cost of capacity) across firms in Section 4, so we defer discussion of the particular allocation mechanism until then. There, we will argue that in our setup, either a Vickrey-Clarke-Groves (VCG-) auction or a uniform-price share auction would induce a capacity allocation that is efficient for the industry, and we carry out the calculations with each payment rule separately. Coasian bargaining over inputs between firms would also yield the efficient allocation, but we do not consider it there because it does not yield tight predictions as to the prices that firms will pay for their capacity.

3 Main Results

We now turn to the analysis of the equilibrium market structure in the model of Section 2, and show that it may be qualitatively different depending on the amount

[14] This observation originates in Borenstein (1988). More recently, Hoppe, Jehiel and Moldovanu (2006) point out that the objective when designing the efficient auction of an input (e.g., licences) should be the weighted sum of the consumer and producer surpluses in the downstream market, and note the difficulty of incorporating the consumer surplus in the auction design as consumers do not submit bids.

of capacity sold in the auction. If the total capacity that is auctioned off is relatively little then the firms behave symmetrically; while if it is large then the only equilibrium is asymmetric, in which exactly one firm ends up with excess capacity and produces a larger quantity, while the other firms constrained by their insufficient capacities produce less. In Section 3.1 we show that the "regime change" (from a symmetric to an asymmetric outcome) happens at a certain capacity threshold and makes the total production drop discontinuously as the total capacity increases. In Section 3.2 we discuss how the market structure in our model differs from the unconstrained Cournot outcome, monopoly, and the structure of a perfectly coordinated cartel. In Section 3.3, we extend the comparison to the case of an infinite number of firms competing for inputs.

3.1 Industry Structure in the Downstream Market

Interestingly, the qualitative results regarding the downstream market structure depend on the total available capacity, K. If the total available capacity is relatively low then the efficient capacity allocation is symmetric, and all firms end up producing at their capacity constraints in the downstream market. However, if K is large—exceeding a threshold that is strictly lower than the total capacity needed to produce the symmetric, unconstrained Cournot output—then the industry structure becomes asymmetric. All but one firm gets the same, low capacity and operates at full capacity, while one firm gets the remaining capacity (which is a bigger share of the total than the share of any other firm), and operates strictly within its capacity constraint.

As a preparation for stating the results formally, we first describe the asymmetric industry structure that prevails when K is sufficiently large. In the efficient capacity auction, each "small" firm buys a capacity $k_1 = \ldots = k_{n-1} = k^*$, while the "large" firm receives the rest, $k_n = K - (n-1)k^*$. The capacity level of each of the small firms, k^*, maximizes

$$P\left((n-1)k + r((n-1)k)\right)\left[(n-1)k + r((n-1)k)\right] - (n-1)c(k) - c(r((n-1)k)). \quad (6)$$

This is the total industry profit given that $(n-1)$ firms produce k and one firm produces the unconstrained best reply, $r((n-1)k)$. The optimal level of k^* is characterized by the first-order condition of the maximization,

$$\left[P'(Q^*)Q^* + P(Q^*)\right]\left[1 + r'((n-1)k^*)\right] = c'(k^*) + r'((n-1)k^*)c'(r((n-1)k^*)), \quad (7)$$

where $Q^* = (n-1)k^* + r((n-1)k^*)$. Note that k^* does not vary with K.

The following lemma states that if *at least one firm is allocated excess capacity* in the efficient auction then the capacity allocation must be the asymmetric one described above. We will later see that a sufficient condition for there being at least one firm with slack capacity is that the total capacity exceed the amount needed for producing the unconstrained Cournot equilibrium outcome, i.e., $K > nq^*$. In the statement and proof of the lemma recall that the firms' indices are ordered so that $k_1 \leq k_2 \leq \ldots \leq k_n$.

Lemma 1 *Suppose that $(k_1, \ldots, k_n)$ is the equilibrium capacity allocation in our game. If at least one firm's capacity constraint is slack, i.e., $k_n > q_n^e(k_1, \ldots k_n)$, then $k_1 = \ldots = k_{n-1} = k^*$ and $k_n = K - (n-1)k^*$.*

The result of Lemma 1 is remarkable because it pins down the industry structure in our model whenever there is any slack capacity in the downstream market. Note that the asymmetric allocation of capacities ($k_1 = \ldots k_{n-1} = k^*$, $k_n = K - (n-1)k^*$) and the corresponding asymmetric production ($q_1^e = \ldots = q_{n-1}^e = k^*$, $q_n^e = r((n-1)k^*)$) do not depend on the total amount of available capacity, K. Observe that in this outcome, the small constrained firms indeed produce less than the unconstrained big firm as $k^* < r((n-1)k^*)$.

The only possibility that we have not considered is that *no firm has slack capacity* in the Cournot game that follows the efficient capacity auction. In this case, since the production technologies are symmetric and exhibit stritly decreasing returns, the efficient capacity allocation must be symmetric. (By distributing the total capacity K, we essentially distribute a fixed total production among the firms because all capacity is fully used. The most efficient way to produce a fixed quantity is by spreading it evenly across the firms.)

In summary, the efficient capacity allocation is either asymmetric with exactly one firm receiving excess capacity and the others all receiving k^*, or symmetric with all capacities binding in the downstream market. Note that the asymmetric outcome can arise as the solution only when it is feasible, that is, when $K \geq Q^* \equiv (n-1)k^* + r((n-1)k^*)$. If $K < Q^*$ then we know the efficient capacity allocation is symmetric, $k_i = K/n$ for all i.

The following proposition states the main result of this subsection: There exists a threshold level of total capacity, $\hat{K}$, such that the efficient capacity allocation is symmetric for $K < \hat{K}$ and asymmetric for $K > \hat{K}$. The threshold $\hat{K}$ falls strictly in between Q^* and nq^*.

Proposition 2 *Define $\hat{K}$ such that*

$$P(\hat{K})\hat{K} - nc(\hat{K}/n) = P(Q^*)Q^* - (n-1)c(k^*) - c(r((n-1)k^*)), \tag{8}$$

where k^* satisfies (7) with $Q^* = (n-1)k^* + r((n-1)k^*)$.

(a) If $K < \hat{K}$ then the efficient capacity allocation is symmetric, that is, each firm receives capacity K/n.

(b) If $K > \hat{K}$ then the efficient capacity allocation is such that all but one firm gets capacity $k^* < q^*$, while exactly one firm gets capacity $K - (n-1)k^*$.

An intriguing consequence of Proposition 2 is that the *capacities* of the firms and the *total output* produced in the downstream market *change discontinuously* as a function of the total available capacity at $K = \hat{K}$. In particular, the capacities of the small firms and the total output fall by discrete amounts at $K = \hat{K}$. This is so because in the asymmetric solution (which is valid for all $K \geq \hat{K}$) the small firms' capacities are k^* each and the total production is Q^*, while in the symmetric solution at $K = \hat{K}$, each firm has capacity $\hat{K}/n$ and the total production is $\hat{K}$. However, we know that $Q^* < \hat{K}$, hence the capacities and the output jump at $K = \hat{K}$. We depict the capacity allocation and the resulting total industry production as a function of K in Figure 2.

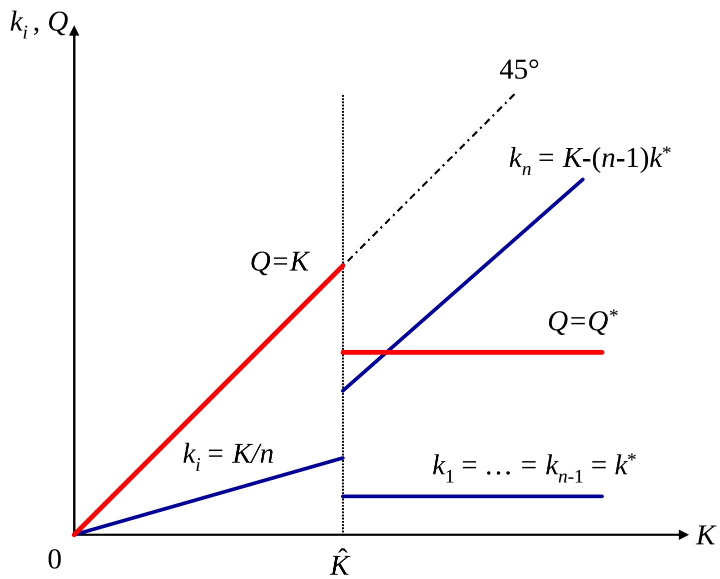

Figure 2: Capacities and total downstream production as a function of K

The most important prediction that follows from Proposition 2 is that the *total surplus* (social welfare) as a function of the available capacity is *maximized at* $K = \hat{K}$. The total surplus is just the sum of the firms' profits and the consumer surplus in the downstream market (payments to the auctioneer cancel). The total surplus is continuous and strictly increasing for $K < \hat{K}$ because K is allocated symmetrically (which is socially desirable), all capacity is fully used in production, and the total production is lower than the Cournot output ($K < \hat{K} < nq^*$). However, the total surplus *falls discretely* as K exceeds $\hat{K}$. This is because the firms' total profit is continuous at $K = \hat{K}$ by equation (8), but the consumer surplus falls discontinuously together with the total output in the downstream market. (The total surplus stays constant for all $K > \hat{K}$.) The policy consequence is that a social planner should restrict the quantity sold in the capacity auction to $\hat{K}$ whenever K exceeds $\hat{K}$.

