

Lumpy Capacity Investment and Disinvestment Dynamics

David Besanko

Kellogg School of Management, Northwestern University, Evanston, Illinois 60208, d-besanko@kellogg.northwestern.edu

Ulrich Doraszelski

Department of Economics, Harvard University, Cambridge, Massachusetts 02138, doraszelski@harvard.edu

Lauren Xiaoyuan Lu

Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, lauren_lu@unc.edu

Mark Satterthwaite

Kellogg School of Management, Northwestern University, Evanston, Illinois 60208, m-satterthwaite@kellogg.northwestern.edu

Capacity addition and withdrawal decisions are among the most important strategic decisions made by firms in oligopolistic industries. In this paper, we develop and analyze a fully dynamic model of an oligopolistic industry with lumpy capacity and lumpy investment/disinvestment. We use our model to suggest answers to two questions: First, what economic factors facilitate preemption races? Second, what economic factors facilitate capacity coordination? With a series of examples we show that low product differentiation, low investment sunkness, and high depreciation tend to promote preemption races. The same examples also show that low product differentiation and low investment sunkness tend to promote capacity coordination. Although depreciation removes capacity, it might impede capacity coordination. Finally, our examples show that multiple equilibria arise over at least some range of parameter values. The distinct structures of these equilibria suggest that firms' expectations play a key role in determining whether or not industry dynamics are characterized by preemption races and capacity coordination. Taken together, our results suggest that preemption races and excess capacity in the short run often go hand-in-hand with capacity coordination in the long run.

Subject classifications: capacity; investment and disinvestment; industry evolution; preemption races; capacity coordination; dynamic stochastic game; Markov perfect equilibrium.

Area of review: Special Issue on Computational Economics.

History: Received March 2008; revision received November 2009; accepted December 2009. Published online in *Articles in Advance* July 1, 2010.

1. Introduction

Capacity addition and withdrawal decisions are among the most important strategic decisions made by firms in oligopolistic industries. Because they are typically lumpy, these decisions to invest or disinvest can have a significant impact on price and profitability in the short run; and because they are usually long-lived, they are a critical determinant of how the competitive environment evolves in the long run. "Mistakes" in the form of overly aggressive or poorly planned capacity additions can result in excess capacity that "spoils" the market for years or even for decades. Periods of excess capacity have occurred in industries ranging from semiconductor memories to hotels and office buildings, and some industries such as railroads, steel, and oil tankers suffered from chronic excess capacity for decades (see, e.g., Lieberman 1987). On the other hand, pursuing an aggressive approach to investment might be a deliberate competitive move. In the early 1970s, for example, DuPont had a market share of 40% in the North American titanium dioxide industry. To maintain its dominance of the industry, DuPont began to construct a plant of twice the normal size in an effort to preempt expansion by rivals.

These observations raise two key questions, both for understanding industry dynamics and for formulating competitive strategy and competition policy in oligopolistic industries. First, what economic factors facilitate preemption races? Second, what economic factors facilitate capacity coordination? While preemption races entail building up excess capacity, by capacity coordination we mean that there is little (if any) excess capacity relative to the benchmark of a capacity cartel. There is thus an apparent tension between preemption races and capacity coordination.

Yet, some industries appear to exhibit both preemption races and capacity coordination. Christensen and Caves (1997) demonstrate the presence of investment rivalry in the North American paper and pulp industry. The fact that the mere announcement of a rival project makes a firm more likely to complete an investment project that it has previously announced points to "some sort of race to add capacity" (p. 48). The evidence for preemption races seems at odds with the evidence for capacity coordination. In the Canadian paper and pulp industry, total industry-wide capacity promptly adjusts to disturbances so as to restore long-run levels of price and capacity utilization

(Bernstein 1992). In the North American newsprint industry, a major sector of the paper and pulp industry, Booth et al. (1991) show that higher concentration in regional markets leads to less capacity expansion, a finding that they interpret as a weak form of capacity coordination. While some industries experience both preemption races and capacity coordination, others seem to sidestep preemption races altogether. Empirical work on the U.K. brick industry, for example, suggests that firms manage to avoid excessive investments and the “unwarranted clustering of expansions” (Wood 2005, p. 47).

The contrasting experiences of the North American newsprint and the U.K. brick industries are all the more surprising because these industries are similar in many ways. In the brick industry, capacity is lumpy, with one large modern factory being able to “supply more than 2% of the national market” and much more of the regional market within a 100-mile radius of the factory where most of its output is sold (Wood 2005, p. 39). Investment is also lumpy because the major piece of equipment of a factory is a kiln, “with the size of that kiln depending on the technology that was available at the time of its commissioning” (p. 39). Capacity and investment are also lumpy in the newsprint industry because an investment project typically involves the installation of one paper machine at an existing or a green-field site. Because these machines are very large, the average project “added 2.6 percent of industry capacity” (Christensen and Caves 1997, p. 56) and “investment... is lumpy due to a minimum efficient scale of production of 220,000 tonnes a year” (Booth et al. 1991, p. 256). Furthermore, products are essentially homogenous in the brick industry, at least within a geographic market, and also in the newsprint industry that experienced a “significant reduction in transportation costs following deregulation of the railroads” after the mid 1960s (Booth et al. 1991, p. 257). What sets the brick industry apart from the newsprint industry is the sunkness of investment. In the brick industry, a kiln cannot be switched off without risking rapid deterioration and “even collapse when re-lit” (Wood 2005, p. 39). The fact that “firms are reluctant to mothball kilns” (p. 39) indicates high investment sunkness. In the newsprint industry, in contrast, “exit... is possible through conversion of machines to produce uncoated groundwood papers that have better printing characteristics and are priced above standard newsprint” (Booth et al. 1991, p. 257), indicating low (or at least lower) investment sunkness.

To explore what industry characteristics facilitate preemption races and capacity coordination and to provide a theoretical explanation for the above observations, we apply the Markov-perfect equilibrium framework of Ericson and Pakes (1995) together with the purification technique of Doraszelski and Satterthwaite (2010) to study the evolution of an oligopolistic industry with lumpy capacity and lumpy investment/disinvestment. Incorporating incomplete information allows us to capture the strategic uncertainty that firms face about their rivals’ investment/disinvestment

decisions. Both lumpiness and stochasticity are realistic features in many industrial settings (see, e.g., Doms and Dunne 1998, Caballero and Engel 1999).

With a series of examples we show that low product differentiation and low investment sunkness tend to promote preemption races and capacity coordination. During a preemption race, firms continue investing as long as their capacities are similar. The race comes to an end once one of the firms gains the upper hand. At this point, the investment process stops and a process of disinvestment starts. During the disinvestment process some of the excess capacity that has been built up during the race is removed. This boosts the extent of capacity coordination.

Low product differentiation appears to be necessary for preemption races and capacity coordination because it intensifies capacity utilization and price competition. To bring prices and profits back up at the end of a preemption race, the loser has an incentive to lead the winner through one or more rounds of capacity withdrawal. Low investment sunkness implies high investment reversibility and promotes preemption races and capacity coordination by allowing firms to remove some of the excess capacity that has been built up during the race. In contrast, if they lack the option to disinvest, then firms have no reason to enter a preemption race in the first place because they anticipate that the industry will be permanently locked into a state of excess capacity and low profitability after the race. This is a dynamic manifestation of the maxim that “exit costs are entry barriers” and might help to explain the absence of preemption races in the U.K. brick industry, where investment is fully sunk.

Taken together, our results suggest that preemption races and excess capacity in the short run often go hand-in-hand with capacity coordination in the long run. The association of these seemingly contradictory behaviors is consistent with observing both preemption races and capacity coordination in the North American newsprint industry, where investment is partially sunk. It is also consistent with Gilbert and Lieberman’s (1987) finding that in the 24 chemical processing industries studied, preemption might be a temporary phenomenon and that “the main role of preemptive activity is to coordinate new investment and to promote efficiency by avoiding excess capacity” (p. 30).