3.2 Comparison with Benchmarks

Now we turn to the comparison of the industry structure in our model to certain benchmarks: (i) the symmetric unconstrained Cournot outcome, (ii) the monopoly, and (iii) the perfectly coordinated collusive cartel. We also discuss the limiting case of the model as the cost function becomes affine (constant returns to scale).

First, by symmetry, $r' \in (-1, 0)$, and the fact that firms 1 through $n-1$ are capacity constrained while firm n is not, it follows that

$$q_1^e = k^* < q^* < r((n-1)k^*) = q_n^e.$$

The small firms each produce less than the per-firm Cournot output while the unconstrained firm produces more than that.

Second, we claim that $k^* > 0$, that is, the outcome of our model always differs from that of a monopoly. To see this, differentiate (6), the total industry profit in the asymmetric solution, in k to get

$$(n-1)\left\{(1 + r'((n-1)k))\left[P'(Q)\,Q + P(Q)\right] - c'(k) - r'((n-1)k)c'\left(r((n-1)k)\right)\right\},$$

where $Q = (n-1)k + r((n-1)k$. For $k = 0$ and $n < \infty$, this simplifies to

$$\begin{aligned}(n-1)\left\{\left[P'(Q^M)Q^M + P(Q^M)\right](1 + r'(0)) - c'(0) - r'(0)c'(Q^M)\right\}\\ = (n-1)\left[c'(Q^M) - c'(0)\right],\end{aligned}$$

where $Q^M \equiv r(0)$ is the monopoly output, and on the second line the first-order condition of profit maximization by a monopoly, equation (4), is used. Since $c'(Q^M) > c'(0)$ by the strict convexity of c, the total industry profit is strictly increasing at $k = 0$. Hence $k^* > 0$.

Third, the outcome of our model is different from that of a collusive cartel unless $K = Q^C$ where Q^C is given by equation (5). This is so because our outcome is asymmetric for $K > \hat{K}$ (hence it cannot coincide with that of the cartel, which is symmetric), and it is symmetric but the total downstream production equals K (not Q^C) for $K < \hat{K}$.

Finally, it may be instructive to consider a limiting case of our model, when the production technology exhibits constant returns, that is, c is affine. While this case is ruled out by our assumption that c is strictly convex—which has been used in the proof of Lemma 1, for example—it is easy to check that Proposition 1 goes through with constant marginal costs as well. Therefore, for any capacity allocation, there exists a unique equilibrium in the follow-up Cournot game. It is interesting to note, by comparing equations (4) and (5), that when c' is constant, the monopoly and cartel outputs are equal, $Q^M = Q^C$.

When c is affine, the efficient allocation of capacities depends on the total available capacity as follows. If K is less than Q^M then *all capacity allocations* are efficient. To see this, note that firm i's unconstrained best response to the other firms' joint production is at least $Q^M - Q_{-i}$ (this is so because the best response would be Q^M for $Q_{-i} = 0$ and $r' > -1$). Since $Q^M > K$ and $Q_{-i} \leq \sum_{j \neq i} k_j$, the unconstrained best response is not feasible: $Q^M - Q_{-i} > K - \sum_{j \neq i} k_j \equiv k_i$. Hence firm i maximizes its profit by producing k_i. Since each firm operates at full capacity, the total industry output and profits are the same no matter how the capacities are allocated. Therefore, all allocations are equally efficient. On the other hand, if K is at least as large as Q^M then the efficient capacity allocation is such that *one firm gets all the capacity.* This follows because for any initial allocation of capacities, the production that maximizes the firms' joint profits is Q^M. However, if more than one firm is allocated a positive capacity then the joint production in the Cournot game exceeds Q^M. The cartel's profit is maximized by shutting down all firms but one. This is the outcome for $K > \hat{K} = Q^M$, therefore, the outcome of our model is monopoly, which can be interpreted as perfect cartelization among the firms by appropriate allocation of inputs. This contrasts with the case when marginal costs are increasing, where firms cannot achieve perfect cartelization through input allocation, but must trade-off productive efficiency against restraining output.

3.3 Market Structure with an Infinite Number of Firms

Our preceding analysis of the market structure is valid for any finite number of firms. In this subsection, we investigate what happens to the market structure as the *number of firms becomes infinitely large*. In particular, we are interested in knowing whether the market structure of our model collapses into monopoly, perfect cartelization, or perhaps perfect competition, in the limit as $n \to \infty$.

If the marginal cost is constant between zero and Q^M, then obviously, for all finite n and in the limit as $n \to \infty$, the outcome of our model is monopoly, which can be interpreted as a perfectly coordinated cartel. Therefore, in what follows, we again do not consider this special limiting case of the model.

In the analysis of the prevailing market structure with an infinite number of firms we will assume that an infinite amount of good can only be sold at zero price, and that the marginal cost of producing the first unit is positive. These two assumptions ensure that as $n \to \infty$, the unconstrained Cournot equilibrium converges to "perfect competition" in the sense that the per-firm production converges to zero, and the total output converges to a quantity where the market's willingness to pay equals the marginal cost of any single infinitesimal firm. To see this, recall that the per-firm output in the unconstrained Cournot equilibrium satisfies $P'(nq^*)q^* + P(nq^*) = c'(q^*)$. As $n \to \infty$, q^* has to go to zero, otherwise $\lim_{n\to\infty} P(nq^*) = 0$ and $\lim_{n\to\infty} P'(nq^*)q^* \leq 0 < \lim_{n\to\infty} c'(q^*)$ yield a contradiction. If $q^* \to 0$ then $\lim_{n\to\infty} P(nq^*) = c'(0)$. We will continue to assume that there is sufficient total capacity to produce the unconstrained Cournot output, that is, $\lim_{n\to\infty} nq^* < K$.

Proposition 3 *Suppose* $\lim_{Q\to\infty} P(Q) = 0$, $0 < c'(0) < c'(Q^M)$ *and* $\lim_{n\to\infty} nq^* < K$. *In our model, as* $n \to \infty$, k^* *converges to zero, however,* $(n-1)k^*$ *tends to a positive number which is less than the limit of the total industry production. The market structure remains different from monopoly, unconstrained Cournot competition, and perfect collusion even as* $n \to \infty$.

4 Tobin's Q, Firm Size, and Demand Cycles

In this section we derive a testable prediction on the relationship between firm size and Tobin's Q: we show that the two are *negatively related*. Tobin's Q is defined as the ratio of the firm's market value to its book value; in our model, it equals the firm's downstream profit divided by the cost of capacity. In order to compute this ratio—in particular, the cost of capacity—we exhibit in Section 4.1 two auction mechanisms that result in a capacity allocation that is efficient for the industry. These auctions are the Vickrey-Clarke-Groves mechanism and the uniform-price share auction. In Section 4.2, we show that the payment schemes corresponding to these auctions (the Vickrey payments and the equilibrium price in the uniform-price share auction, respectively) indeed imply that Tobin's Q and firm size are negatively related. Then, in Section 4.3, we investigate the comparative statics of our model with respect to demand fluctuations. In particular, we show that a small slump in demand may lead to a relatively large drop in the downstream production, and that the industry becomes more asymmetrical and concentrated during a contraction than it is during a demand-driven expansion.

4.1 Payment Rules in the Capacity Market

In the context of our model, a *Vickrey-Clarke-Groves (VCG) auction* with bids that are contingent on the entire allocation constitutes an efficient auction. This mechanism works as follows. Participants are requested to submit their monetary valuations for all possible allocations of the goods. The auctioneer chooses the allocation that maximizes the sum of the buyers' reported valuations. Then, each buyer pays the difference between the other buyers' total valuation in the hypothetical case that the goods were allocated efficiently among them (excluding him) and in the allocation actually selected by the auctioneer. The rules induce all participants to submit their valuations for every allocation honestly, and the outcome of the auction is efficient.[15]

For future reference, we introduce notation for the capacity allocation and the payments in the VCG auction. Suppose that the valuation submitted for allocation $(k_1, \ldots, k_n)$ by firm i is $b_i(k_1, \ldots, k_n)$. In the VCG auction, the auctioneer determines the allocation maximizes $\sum_{i=1}^{n} b_i(k_1, \ldots, k_n)$. Denote this allocation by $(k_1^*, \ldots, k_n^*)$. The price paid by firm i, also called the *Vickrey payment*, is calculated as

$$\max\left\{ \sum_{j\neq i} b_j(k_1, \ldots, k_n) \middle| \sum_{j\neq i} k_j = K \right\} - \sum_{j\neq i} b_j(k_1^*, \ldots, k_n^*). \tag{9}$$

It is routine to check that these rules induce firm i to bid $b_i(k_1, \ldots, k_n) = \Pi_i(k_1, \ldots, k_n)$, i.e., all firms bid honestly. Each firm that gets something in the efficient capacity allocation pays a positive price. Finally, each firm obtains a non-negative payoff from participation. These results are established in Krishna (2002), Chapter 5.3.

[15] See Krishna (2002), Chapter 5.3 for a more complete discussion.

There are other auction forms—simpler and more widely used in practice—that also yield an efficient capacity allocation in the context of our model. In particular, the *uniform-price share auction* (first analyzed by Wilson (1979)) is one such mechanism. In the uniform-price share auction for K units of capacity each firm i is required to submit an inverse demand schedule, $p_i(k_i)$, $k_i \in [0, K]$, which specifies the highest unit price firm i is willing to pay in exchange for k_i units of capacity. The auctioneer aggregates the demands and computes a market clearing price. A price level, say p^*, is called market clearing if there exists a capacity vector $(k_1, \ldots, k_n)$ such that $\sum_i k_i = K$ and $p_i(k_i) = p^*$ for all i. Each firm i is then required to buy k_i units of capacity at unit price p^*.