Of course, none of this means that preemption cannot have lasting effects on industry structure. In fact, our examples show how a preemption race leads to an asymmetric industry structure in the long run. This is consistent with the dominance of DuPont of the North American titanium dioxide industry that can be traced back to the preemptive strategy of capacity accumulation that DuPont initiated in the early 1970s (Ghemawat 1984, Hall 1990).

Strategic capacity investment decisions are a classic question in industrial economics. Spence (1977) and Dixit (1980), among others, construct static investment models to show that entry can be deterred by incumbents building up excess capacity. However, industry dynamics and

the transitory nature of preemption that is observed empirically cannot be captured by these static models. Brock and Scheinkman (1985), Staiger and Wolak (1992), and Compte et al. (2002) develop dynamic collusion models in which firms operate under capacity constraints. In these models, the symmetry and stationarity assumptions on firm behavior imply that market shares remain relatively stable. In contrast, we show that a preemption race leads to an asymmetric industry structure in the long run. Our paper is most closely related to Besanko and Doraszelski (2004) but differs along three dimensions. First, in light of the empirical literature, we treat investment/disinvestment as lumpy. Second, our model is flexible enough to characterize fully or partially sunk investment. Third, we apply the homotopy method to explore the equilibrium correspondence in a systematic fashion and, in contrast to Besanko and Doraszelski (2004), find ample evidence of multiplicity.

Our paper is organized in four sections, including the introduction. Section 2 lays out the model, §3 presents the results, and §4 concludes.

2. Model

We model the evolution of an oligopolistic industry using a discrete-time, infinite-horizon dynamic stochastic game.

Setup and Timing. There are two firms, indexed by 1 and 2, with potentially different capacities \bar{q}_i and \bar{q}_j , respectively. Capacity is lumpy so that \bar{q}_i and \bar{q}_j take on one of M values, $0, \Delta, 2\Delta, \dots, (M-1)\Delta$, where $\Delta > 0$ measures the lumpiness of capacity. For notational simplicity, we take (i, j) to mean (\bar{q}_i, \bar{q}_j) . We refer to $(i, j) \in \{0, 1, 2, \dots, (M-1)\}^2$ as the state of the industry; in state (i, j) firm 1 has a capacity of $i\Delta$ units and firm 2 has a capacity of $j\Delta$ units.

At the beginning of a period, firms first learn their cost/benefit of capacity addition/withdrawal. Its cost/benefit is private to a firm and hence unknown to its rival. Then firms make investment/disinvestment decisions. They next compete in the product market. At the end of the period, the investment/disinvestment decisions are implemented, and previously installed capacity is subjected to depreciation. We first give details on the product market competition and then turn to the dynamic framework.

Demand. The two firms compete in a differentiated product market by setting prices subject to capacity constraints. Demand is derived from the utility-maximization problem of a representative consumer:

$$\max_{q_0, q_1, q_2 \geq 0} q_0 + Aq_1 + Aq_2 - \frac{B}{2}q_1^2 - \frac{B}{2}q_2^2 - \gamma Bq_1q_2,$$

subject to the budget constraint $q_0 + p_1q_1 + p_2q_2 = y$, where q_1 and q_2 are the quantities of goods 1 and 2 as purchased from firms 1 and 2 at prices p_1 and p_2 , q_0 is the numéraire good, and y the consumer's income. $\gamma \in [0, 1)$ measures the degree of product differentiation: If $\gamma = 0$,

then goods 1 and 2 are independent; as γ approaches 1, the two goods become homogeneous (perfect substitutes). Solving the consumer's problem, the demand function for firm 1 is¹

$$q_1(p_1, p_2) = \frac{1}{1-\gamma^2}(a(1-\gamma) - bp_1 + \gamma bp_2),$$

where $a = A/B > 0$ and $b = 1/B > 0$.