Proposition 4 *There exists an equilibrium in the uniform-price share auction that implements the efficient capacity allocation.*

It is well known that the uniform-price share auction exhibits multiple equilibria, as far as the allocation of goods and the unit price are concerned (see Wilson (1979)). A straightforward argument as to why we would expect the capacity allocation $(k_1^*, \ldots, k_n^*)$ to emerge as the focal equilibrium is that this allocation is efficient, that is, it maximizes the firms' joint profits. As remarked above, Coasian bargaining between firms could also implement the efficient capacity allocation, but since it does not generate a clear prediction for the amount that firms will pay for their capacity, we do not consider it here.

4.2 Tobin's Q and Firm Size

We now turn to the relationship between Tobin's Q and firm size. If $K < \hat{K}$ then all firms are identical. Therefore, in this subsection we confine attention to the case where $K > \hat{K}$ and so the capacity allocation is asymmetric.

Tobin's Q is formally defined as the ratio of the firm's market value to its book value: $\tau_i = \Pi_i^*/B_i$, where Π_i^* is the firm's equilibrium profit (representing its market value) and B_i is the firm's total payment for capacity (representing its book value). The market value is given by $\Pi_i^* = \Pi_i(k_1, \ldots, k_n)$ where $k_1 = \ldots = k_{n-1} = k^*$, $k_n = K - (n-1)k^*$ and k^* solves (7). The book value B_i depends on the specific payment rule used in the capacity auction.

We now derive Tobin's Q for each one of the two auctions considered in Section 4.1. First, suppose that each firm is required to make the so-called Vickrey payments in the capacity auction, as defined in equation (9). Denote Γ_{-1}^* the total profit of a subset of $(n-1)$ firms when the total capacity K is allocated *only among them* (i.e., excluding one firm) in a VCG auction. After a relabeling of the firms from firm 1 to $(n-1)$, this allocation is $k_1 = \ldots k_{n-2} = k_{-1}^*$ and $k_{n-1} = K - (n-2)k_{-1}^*$ where k_{-1}^* solves

$$\left[P'(Q_{-1}^*)Q_{-1}^* + P(Q_{-1}^*)\right]\left[1 + r'((n-2)k_{-1}^*)\right] \\ = c'(k_{-1}^*) + r'((n-2)k_{-1}^*)c'(r((n-2)k_{-1}^*)),$$

with $Q_{-1}^* = (n-2)k_{-1}^* + r((n-2)k_{-1}^*)$. This equation is just (7) for $n-1$ firms instead of n. The Vickrey payment that a small firm makes in the VCG capacity auction (i.e., its book value) is

$$B_i^V = \Gamma_{-1}^* - (n-2)\Pi_1^* - \Pi_n^* \text{ for } i = 1, \ldots, n-1.$$

The corresponding payment for capacity that the large firm makes is

$$B_n^V = \Gamma_{-1}^* - (n-1)\Pi_1^*.$$

Second, suppose that capacity is allocated in the uniform-price share auction, and that the equilibrium price of a unit of capacity is $p^* > 0$. Then a small firm's book value is simply $B_i^U = p^*k^*$, $i = 1, \ldots, n-1$, while the large firm's book value is $B_n^U = p^*(K - (n-1)k^*)$.

Depending on the type of auction used for allocating capacities, Tobin's Q for firm i is either $\tau_i^V = \Pi_i^*/B_i^V$ (under Vickrey payments) or $\tau_i^U = \Pi_i/B_i^U$ (under the uniform-price share auction). The following proposition establishes Tobin's Q and firm size for both payment rules.

Proposition 5 *Suppose $K > \hat{K}$. In equilibrium, Tobin's Q satisfies $\tau_1^\alpha = \ldots = \tau_{n-1}^\alpha > \tau_n^\alpha$ for $\alpha \in \{V, U\}$. That is, under both Vickrey payments and uniform capacity prices, there is a negative relationship between firm size (as measured by either book or market value, capacity, output, or sales) and Tobin's Q.*

Our model can thus help to explain the 'size-discount puzzle'. While standard models predict that more efficient firms are larger, there is a negative relationship between Tobin's Q (as a measure of firm efficiency) and various measures of firm size (sales, book value) in the data. This empirical puzzle was first pointed out by Lang and Stulz (1994); see also Eeckhout and Jovanovic (2002).

4.3 Output and Market Structure Over the Business Cycle

In this subsection, we consider the comparative statics of our model with respect to the level of demand. We show that if K is just below the threshold level of capacity $\widehat{K}$, then a small slump in demand will be reinforced by a large contraction of output. Further, the change in output is asymmetric: all but one firm downsize, relinquishing their capacity to one large firm, which will then exhibit a low Tobin's Q. If on the other hand, K is just above $\widehat{K}$, then a small increase in demand will induce a large expansion of output. But, again, the expansion of output is asymmetric: the small firms will grow at the expense of the large firm. This result, stated formally in Proposition 6 below, has potentially important implications for the economic effects of business cycles as it shows that an "efficient" allocation of capacity will result in a magnification of the business cycle. It also implies that industrial concentration measures should tend to rise in recession periods.

Let $P(Q;\theta)$ denote inverse demand if the state of demand is given by $\theta \geq 0$. Conditional on θ, we make the same assumptions on the shape of inverse demand as in section 2 above. Further, we assume that an increase θ will be associated with (i) an increase in demand, $\partial P(Q;\theta)/\partial\theta > 0$ for all $Q > 0$, $\lim_{\theta\to 0} P(Q;\theta)Q/n < c'(Q/n)$ for $Q > 0$, and $\lim_{\theta\to\infty} P(Q;\theta) + Q\partial P(Q;\theta)/\partial Q > c'(Q/n)$ for $Q > 0$; and (ii) less price-elastic demand, $\partial^2 P(Q;\theta)/\partial Q\partial\theta \geq 0$ for all $Q > 0$. These assumptions subsume the special case where an increase in the level of demand means a replication of the population of consumers, leaving consumers' tastes and incomes unchanged, and so inverse demand can be written as $P(Q;\theta) \equiv \widetilde{P}(Q/\theta)$ and satisfies $\widetilde{P}'(\cdot) < 0$.

The following proposition shows that the equilibrium industry structure is more asymmetric in a demand slump ($\theta < \widehat{\theta}$) than during a boom ($\theta > \widehat{\theta}$).

Proposition 6 *There exists a threshold demand level $\widehat{\theta}(K)$ such that $\widehat{K}(\theta) < K$ if and only if $\theta < \widehat{\theta}(K)$ and $\widehat{K}(\theta) > K$ if and only if $\theta > \widehat{\theta}(K)$. That is, if demand is low, $\theta < \widehat{\theta}(K)$, the efficient auction induces the asymmetric capacity allocation $(k^*(\theta), ..., k^*(\theta), K - k^*(\theta))$, while if demand is high, the efficient auction induces the symmetric capacity allocation $(K/n, ..., K/n)$.*

The following corollary is an immediate implication of Proposition 6.

Corollary 1 *Consider two demand levels θ_0 and θ_1 such that $\theta_0 < \widehat{\theta}(K) < \theta_1$ and $\theta_1 - \theta_0$ is arbitrarily small. Then, an increase in the demand parameter from θ_0 to θ_1 leads to a discrete increase in industry output as the capacity allocation switches from being asymmetric to being symmetric.*

In other words, around the capacity threshold, a small change in demand leads to a disproportionate change in output. This result may be of interest to macro-economists, since it suggests that cyclical changes in demand will be exaggerated in industries where capacity is scarce and is allocated efficiently. This may be surprising because one might think that industries where the input is scarce and most firms are capacity-constrained would be industries where cyclical changes in demand would tend to be dampened rather than amplified, but in fact we show that the reverse is true. It is also possible to study the cyclical behavior of marginal costs and mark-ups as demand fluctuates in our model (see Rotemberg and Woodford (1999) for a survey of the evidence and macro-economic literature on this topic). Interestingly, contrary to the usual supposition in macro-economics, average mark-ups do not have not be counter-cyclical for the business cycle to be amplified in our set-up, because we do not have a representative firm model. Indeed, examples can be constructed where the average mark-up rises discretely as the demand parameter θ crosses the threshold $\widehat{\theta}(K)$ from below. Despite this discrete increase in mark-ups, output also jumps up at this point because the change in industry structure results in a discrete increase in productive efficiency at the same time.

5 Differentiated Bertrand Competition

In this section we show that our main results extend to the case where the downstream industry is modeled as differentiated-products Bertrand competition, and firms compete in strategic complements instead of substitutes. The purpose of analyzing this extension is to demonstrate that our results are not due to some particular property of the Cournot model.

Assume that there are two firms that simultaneously set prices, denoted by p_i $(i = 1, 2)$. The demands for their goods are $q_1 = Q(p_1, p_2)$ and $q_2 = Q(p_2, p_1)$, respectively, where Q is decreasing in its first and increasing in its second argument. The firms have capacity constraints k_i $(i = 1, 2)$, which are determined in the first stage of the game.

We model differentiated Bertrand competition subject to capacity constraints as in Maggi (1996). If firm i faces a demand $q_i \leq k_i$ then its cost is $c(q_i)$; if $q_i > k_i$ then its cost is $c(q_i) + \theta(q_i - k_i)$, where c is a strictly increasing function and θ is a large positive number. Verbally, this means that the firms (constrained or not) always serve the entire demand they face; however, producing beyond their respective capacity constraints carries a drastic monetary penalty. This assumption allows us to ignore the issue of *rationing* when demand exceeds capacity.[16] Therefore, we can focus on the main qualitative difference between Bertrand and Cournot models: strategic complements vs. strategic substitutes.

We assume that for all capacity allocations (k_1, k_2) with $k_1 + k_2 = K$, there exist prices (p_1, p_2) such that $Q(p_1, p_2) = k_1$ and $Q(p_2, p_1) = k_2$. In order to ensure that the price vector that gives rise to demands that equal the capacities is *uniqe*, we assume that for all (p_1, p_2), $Q_1(p_1, p_2) + Q_2(p_1, p_2) < 0$, where Q_i denotes $\partial Q/\partial p_i$ for $i = 1, 2$. As a result of this assumption, firm 1's "iso-demand curve," $p_2(p_1)$, defined implicitly by $Q(p_1, p_2(p_1)) \equiv k_1$, has a slope greater than one: $p_2' = -Q_1/Q_2 > 1$. Therefore, the iso-demand curves intersect only once, hence the point (p_1, p_2) where $Q(p_1, p_2) = k_1$ and $Q(p_2, p_1) = k_2$ is unique.

Firm 1's profit when its capacity constraint is slack is $\pi(p_1, p_2) = p_1 Q(p_1, p_2) - c(Q(p_1, p_2))$. Assume that π is strictly concave in p_1, and define firm 1's unconstrained reaction function as $r(p_2) = \arg\max_{p_1} \pi(p_1, p_2)$. By symmetry, r is the unconstrained reaction function of firm 2 as well. Assume that r is differentiable with $r' \in (0, 1)$, which implies that there exists a unique equilibrium without capacity constraints, where both firms set $p^B = r(p^B)$. These assumptions could be expressed in terms of the true fundamentals (the functions Q and c), but, in the interest of brevity, we keep them in this form.[17] Denote the per-firm equilibrium output in the unconstrained differentiated Bertrand model by $q^B = Q(p^B, p^B)$.

If $Q(r(p_2), p_2) > k_1$, that is, firm 1's best response to firm 2's price yields a demand for firm 1's good that exceeds its capacity, then by the concavity of $\pi(p_1, p_2)$, the optimal (constrained) response for firm 1 is to set $p_1 > r(p_2)$ such that $Q(p_1, p_2) = k_1$. By symmetry, the same is true for firm 2: In case its unconstrained best response is not feasible, $Q(r(p_1), p_1) > k_2$, then its constrained best response to p_1 is $p_2 > r(p_1)$ such that $Q(p_2, p_1) = k_2$.

Our first result is that for all initial capacity allocations, there is an equilibrium in the ensuing differentiated Bertrand model with capacity constraints.

Lemma 2 *For all k_1, k_2 with $0 < k_1 \leq k_2$ and $k_1 + k_2 = K$, there exists an equilibrium in the capacity-constrained Bertrand game.*

The next issue is to determine the capacity allocation that maximizes the sum of the firms' profits subject to the constraint that for any initial capacity allocation

[16] Rationing naturally does not arise in the Cournot model with capacity constraints. Using this model, we essentially assume it away in the Bertrand model.

[17] The reader may consult chapter 6.2 of Vives (1999) for details.

(k_1, k_2) the equilibrium described in the previous lemma is played.

If the capacity allocation leads to a downstream equilibrium in which both firms are constrained then their joint profit is

$$p_1^* Q(p_1^*, p_2^*) - c\left(Q(p_1^*, p_2^*)\right) + p_2^* Q(p_2^*, p_1^*) - c\left(Q(p_2^*, p_1^*)\right),$$

where (p_1^*, p_2^*) is such that $k_1 = Q(p_1^*, p_2^*)$ and $k_2 = Q(p_2^*, p_1^*)$, as in Case 2 of the lemma. Change variables so that $p_1^* = P_1(k_1, k_2)$ and $p_2^* = P_2(k_1, k_2)$ and rewrite the joint profit as

$$P_1(k_1, k_2)k_1 - c(k_1) + P_2(k_1, k_2)k_2 - c(k_2). \tag{10}$$

We will assume that this expression is maximized in k_1 and $k_2 \equiv K - k_1$ at $k_1 = k_2 = K/2$. While (10) is symmetric in k_1 and k_2, this amounts to an additional (though mild) assumption. The assumption is made in the spirit of the original (Cournot) model, where the firms' joint profit maximizing quantity choice is symmetric as well.

Let K^C denote the joint production of a "cartel," that is, the value of K that maximizes $\left[P_1(K/2, K/2) + P_2(K/2, K/2)\right] K/2 - 2c(K/2)$. By definition (and the assumption in the previous paragraph), if the total capacity is K^C, then the optimal capacity allocation is $k_1 = k_2 = K^C/2$.

The symmetric allocation can be optimal only for K not exceeding the joint production in the unconstrained Bertrand equilibrium, $2q^B$. We now argue that even at $K = 2q^B$ it is strictly better for the firms to allocate the total capacity asymmetrically, so that the smaller firm (denoted by firm 1) becomes capacity constrained while the other firm becomes unconstrained in the ensuing equilibrium.

Suppose towards contradiction that each firm has capacity q^B and plays the unconstrained equilibrium by setting price p^B. Recall that $q^B = Q(p^B, r(p^B))$. Now reduce k_1 and increase k_2 by the same infinitesimal amount, $dk = [Q_1(p^B, p^B) + Q_2(p^B, p^B)r'(p^B)]dp$. By construction, firm 1 remains exactly capacity constrained if it increases its price by dp and firm 2 increases it by $r'(p^B)dp$. On the other hand, the same change in prices makes firm 2 unconstrained because the total demand decreases (as both prices go up) while the total capacity remains the same. Therefore, the resulting prices, $p^B + dp$ and $r(p^B + dp)$, form an equilibrium where firm 1 is constrained and firm 2 is unconstrained. We just need to show that the joint profit is higher in the new equilibrium. The change in the joint profit can be written as

$$\left.\frac{d\left[\pi(p_1, r(p_1)) + \pi(r(p_1), p_1)\right]}{dp_1}\right|_{p_1 = p^B} = \left[\pi_1(p^B, p^B) + \pi_2(p^B, p^B)\right]\left[1 + r'(p^B)\right],$$

where π_j denotes the derivative with respect to the jth argument. But this expression is positive because $\pi_1(p^B, p^B) = 0$ by the equilibrium condition, while $\pi_2(p^B, p^B) > 0$ and $r' > 0$.

We have so far established that for a total capacity level of $K = K^C$ the optimal capacity allocation is symmetric, while for $K = 2q^B$, the optimal allocation is asymmetric. By continuity (i.e., since the problem of optimal capacity allocation is continuous in K), there must exist an intermediate value of K, call it $\hat{K}$, where the optimal capacity allocation changes from symmetric to asymmetric. At such K, the joint profit of the firms is the same from splitting $\tilde{K}$ equally and allocating it optimally in an asymmetric fashion. If the production technology exhibits strictly decreasing returns (i.e., c is strictly convex) then there is a discrete drop in the social surplus as K increases past $\tilde{K}$. This is the exact same phenomenon that we found in the Cournot model. We summarize our findings regarding the differentiated Bertrand model in the following proposition.

Proposition 7 *In the differentiated Bertrand model, for some (low) values of K the efficient capacity allocation is symmetric, while for some other (high) values of K it is asymmetric. There exists a threshold value $\tilde{K} \in (K^C, 2q^B)$ where the efficient capacity allocation changes from symmetric to asymmetric. Around $\tilde{K}$, a small increase in the total available capacity reduces the social surplus.*

6 Discussion and Extensions

In this section we discuss the robustness of our results to various alternative specifications of the model.

Alternative models of capacity. In this paper, we have modeled the scarce input for which firms compete as capacity. Each firm's production is determined by a generalized "Leontief-type" technology with decreasing returns to scale and "capacity" as an essential and constraining input. It is an interesting question whether our results extend to competition for other types of potentially scarce inputs.

We have investigated the robustness of our results to two different ways of modeling inputs. First, one could imagine that a firm's capacity does not provide a strict upper bound on its production, but instead exceeding capacity simply increases its marginal cost of production. For concreteness, suppose that each firm's marginal cost is zero up to its capacity, and linearly increasing with output beyond that point (with no jumps).[18] In this variant, direct calculations show that there exists a capacity threshold $\hat{K}$ below which the equilibrium allocation is symmetric and all firms are "constrained" (produce beyond their capacities), while for a total capacity above $\hat{K}$ the allocation is asymmetric with one large unconstrained firm and $(n-1)$ identical constrained firms. As in our original model, output and social welfare drop discontinuously as total capacity crosses the threshold.[19]

Second, we consider a Cobb-Douglas technology with decreasing returns to scale where one of the inputs is scarce and is allocated efficiently prior to the production stage. In contrast to our base model, this variant allows firms to substitute other inputs (with perfectly elastic supply) for the scarce resource. Despite this ability to substitute our main results continue to hold. When the available amount of the scarce input is small, it will be allocated symmetrically across firms, while when it is large, the allocation is asymmetric with one firm acquiring all of the scarce input. Again, output and social welfare drop discretely as input availability crosses the threshold between symmetric and asymmetric allocations.[20] Intuitively, when the input is in very limited supply, the most important consideration for the firms is to ensure productive efficiency; whereas when the input is relatively plentiful, it is more important for them to limit production by allocating it asymmetrically.