We do not explicitly model a firm's decision to enter or exit the industry. Instead, we simply assume that a firm's demand is zero if its capacity is zero. The firm's rival then faces the entire market demand as long as its capacity is nonzero. For example, if firm 1 has nonzero capacity and firm 2 has zero capacity, then the demand functions are $q_1(p_1, p_2) = a - bp_1$ and $q_2(p_1, p_2) = 0$ and, if both firms have zero capacity, then $q_1(p_1, p_2) = q_2(p_1, p_2) = 0$.

Soft Capacity Constraints. We assume that capacity constraints are "soft" in that they allow firms to produce any quantity, albeit at an exploding cost. In the real world, hiring temporary workers, adding shifts, or expediting material deliveries to alleviate capacity constraints are common and often costly. The production cost function of firm 1 depends on the its quantity q_1 and its capacity \bar{q}_i :

$$C(q_1, \bar{q}_i) = \frac{1}{1+\nu} \left(\frac{q_1}{\bar{q}_i} \right)^\nu q_1,$$

where $\nu \geq 0$ measures the "hardness" of the capacity constraint. The larger is ν , the closer we are to "hard" capacity constraints because the marginal cost of production $(q_1/\bar{q}_i)^\nu$ tends to either zero if $q_1 < \bar{q}_i$ or infinity if $q_1 > \bar{q}_i$.

Price Competition. Soft capacity constraints allow us to impose a "common carrier requirement": A firm is obliged to satisfy all of its demand and cannot turn away customers. This avoids specifying a rationing scheme and gives rise to a Nash equilibrium in pure strategies in the product market game (see, e.g., Maggi 1996). Suppose firms' capacities are (\bar{q}_i, \bar{q}_j) in state (i, j) . Because firms produce to satisfy demand, the profit-maximization problem of firm 1 is

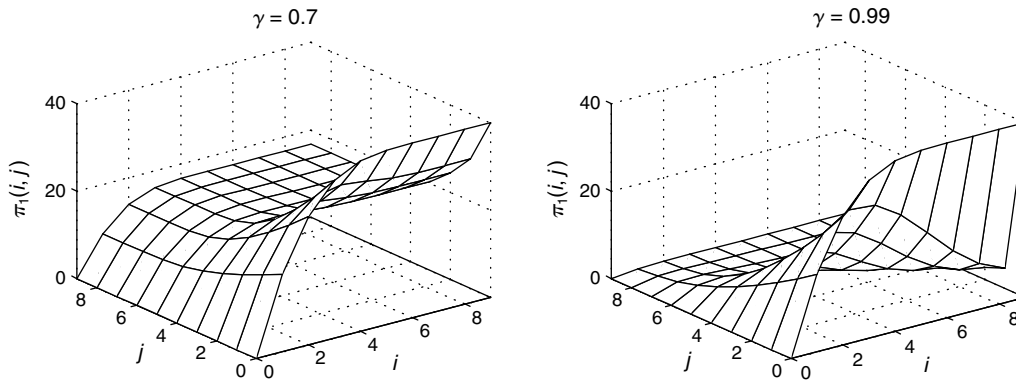
$$\max_{p_1, q_1(p_1, p_2) \geq 0} p_1 q_1(p_1, p_2) - C(q_1(p_1, p_2), \bar{q}_i).$$

Solving both firms' problems yields a Nash equilibrium in prices $(p_1(i, j), p_2(i, j))$. The single-period profit function of firm 1 thus derived is

$$\pi_1(i, j) = p_1(i, j)q_1(p_1(i, j), p_2(i, j)) - C(q_1(p_1(i, j), p_2(i, j)), \bar{q}_i).$$

Figure 1 displays the profit $\pi_1(i, j)$ of firm 1 in the Nash equilibrium of the product market game for different degrees of product differentiation. In each panel, the x - and y -axes are the capacities of firms 1 and 2 as indexed by i and j , respectively, and the z -axis is the profit of firm 1 in state

Figure 1. Profit $\pi_1(i, j)$ in the Nash equilibrium of the product market game.