Endogenizing the input supply. In this paper we have taken the upstream supply of capacity as given. For certain inputs (e.g., airport landing slots, spectrum, forestry tracts, tradeable pollution rights) supply is indeed inelastic, and the seller almost invariably allocates the entire stock among downstream firms. Our assumption is therefore appropriate for these applications.

However, our model can also be thought of as the final subgame of a larger game where upstream firms produce the scarce input in the first stage. Suppose, for example, that competing upstream producers choose quantities of the input which are then pooled and allocated to downstream firms through one of the auction processes described in Section 4.1. Depending on the number of upstream firms and their cost structures, it is clear that the equilibrium capacity production can be at, above, or below the capacity threshold $\hat{K}$. Therefore, for the purpose of our analysis we remain agnostic as to the amount of available capacity, and simply consider all possible cases.

Alternative (dynamic) input allocation mechanisms. In light of the finding

[18]The kink in the marginal cost curve turns out not to be important because in equilibrium firms never operate at this point.

[19]The calculations described in this paragraph are available from the authors upon request.

[20]Numerical calculations supporting these claims are available from the authors upon request.

that the "efficient capacity auction" yields an asymmetric and socially undesirable outcome in the downstream market, it is important to know for policy purposes whether other auctions (which are not efficient from the perspective of the capacity buyers) would yield socially better outcomes.

Dynamic auctions, where each unit of capacity is auctioned off separately over time, may be good candidates for such mechanisms. Suppose, for example, that the total capacity to be sold is divided into small units. At each point in time, one "unit" of capacity is sold at a second-price auction, with no discounting between periods. At first glance, it may seem surprising that this mechanism does not yield the same "efficient" result as the ones considered in Section 4.1. In fact, if K is sufficiently large and there are *constant returns to scale*, then the dynamic auction proposed above is socially more desirable than those auctions.[21] The brief intuition for this result is the following. Under constant returns to scale, an "efficient" capacity auction would allocate all available capacity to one firm. In order to get the same result in the dynamic auction, one firm would have to outbid all the others for each capacity unit, and pay the marginal profit of the first capacity unit every time. Since the marginal profit of capacity is decreasing, this is unprofitable for the large firm and monopoly cannot be sustained. It is an open question whether a dynamic auction would do socially better than the mechanisms studied in Section 4.1 under decreasing returns to scale.

7 Conclusion

In this paper we have examined the behavior of an industry requiring a scarce input ("capacity") which is in fixed supply, when the input is allocated through an efficient auction or other equivalent process, such as Coasian bargaining. After the input is allocated, firms compete subject to the capacity constraints imposed by their prior purchases in a Cournot (or, in an extension, differentiated goods Bertrand) game.

We have shown that under these circumstances, firms with ex ante symmetric production technologies end up in either a symmetric or an asymmetric equilibrium, depending on whether the available amount of input is smaller or larger than a certain threshold, respectively. The asymmetric equilibrium features one large firm which hoards input, with all other firms relatively small and constrained by their input

[21] This is established in a related model by Krishna (1993). However, dynamic auctions of capacity can lead to entry deterrence in an oligopoly, see Dana and Spier (2006).

purchases: thus the input is allocated in a way that is productively inefficient. This implies that, around the capacity threshold, an *increase* in the amount of input available will tighten rather than ease the input constraints which most firms face, and will lead to a *drop* in total output.

The inuition behind these results is that when the input is extremely scarce, the firms' priority is efficient production. Instead, when the input is abundant, production efficiency is sacrificed in favor of lower production and higher prices, which are attained by a wasteful asymmetric input allocation. This intuition does not rely on the scarce resource being capacity. Indeed, as discussed in Section 6, our main results generalize to the case where there is substitutability between the scarce resource and other inputs.

We showed that our model yields testable implications on the cross-sectional relationship between firm size and profitability (Tobin's Q) which seem to be consistent with available evidence. We also explained how small changes in demand will be amplified into much larger changes in output in our model as firms switch from asymmetric to symmetric equilibria. Our model might thus form the basis of an interesting macro-economic model of business cycles.

Our model also has implications for micro-economic policy, since it suggests that allocating input through efficient auctions may be misguided when the bidders are competing firms (see also the references cited in footnotes 7 and 14). More suprisingly, it shows that trying to increase input availability can easily be a misguided policy measure in such markets— even though firms face binding capacity (or input) constraints, an increase in input availability will lead to a reduction in output if it leads to a change in industry structure. Rather than encourage entry into the upstream market, it might be preferable to change the method by which the input is allocated. A more in-depth analysis of these issues is beyond the scope of this paper, but constitutes an interesting avenue for future research.

8 Appendix: Omitted Proofs

Proof of Proposition 1. If the total industry production is Q and firm i's production is q_i, then firm i's marginal profit is

$$\left.\frac{\partial \pi_i(q_i, Q_{-i})}{\partial q_i}\right|_{Q_{-i}=Q-q_i} = P'(Q)q_i + P(Q) - c'(q_i). \tag{11}$$

This expression is strictly decreasing in q_i because $P' < 0$ and $c'' \geq 0$, and it becomes negative if q_i is sufficiently large. Therefore, in equilibrium, if the total production is Q and firm i's capacity constraint is slack, then firm i produces a quantity $q^U(Q)$ such that

$$q^U(Q) = \min\left\{q_i \geq 0 \mid P'(Q)q_i + P(Q) - c'(q_i) \leq 0\right\}. \tag{12}$$

If firm i's capacity constraint is less than $q^U(Q)$ then it produces k_i. Note that all firms whose capacity constraints are slack produce the same output, $q^U(Q)$.

The function $q^U(Q)$ is continuous, and by the Implicit Function Theorem its derivative is

$$\frac{dq^U(Q)}{dQ} = -\frac{P''(Q)q^U(Q) + P'(Q)}{P'(Q) - c''(q^U(Q))}.$$

If $q^U(Q) \leq Q$ then $P''(Q)q^U(Q) + P'(Q) < 0$ by assumption. This, combined with $P' < 0$ and $c'' \geq 0$, implies that $dq^U(Q)/dQ < 0$ whenever $q^U(Q) \leq Q$.

Define

$$h(Q) = \sum_{i=1}^{n} \min\left\{k_i, q^U(Q)\right\} - Q. \tag{13}$$

Clearly, $Q^* \in [0, K]$ and $h(Q^*) = 0$ if and only if Q^* is the total production in a capacity-constrained Cournot equilibrium.

We claim that there exists a unique $Q^* \in [0, K]$ that satisfies $h(Q^*) = 0$. To see this, first note that $q^U(0) > 0$ by equation (12), hence $h(0) > 0$ by equation (13). If $Q \geq K \equiv \sum_i k_i$ then equation (13) yields $h(Q) \leq 0$. Since $q^U(Q)$ is continuous, $h(Q)$ is continuous as well. Therefore, by the Intermediate Value Theorem, there exists $Q^* \in (0, K]$ such that $h(Q^*) = 0$. If $Q < K$ then, by (13), $h(Q) \leq 0$ implies that $q^U(Q) < k_i$ for some i, and therefore $q^U(Q) \leq Q$. As a result, $q^U(Q)$ is strictly decreasing, and so is $h(Q)$, for all $Q \in [Q^*, K)$. Since $h(Q^*) = 0$, we have $h(Q) < 0$ for all $Q \in (Q^*, K]$. Therefore, any $Q^* \in [0, K]$ such that $h(Q^*) = 0$ is unique. ■

Proof of Lemma 1. We will argue that if some firm or firms have excess capacity and $(k_1, \ldots, k_n)$ differs from the proposed asymmetric capacity allocation, then there exists some perturbation that increases the total industry profit thereby contradicting the efficiency of $(k_1, \ldots, k_n)$.

First, we show that under the hypothesis of the lemma, there is *at least one* firm whose capacity constraint is binding in the downstream market. Suppose towards contradiction that all firms are unconstrained. Then they each produce q^*, where $q^* < k_i$. Redistribute capacities so that for all $i < n$, $k_i = q^*$, and $k_n = K - (n-1)q^*$. This change does not affect the downstream equilibrium production of any firm. Then, carry out the following perturbation: Reduce the capacity of each firm except firm n by an infinitesimal amount, dq, and increase k_n by $(n-1)dq$. As a result, the total production changes: Firm n gains $dq_n = r'((n-1)q^*)(n-1)dq$, while the other firms lose a combined $dQ_{-n} = (n-1)dq$. Since $r' > -1$, the change in total production is negative, that is, $dq_n + dQ_{-n} < 0$. The change in the total industry profit is,

$$\begin{aligned} d\Pi = &\frac{\partial \pi_n(q^*, (n-1)q^*)}{\partial q_n} dq_n + \frac{\partial \pi_n(q^*, (n-1)q^*)}{\partial Q_{-n}} dQ_{-n} \\ &+ \sum_{i=1}^{n-1} \left[\frac{\partial \pi_i(q^*, (n-1)q^*)}{\partial q_i} dq + \frac{\partial \pi_i(q^*, (n-1)q^*)}{\partial Q_{-i}} \left(dq_n + \frac{n-2}{n-1} dQ_{-n} \right) \right]. \end{aligned}$$

q^* is the unconstrained Cournot equilibrium production, therefore

$$\frac{\partial \pi_i(q^*, (n-1)q^*)}{\partial q_i} = 0 \text{ for all } i = 1, \ldots, n.$$

By symmetry,

$$\frac{\partial \pi_i(q^*, (n-1)q^*)}{\partial Q_{-i}} = \frac{\partial \pi_j(q^*, (n-1)q^*)}{\partial Q_{-j}} \quad \text{for all } i, j = 1, \ldots, n.$$

Using these facts, the expression for $d\Pi$ simplifies to

$$\begin{aligned} d\Pi &= \frac{\partial \pi_n(q^*, (n-1)q^*)}{\partial Q_{-n}} dQ_{-n} + \sum_{i=1}^{n-1} \frac{\partial \pi_i(q^*, (n-1)q^*)}{\partial Q_{-i}} \left(dq_n + \frac{n-2}{n-1} dQ_{-n} \right) \\ &= \frac{\partial \pi_n(q^*, (n-1)q^*)}{\partial Q_{-n}} (n-1) \left(dQ_{-n} + dq_n \right). \end{aligned}$$

By $\partial \pi_n / \partial Q_{-n} < 0$ and $dq_n + dQ_{-n} < 0$, the change in total industry profit is positive, that is, $d\Pi > 0$. The perturbation of capacities increases the firms' total profit, hence the original distribution of capacities was not efficient, which is a contradiction.

For $n = 2$, the previous argument establishes that *exactly one firm* has excess capacity. We now prove that the same is true for $n > 2$ as well. Suppose towards contradiction that more than one firm has excess capacity, i.e., due to the way firms are indexed, $q^e_{n-1}(k_1, \ldots, k_n) < k_{n-1}$. Note that the capacity of firm 1 is binding, therefore $q^e_1(k_1, \ldots, k_n) = k_1 < q_{n-1}$. Redistribute all excess capacity from firms 2 through $n-1$ to firm n; this obviously does not change the production levels. Denote the new capacity levels by $(\tilde{k}_1, \ldots, \tilde{k}_n)$. Now decrease $\tilde{k}_{n-1} = q^e_{n-1}(k_1, \ldots, k_n)$ by dq and increase $\tilde{k}_1 = k_1$ by dq. Since firm 1's capacity is a binding constraint for its producton, q_1^e increases by dq as well. As a result, the total production of all firms is unchanged. However, as the cost functions are strictly convex and the distribution of production among the firms has become less asymmetrical (we have increased q_1^e, decreased q_{n-1}^e, and $q_1^e < q_{n-1}^e$ at the initial capacity levels), the total industry profit increases. The original allocation of capacities was not maximizing the total industry profit, which is a contradiction.

We conclude that if the capacity auction is efficient and there is a firm with excess capacity in the downstream market then it is firm n (i.e., there can only be one firm with slack capacity). Due to symmetry, the allocation of capacities that maximizes the total downstream industry profit subject to the constraint that firm n best-responds to the joint production of the other firms is the same for firms 1 through $n-1$, that is, $k_1 = \ldots = k_{n-1} = k^*$. The capacity-constrained firms each produce k^*, while the unconstrained firm produces $r((n-1)k^*)$. The optimal capacity constraint, k^*, maximizes the total industry profit, (6). ■

Proof of Proposition 2. We already know that the efficient capacity allocation is either symmetric, where $k_i = K/n$ for all i and all capacity constraints bind, or asymmetric as in Lemma 1, where $k_i = k^*$ for $i < n$ and $k_n = K - (n-1)k^*$. Note that the former allocation is the efficient one when the latter is not feasible, that is, $K \leq Q^*$.

Recall that we say that the capacity allocation is efficient when it maximizes the total industry profit in the capacity-constrained Cournot game. In the downstream market following the symmetric capacity allocation the total industry profit is $P(K)K - nc(K/n)$, which is strictly concave in K. Moreover,

$$P(Q^*)Q^* - nc(Q^*/n) > P(Q^*)Q^* - (n-1)c(k^*) - c(r((n-1)k^*)) \qquad (14)$$

because c is strictly convex and $k^* < Q^*/n < r((n-1)k^*)$. On the other hand, if the total capacity equals the total output in the unconstrained Cournot equilibrium, $K = nq^*$, then at least one firm must be unconstrained in any capacity allocation, hence by Lemma 1 the efficient allocation is the asymmetric one, and so

$$P(nq^*)nq^* - nc(q^*) < P(Q^*)Q^* - (n-1)c(k^*) - c(r((n-1)k^*)).$$

Therefore, there exists $\hat{K} \in (Q^*, nq^*)$ such that if $K > \hat{K}$, the asymmetric allocation is efficient, while if $K < \hat{K}$, the symmetric allocation is efficient. At $K = \hat{K}$, the two allocations generate the same industry profits, that is, $\hat{K}$ is defined by (8). ■

Proof of Proposition 3. Under our assumptions, the per-firm Cournot output converges to zero as the number of firms goes to infinity. Since k^* is less than q^* for any given n, it must also converge to zero.

We claim that $(n-1)k^*$ cannot converge to zero as $n \to \infty$. If it did then $\lim_{n\to\infty} r((n-1)k^*) = Q^M$. By equation (4),

$$\begin{aligned}\left[P'(Q^M)Q^M + P(Q^M)\right][1+r'(0)] &= c'(Q^M)[1+r'(0)] \\ &> c'(0) + r'(0)c'(Q^M),\end{aligned}$$

where $c'(0) < c'(Q^M)$ is used on the second line. The strict inequality contradicts (7), the first-order condition characterizing k^*, for n sufficiently large.

Finally, we claim that if the total industry production converges to $\bar{Q}^*$ as n goes to infinity then $\lim_{n\to\infty}(n-1)k^* < \bar{Q}^*$. In other words, the output of the unconstrained firm does not shrink to zero as the number of firms grows large. (Its output is greater than q^* for any finite n, but q^* goes to zero as n goes to infinity.) Suppose towards contradiction that $r(\bar{Q}^*) = 0$. By the definition of the best-response function, equation (3), $P(\bar{Q}^*) = c'(0)$. This contradicts the first-order condition that defines k^* for n sufficiently large, because as $n \to \infty$, by (7), $P'(\bar{Q}^*)\bar{Q}^* + P(\bar{Q}^*) = c'(0)$, and hence $P(\bar{Q}^*) > c'(0)$. ■

Proof of Proposition 4. Pick a positive p^* such that $p^*k_i^* < \Pi_i(k_1^*, \ldots, k_n^*)$ for all i. Recall that $k_1^* = \ldots = k_{n-1}^* = k^*$ and $k_n^* = K - (n-1)k^* > k^*$. We will define an equilibrium where, given the other $(n-1)$ firms' equilibrium strategies (inverse demand schedules), each firm is indifferent to use any strategy in response, therefore they each use their proposed equilibrium strategy. In this equilibrium, firms $i = 1, \ldots, n-1$ submit the same schedule, $p_1^*(\cdot)$, while firm n submits $p_n^*(\cdot)$, and the induced allocation of capacity is $(k_1^*, \ldots, k_n^*)$.

Denote $\Pi_i^* = \Pi_i(k_1^*, \ldots, k_n^*)$ for $i = 1, \ldots, n$. Let

$$p_1^*(k_1) = \frac{\Pi_1(k_1, \ldots, k_1, K-(n-1)k_1) - \Pi_n^* + p^*k_n^*}{K-(n-1)k_1}. \tag{15}$$

We claim that this inverse demand bid function makes firm n indifferent to submit any bid function. To see this, note that if firm n's bid results in it getting capacity k_n then the other firms each receive capacity $k_1 = (K-k_n)/(n-1)$, and the unit price of capacity becomes $p_1^*((K-k_n)/(n-1))$. Using (15), firm n's profit is

$$\Pi_1(k_1, \ldots, k_1, k_n) - p_1^*(k_1)\,k_n = \Pi_n^* - p^*k_n^*.$$

Hence firm n is indifferent between inducing any capacity k_n and k_n^*. Now we construct an inverse demand schedule for firm n that makes any other firm (say, firm $(n-1)$) indifferent to submitting any demand schedule (given that the other $(n-2)$ firms use p_1^*), and, together with p_1^* defined in (15), induces the allocation $(k_1^*, \ldots, k_n^*)$.

For all $k_1 \leq K/(n-2)$, define $k_n(k_1)$ as the lowest non-negative number such that

$$\Pi_{n-1}(k_1, \ldots, k_1, k_{n-1}, k_n) - p_1^*(k_1)k_{n-1} \leq \Pi_{n-1}^* - p^* k_{n-1}^*, \tag{16}$$

where $k_{n-1} \equiv K - k_n - (n-2)k_1$. Such $k_n(k_1)$ is well-defined because at $k_n = K - (n-2)k_1$, the left-hand side of (16) becomes zero, while the right-hand side is a positive constant, so (16) holds as a strict inequality. Note also that if $k_n(k_1)$ is positive then (16) holds as an equality. Now let $p_n^*(k_n(k_1)) \equiv p_1^*(k_1)$. Defining p_n^* this way guarantees that when firms $i = 1, \ldots, n-2$ submit p_1^* and firm n submits p_n^*, the best response of the remaining firm, firm $n-1$, is to submit p_1^* as well. This is so because by submitting an inverse demand schedule, firm $n-1$ can induce any capacity allocation $(k_1, \ldots, k_1, k_{n-1}, k_n)$ where $k_n = k_n(k_1)$ and the unit price of capacity is $p_1^*(k_1) \equiv p_n^*(k_n)$. In particular, if firm $n-1$ submits p_1^* then the induced allocation is $(k_1^*, \ldots, k_n^*) = (k^*, \ldots, k^*, K-(n-1)k^*)$ and the unit price is $p^* = p_1^*(k^*)$. By (16), the net profit of firm $n-1$ is maximized by inducing exactly this allocation. ■

Proof of Proposition 5. In equilibrium, firms 1 to $(n-1)$ are identical, therefore $\tau_1^\alpha = \ldots = \tau_{n-1}^\alpha$ for $\alpha \in \{V, U\}$.