Note. $\gamma = 0.7$ (left panel) and $\gamma = 0.99$ (right panel).

(i, j) . Product differentiation plays a key role in shaping the single-period profit function. When product differentiation is high ($\gamma = 0.7$, left panel), a firm's profit from product market competition plateaus in its capacity. Because of market power, marginal revenue is less than price, giving the firm an incentive to idle any unneeded capacity. In contrast, when product differentiation is low ($\gamma = 0.99$, right panel), a firm's profit first increases and then decreases in its capacity. The reason is that when goods are almost homogenous, marginal revenue almost equals price, which gives a firm an incentive to utilize its available capacity. If the total capacity in the industry is near the joint profit maximum, then the firms do well, splitting profits roughly in proportion to their capacities. But any excess capacity depresses prices and profits. Thus, holding fixed its rival's capacity, a firm's profit starts at zero with zero capacity, rises to a peak, and then falls to almost zero as its capacity additions drive the industry into a capacity glut. This same peak also arises in models of product market competition with homogenous goods and hard capacity constraints (Kreps and Scheinkman 1983, Deneckere and Kovenock 1996, Allen et al. 2000).

Investment/Disinvestment. We next turn to the dynamic framework. We treat not only capacity but also investment/disinvestment as lumpy. In particular, in our model a firm cannot expand its capacity unless it pays the entire cost of Δ units of capacity. Once this cost is paid, however, the firm is guaranteed to expand its capacity by Δ units. Hence, investment is certain.

A firm chooses from three mutually exclusive actions: Invest to add Δ units of capacity, disinvest to withdraw Δ units of capacity, or remain inactive. Before making its investment/disinvestment decisions, the firm is privately informed about its cost/benefit. The cost of capacity addition of firm 1 is $\eta_{e,1} = \phi_e + \epsilon_e \theta_1$ and its benefit of capacity withdrawal is $\eta_{w,1} = \phi_w + \epsilon_w \theta_1$, where θ_1 is a random variable drawn from a $Beta(3, 3)$ distribution with mean zero. We assume that θ_1 is drawn anew each period and that draws are independent across periods and firms. This captures the changing nature of project opportunities. ϕ_e is

the expected cost of adding capacity and ϕ_w is the expected benefit of withdrawing capacity and $\epsilon_e > 0$ and $\epsilon_w > 0$ are scale parameters.

We assume $\min_{\theta_1} \{\eta_{e,1}\} > 0$; hence, adding capacity is costly. In contrast, we impose no restrictions on the sign of $\eta_{w,1}$ so that withdrawing capacity can either yield a scrap value if $\eta_{w,1} > 0$ or require a costly payment if $\eta_{w,1} < 0$. We finally assume that $\min_{\theta_1} \{\eta_{e,1}\} > \max_{\theta_1} \{\eta_{w,1}\}$ in order to rule out arbitrage opportunities. The difference between ϕ_e and ϕ_w measures the sunkness of investment; in particular, investment is fully sunk if $\phi_w = -\infty$.

Note that θ_1 is private to firm 1 and hence unknown to firm 2. The motivation for incorporating incomplete information into our model is twofold. First, in reality, a firm hardly ever knows a rival's exact cost/benefit of capacity addition/withdrawal. The firm thus faces strategic uncertainty when it comes to its rival's investment/disinvestment decisions. Second, incorporating incomplete information allows us to apply the purification technique developed by Doraszelski and Satterthwaite (2010) that ensures that the MPE is computable. Indeed, because the cost of adding capacity is perfectly positively correlated with the benefit of withdrawing capacity, a firm's optimal investment/disinvestment decisions are given by cutoff rules. In what follows, we let

$$e_1(i, j, \theta_1) = \mathbf{1}\{\text{add } \Delta \text{ units of capacity}\},$$