(i) Suppose $\alpha = V$ (Vickrey auction). Now $\tau_1 > \tau_n$ is equivalent to

$$\frac{\Pi_1^*}{\Gamma_{-1}^* - (n-2)\Pi_1^* - \Pi_n^*} > \frac{\Pi_n^*}{\Gamma_{-1}^* - (n-1)\Pi_1^*}.$$

Cross-multiplying and rearranging yields, equivalently,

$$\Pi_n^{*2} + (n-2)\Pi_n^*\Pi_1^* - (n-1)\Pi_1^{*2} > \Pi_n^*\Gamma_{-1}^* - \Pi_1^*\Gamma_{-1}^*.$$

Factoring out $(\Pi_n^* - \Pi_1^*)$ yields

$$(\Pi_n^* - \Pi_1^*)\left[\Pi_n^* + (n-1)\Pi_1^*\right] > (\Pi_n^* - \Pi_1^*)\,\Gamma_{-1}^*.$$

Since $\Pi_n^* > \Pi_1^*$, this is equivalent to $\Pi_n^* + (n-1)\Pi_1^* > \Gamma_{-1}^*$, which holds because the VCG allocation is efficient for the firms.

(ii) Suppose $\alpha = U$ (uniform-price share auction). Note that $k^* < K - (n-1)k^*$.

Firms 1 to $n-1$ produce each an output of k^*, while the large firm n produces $r\left((n-1)k^*\right) \in (k^*, K-(n-1)k^*)$. Let $Q^* = (n-1)k^* + r((n-1)k^*)$ denote industry output. Then, $\Pi_1^* = P(Q^*)k^* - c(k^*)$, while $\Pi_n^* = P(Q^*)r((n-1)k^*) - c(r((n-1)k^*))$.

We need to show that $\tau_1 > \tau_n$, that is,

$$\frac{P(Q^*)k^* - c(k^*)}{p^* k^*} > \frac{P(Q^*)r((n-1)k^*) - c(r((n-1)k^*))}{p^*\left[K-(n-1)k^*\right]}.$$

Multiplying both sides by $p^* > 0$, we get

$$P(Q^*) - \frac{c(k^*)}{k^*} > P(Q^*)\frac{r((n-1)k^*)}{K-(n-1)k^*} - \frac{c(r((n-1)k^*))}{K-(n-1)k^*}.$$

This inequality indeed holds because $k^* < r((n-1)k^*) < K-(n-1)k^*$ and c is strictly convex. Hence, $\tau_1 > \tau_n$. ■

Proof of Proposition 6. Let

$$\varphi(K;\theta) \equiv P(K;\theta)K - nc(K/n) - \{P(Q^*;\theta)Q^* - (n-1)c(k^*) - c(r((n-1)k^*;\theta))\}, \quad (17)$$

and

$$\psi\left(q;\theta\right) \equiv P((n-1)k^* + q;\theta) + q\frac{\partial P((n-1)k^*+q;\theta)}{\partial Q} - c'(q), \quad (18)$$

where $r((n-1)k^*;\theta)$ is defined by the first-order condition $\psi\left(r((n-1)k^*;\theta);\theta\right) = 0$, $Q^* \equiv (n-1)k^* + r((n-1)k^*;\theta)$, and k^* (which depends on θ) maximizes the expression in curly brackets in equation (17). As we have shown before, the threshold capacity level $\widehat{K}$ is uniquely defined by $\varphi(\widehat{K};\theta) = 0$.

We first show that $d\widehat{K}/d\theta > 0$. Since $\partial\varphi(\widehat{K};\theta)/\partial K < 0$, it follows from the implicit function theorem that $d\widehat{K}/d\theta > 0$ if and only if $\partial\varphi(\widehat{K};\theta)/\partial\theta > 0$. Applying the envelope theorem (as k^* maximizes the expression in curly brackets above), we obtain

$$\begin{aligned}\frac{\partial\varphi(\widehat{K};\theta)}{\partial\theta} &= \widehat{K}\frac{\partial P(\widehat{K};\theta)}{\partial\theta} - Q^*\frac{\partial P(Q^*;\theta)}{\partial\theta} \\ &\quad - \left[P(Q^*;\theta) + Q^*\frac{\partial P(Q^*;\theta)}{\partial Q} - c'(r((n-1)k^*;\theta))\right]\frac{\partial r((n-1)k^*;\theta)}{\partial\theta}.\end{aligned}$$

From (18), the first-order condition $\psi\left(r((n-1)k^*;\theta);\theta\right) = 0$ and $Q^* > r((n-1)k^*;\theta)$ it follows that the expression in brackets is negative. Since $r((n-1)k^*;\theta)$ is the large firm's best response, we have $\partial\psi\left(r((n-1)k^*;\theta);\theta\right)/\partial q < 0$, and so (from the implicit function theorem), $\partial r((n-1)k^*;\theta)/\partial\theta > 0$ if and only if $\partial\psi\left(r((n-1)k^*;\theta);\theta\right)/\partial\theta >$ 0. Indeed,

$$\frac{\partial\psi\left(r((n-1)k^*;\theta);\theta\right)}{\partial\theta} = \frac{\partial P(Q^*;\theta)}{\partial\theta} + r((n-1)k^*;\theta)\frac{\partial^2 P((n-1)k^*+q;\theta)}{\partial Q\partial\theta} < 0.$$

Hence, $\partial r((n-1)k^*;\theta)/\partial\theta$. We now claim that $\widehat{K}\partial P(\widehat{K};\theta)/\partial\theta < Q^*\partial P(Q^*;\theta)/\partial\theta$. To see this, recall that $\widehat{K} < Q^*$. From our assumption on the cross-partial derivative of inverse demand, it then follows $0 < \partial P(\widehat{K};\theta)/\partial\theta < \partial P(Q^*;\theta)/\partial\theta$. Hence, $\partial\varphi(\widehat{K};\theta)/\partial\theta > 0$, and so $d\widehat{K}/d\theta > 0$.

We now show that $\widehat{K} \to 0$ as $\theta \to 0$. Our assumptions on inverse demand imply that for any fixed $K > 0$, $\lim_{\theta\to 0}\varphi(K;\theta) < 0$. The assertion then follows from the observation that $\varphi(K;\theta)$ is strictly concave in K. Next, we show that $\widehat{K} \to \infty$ as $\theta \to \infty$. To see this, note that $\varphi(K;\theta)$ is maximized at $K = Q^C$, the perfectly collusive cartel output, which is implicitly defined by

$$P(Q^C;\theta) + Q^C\frac{\partial P(Q^C;\theta)}{\partial Q} - c'(Q^C/n) = 0.$$

Observe that $Q^C \to \infty$ as $\theta \to \infty$. Otherwise, if Q^C were bounded from above, the l.h.s. of the above equation would become strictly positive for θ sufficiently large; a contradiction. Since $\widehat{K} > Q^C$, the assertion is indeed correct.

Summing up, we have shown that $\widehat{K}$ is strictly increasing with θ, $\widehat{K} \to 0$ as $\theta \to 0$, and $\widehat{K} \to \infty$ as $\theta \to \infty$. Hence, there exists a unique $\widehat{\theta}$ such that $K > \widehat{K}$ if and only if $\theta < \widehat{\theta}$ and $K < \widehat{K}$ if and only if $\theta < \widehat{\theta}$. ■

Proof of Lemma 2. If $k_1 \geq q^B$ then both firms are capable of producing the unconstrained Bertrand equilibrium output. It is immediate that both firms setting p^B forms an equilibrium.[22]

In the rest of the proof assume $k_1 < q^B$. Find p_1^0 such that $Q(p_1^0, r(p_1^0)) = k_1$. Note that $p_1^0 > p^B$ because $Q(p^B, r(p^B)) = q^B > k_1$ and $Q(p_1, r(p_1))$ is decreasing in p_1.[23] We distinguish two cases depending on whether or not k_2 exceeds $Q(r(p_1^0), p_1^0)$.

Case 1: $Q(r(p_1^0), p_1^0) \leq k_2$. We claim that $(p_1^0, r(p_1^0))$ is an equilibrium.

[22] The same prices form an equilibrium when the firms do not have capacity constraints. The only action that is not available to a firm without capacity constraint that is available to it with capacity constraint is decreasing its price so much that the capacity constraint becomes binding. However, such a move clearly cannot be profitable. Therefore there is no profitable deviation from equilibrium for either firm as long as their capacities exceed the equilibrium output without capacity constraints.

[23] This is so because $dQ(p_1, r(p_1))/dp_1 = Q_1 + Q_2 r' < Q_1 + Q_2 < 0$.

Firm 2 is best responding to firm 1's price without violating its capacity constraint, therefore it has no profitable deviation.

Firm 1's unconstrained best response to $r(p_1^0)$ would be $r(r(p_1^0))$. Since $p_1^0 > p^B$ and $r' \in (0,1)$, we have $p_1^0 > r(p_1^0) > p^B$, which then implies (by the same argument) that $r(p_1^0) > r(r(p_1^0)) > p^B$. But then $Q(r(r(p_1^0)), r(p_1^0)) > k_1$, that is, firm 1's best response to $r(p_1^0)$ violates its capacity constraint, because $Q(p_1^0, r(p_1^0)) = k_1$, $p_1^0 > r(r(p_1^0))$, and Q is decreasing in its first argument. Therefore firm 1's constrained best response to $r(p_1^0)$ is p_1^0, the price for which the capacity constraint holds as an equality.

Case 2: $Q(r(p_1^0), p_1^0) > k_2$. In this case, find (p_1^*, p_2^*) such that $Q(p_1^*, p_2^*) = k_1$ and $Q(p_2^*, p_1^*) = k_2$. We claim that (p_1^*, p_2^*) is an equilibrium.

First note that $p_1^0 < p_1^*$ and $p_2^* < p_1^*$. The first inequality holds because $Q(p_1^0, r(p_1^0)) = Q(p_1^*, p_2^*) = k_1$, $Q(r(p_1^0), p_1^0) > Q(p_2^*, p_1^*) = k_2$, and $Q_1 + Q_2 < 0$. Intuitively (graphically), we move along firm 1's iso-demand curve starting from $(p_1^0, r(p_1^0))$ in the direction where firm 2's demand decreases, so $p_1^* > p_1^0$ and $p_2^* > r(p_1^0)$. The second inequality follows because $k_1 < k_2$, and the firms are symmetric.

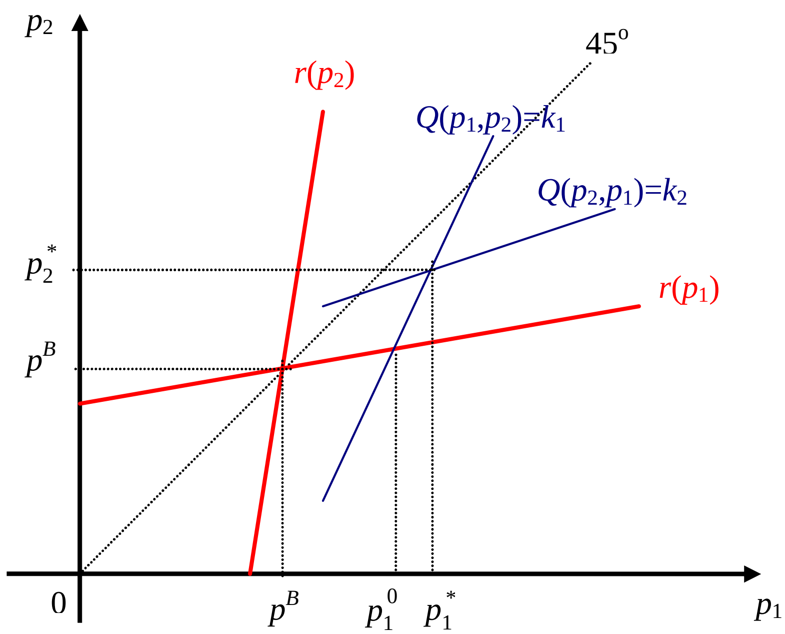

Figure 3: Illustration for Lemma 2

Now we verify that both firms play constrained best responses. As for firm 1,

$r(p_2^*) < p_1^*$ because $p_1^* > p^B$ and $p_1^* > p_2^*$. Therefore firm 1's unconstrained best response to p_2^* would violate its capacity constraint, hence its constrained best response is indeed p_1^*. As for firm 2, $r(p_1^*) < p_2^*$ as well; this is so because as we increase p_1 from p_1^0 to p_1^* while keeping $Q(p_1, p_2)$ constant (at k_1), the change in p_2 is greater than the increase in 2's best response. (Graphically, firm 1's iso-demand curve is steeper than firm 2's reaction curve. See the figure.) By $r(p_1^*) < p_2^*$, the unconstrained best response of firm 2 violates its capacity constraint, hence its constrained best response is p_2^*. ■

References

[1] d'Aspremont, Claude and Alexis Jacquemin (1988), "Cooperative and Noncooperative R&D in Duopoly with Spillovers," *American Economic Review*, vol. 78(5), 1133-1137

[2] Benoît, Jean-Pierre and Vijay Krishna (1987), "Dynamic Duopoly: Prices and Quantities," *Review of Economic Studies*, vol. 54(1), 23-35.

[3] Besanko, David and Ulrich Doraszelski (2004), "Capacity Dynamics and Endogenous Asymmetries in Firm Size," *Rand Journal of Economics*, vol. 35(1), 23-49.

[4] Borenstein, Severin (1988), "On the Efficiency of Competitive Markets for Operating Licences," *Quarterly Journal of Economics*, vol. 103 (2), 357-385.

[5] Borenstein, Severin (1989), "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry," *RAND Journal of Economics*, vol. 20(3), 344-365.

[6] Borenstein, Severin (1991), "The Dominant-Firm Advantage in Multiproduct Industries: Evidence from the U.S. Airlines," *Quarterly Journal of Economics*, vol. 106(4), 1237-1266.

[7] Cave, Jonathan and Stephen W. Salant (1995), "Cartel Quotas under Majority Rule," *American Economic Review*, vol. 85 (1), 82-102.

[8] Dana, James, and Kathryn Spier (2006), "Entry Deterrence in a Duopoly Market," forthcoming, *BE Press*.

[9] Davidson, Carl and Raymond Deneckere (1990), "Excess Capacity and Collusion," *International Economic Review*, vol. 31(3), 521-541.

[10] Eeckhout, Jan and Boyan Jovanovic (2002), "Knowledge Spillovers and Inequality", *American Economic Review*, vol. 92(5), 1290-1307.

[11] Ghemawat, Pankaj (1990), "The Snowball Effect," *International Journal of Industrial Economics*, vol. 8(), 335-351.

[12] Hoppe, Heidrun, Philippe Jehiel, and Benny Moldovanu (2006), "License Auctions and Market Structure", *Journal of Economics & Management Strategy*, vol. 15(2), 371-398.

[13] Jegadeesh, Narasimhan (1993) "Treasury Auction Bids and the Salomon Squeeze," *Journal of Finance*, vol. 48(4), 1403-1419.

[14] Jehiel, Philippe and Benny Moldovanu (2000), "Auctions with Downstream Interaction among Buyers", *RAND Journal of Economics*, vol. 31 (4), 768-791.

[15] Jehiel, Philippe, Benny Moldovanu, and Ennio Stacchetti (1996), "How (not) to Sell Nuclear Weapons," *American Economic Review*, vol. 86(4), 814-829.

[16] Katz, Michael L., and Carl Shapiro (1986), "How to Licence Intagible Property," *Quarterly Journal of Economics*, vol. 101(3), 567-590.

[17] Krishna, Kala (1993), "Auctions with Endogenous Valuations: The Persistence of Monopoly Revisited," *American Economic Review*, vol. 83(1), 147-160.

[18] Krishna, Vijay (2002), *Auction Theory*, San Diego: Academic Press

[19] Lang, Larry H. P. and René M. Stulz (1994), "Tobin's q, Corporate Diversification, and Firm Performance," *Journal of Political Economy*, vol. 102(6), 1248-1280.

[20] Maggi, Giovanni, "Strategic Trade Policies with Endogenous Mode of Competition," *American Economic Review,* vol. 86(1), 237-258.

[21] McAfee, R. Preston (1998), "Four Issues in Market Design," *Revista Analisis Economico,* vol. 13(1), 7-24.

[22] Monopolies and Mergers Commission (1990) "The Supply of Petrol: A report on the supply in the United Kingdom of petrol by wholesale".

[23] Nocke, Volker (2007), "Collusion and (Under-)Investment in Quality," *RAND Journal of Economics*, forthcoming.

[24] Rey, Patrick and Jean Tirole (2007) "A Primer on Foreclosure" in the *Handbook of Industrial Organization*, vol. III, Mark Armstrong and Robert Porter, eds., North Holland.

[25] Riordan, Michael (1998) "Anti-Competitive Vertical Integration by a Dominant Firm," *American Economic Review,* vol. 88 (5), 1232-1248.

[26] Rotemberg, Julio and Michael Woodford (1999) "The Cyclical Behavior of Prices and Costs" in the *Handbook of Macroeconomics*, John Taylor and Michael Woodford, eds., Amsterdam: North-Holland Elsevier.

[27] Saloner, Garth (1987), "Cournot Duopoly with Two Production Periods," *Journal of Economic Theory*, vol. 42(1), 183-187.

[28] Salinger, Michael A. (1988) "Vertical Mergers and Market Forclosure," *Quarterly Journal of Economics*, vol.103(2), 345-356.

[29] Spar, Debora L. (2006), "Markets: Continuity and Change in the International Diamond Market," *Journal of Economic Perspectives*, vol. 20(3), 195-208.

[30] Stahl, Dale O. (1988), "Bertrand Competition for Inputs and Walrasian Outcomes," *American Economic Review*, vol 78(1), 189-201.

[31] Vickrey, William (1961), "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance*, vol.16(1), 8-37.

[32] Vives, Xavier (1999), *Oligopoly Pricing: Old Ideas and New Tools*, Cambridge, MA: MIT Press.

[33] Wilson, Robert (1979), "Auctions of Shares," *Quarterly Journal of Economics*, vol. 93(4), 675-689.

[34] Yanelle, Marie-Odile (1997) "Banking Competition and Market Efficiency," *Review of Economic Studies*, vol. 64(2), 215-239