$$w_1(i, j, \theta_1) = \mathbf{1}\{\text{withdraw } \Delta \text{ units of capacity}\}$$

denote the investment/disinvestment decisions. Because these decisions are mutually exclusive, we must have $e_1(i, j, \theta_1)w_1(i, j, \theta_1) = 0$. Recalling that firm 2 cannot exactly pinpoint the investment/disinvestment decisions of firm 1 because θ_1 is private to firm 1, we further let, with a slight abuse of notation,

$$e_1(i, j) = \int e_1(i, j, \theta_1) dF(\theta_1),$$

$$w_1(i, j) = \int w_1(i, j, \theta_1) dF(\theta_1)$$

denote the investment/disinvestment probabilities of firm 1 as viewed from the perspective of firm 2 (or, alternatively, from the perspective of an outside observer of the industry).

State-to-State Transitions. While a firm invests to add and disinvests to withdraw capacity, its capacity is also subject to depreciation. We think of depreciation as being of a physical nature such as obsolescence, equipment breakdowns, or natural disasters, and assume that a firm is subjected to depreciation with probability $\delta \in [0, 1]$.

Putting investment, disinvestment, and depreciation together, the state-to-state transitions of firm 1 are described by the following transition probabilities: If $i \in \{2, \dots, M - 2\}$, then

$$\Pr(i' | i, e_1, w_1) = \begin{cases} (1 - \delta)e_1 & \text{if } i' = i + 1, \\ \delta e_1 + (1 - \delta)(1 - e_1 - w_1) & \text{if } i' = i, \\ \delta(1 - e_1 - w_1) + (1 - \delta)w_1 & \text{if } i' = i - 1, \\ \delta w_1 & \text{if } i' = i - 2. \end{cases}$$

If $i \in \{0, 1, M - 1\}$, then we have to appropriately modify the above transition probability because firm 1 cannot further decrease (increase) its capacity when its previously installed capacity is already at the minimum (maximum); see the online appendix for details (<http://or.journal.informs.org>).

Bellman Equation. Let $V_1(i, j, \theta_1)$ denote the expected net present value of firm 1's cash flow if the industry is currently in state (i, j) and the firm has drawn θ_1 for its cost of capacity addition/withdrawal. In what follows, we first characterize the value function $V_1(i, j, \theta_1)$ under the presumption that firm 1 behaves optimally. In a second step, we derive the optimal investment/disinvestment decisions of firm 1, $e_1(i, j, \theta_1)$ and $w_1(i, j, \theta_1)$. Throughout, we take the investment/disinvestment probabilities of firm 2, $e_2(i, j)$ and $w_2(i, j)$, as given.

The value function $V_1(i, j, \theta_1)$ is defined recursively as the solution to the Bellman equation:

$$\begin{aligned} V_1(i, j, \theta_1) = & \max_{\substack{e_1, w_1 \in \{0, 1\}, \\ e_1 w_1 = 0}} \pi_1(i, j) + e_1 \{-\eta_{e,1} + \beta W_1^+(i, j)\} \\ & + (1 - e_1 - w_1)\beta W_1^0(i, j) \\ & + w_1 \{\eta_{w,1} + \beta W_1^-(i, j)\}, \end{aligned}$$

where $\beta \in [0, 1)$ is the discount factor. The continuation values are

$$\begin{aligned} W_1^+(i, j) &= (1 - \delta)W_1(i + 1, j) + \delta W_1(i, j), \\ W_1^0(i, j) &= (1 - \delta)W_1(i, j) + \delta W_1(i - 1, j), \\ W_1^-(i, j) &= (1 - \delta)W_1(i - 1, j) + \delta W_1(i - 2, j), \end{aligned}$$

where the superscripts $+$, 0 , and $-$ denote investment, inaction, and disinvestment, and

$$W_1(i, j) = \sum_{j'=0}^{M-1} V_1(i, j') \Pr(j' | j, e_2(i, j), w_2(i, j)).$$

While $V_1(i, j, \theta_1)$ is the value function *after* firm 1 has drawn θ_1 , $V_1(i, j) = \int V_1(i, j, \theta_1) dF(\theta_1)$ is the expected (or integrated) value function, i.e., the value function *before* firm 1 has drawn θ_1 . Note that the above expressions assume $(i, j) \in \{2, \dots, M - 2\}^2$; the appropriate modifications at the boundaries of the state space are obvious and therefore omitted for brevity.

Turning to the optimal investment/disinvestment decisions, $e_1(i, j, \theta_1) = 1$ if and only if

$$-\eta_{e,1} + \beta W_1^+(i, j) \geq \max\{\eta_{w,1} + \beta W_1^-(i, j), \beta W_1^0(i, j)\}$$

or equivalently,

$$\theta_1 \leq \underline{\theta}_1(i, j) = \min \left\{ \frac{-\phi_e - \phi_w + \beta(W_1^+(i, j) - W_1^-(i, j))}{\epsilon_e + \epsilon_w}, \frac{-\phi_e + \beta(W_1^+(i, j) - W_1^0(i, j))}{\epsilon_e} \right\}.$$

Similarly, $w_1(i, j, \theta_1) = 1$ if and only if

$$\eta_{w,1} + \beta W_1^-(i, j) \geq \max\{-\eta_{e,1} + \beta W_1^+(i, j), \beta W_1^0(i, j)\}$$

or equivalently,

$$\theta_1 \geq \bar{\theta}_1(i, j) = \max \left\{ \frac{-\phi_e - \phi_w + \beta(W_1^+(i, j) - W_1^-(i, j))}{\epsilon_e + \epsilon_w}, \frac{-\phi_w + \beta(W_1^0(i, j) - W_1^-(i, j))}{\epsilon_w} \right\}.$$

Hence, the optimal investment/disinvestment decisions are determined by the cutoffs $\underline{\theta}_1(i, j)$ and $\bar{\theta}_1(i, j)$. Notice that $\theta_1(i, j) \leq \bar{\theta}_1(i, j)$ by construction so that $e_1(i, j, \theta_1)w_1(i, j, \theta_1) = 0$ almost surely.² The implied investment/disinvestment probabilities can be written as

$$e_1(i, j) = \int \mathbf{1}[\theta_1 \leq \underline{\theta}_1(i, j)] dF(\theta_1) = F(\underline{\theta}_1(i, j)),$$

$$w_1(i, j) = \int \mathbf{1}[\theta_1 \geq \bar{\theta}_1(i, j)] dF(\theta_1) = 1 - F(\bar{\theta}_1(i, j)).$$

It remains to determine the expected value function $V_1(i, j)$. Substituting the optimal investment/disinvestment decisions into the Bellman equation and integrating both sides with respect to θ_1 gives

$$\begin{aligned} V_1(i, j) = & \pi_1(i, j) + e_1(i, j) \left\{ -\phi_e - \frac{\epsilon_e}{e_1(i, j)} \right. \\ & \cdot \int \theta_1 \mathbf{1}[\theta_1 \leq \underline{\theta}_1(i, j)] dF(\theta_1) + \beta W_1^+(i, j) \left. \right\} \\ & + (1 - e_1(i, j) - w_1(i, j))\beta W_1^0(i, j) \\ & + w_1(i, j) \left\{ \phi_w + \frac{\epsilon_w}{w_1(i, j)} \right. \\ & \cdot \int \theta_1 \mathbf{1}[\theta_1 \geq \bar{\theta}_1(i, j)] dF(\theta_1) + \beta W_1^-(i, j) \left. \right\}. \end{aligned}$$

Closed-form expressions for the above integrals are given in the online appendix. Note that an optimizing firm cares about the expected cost/benefit of adding/withdrawing capacity conditional on investing/disinvesting. The first integral, when scaled by $\epsilon_e/e_1(i, j)$, can be thought of as the deviation from the unconditional expectation of the cost of adding capacity ϕ_e . Hence, ϕ_e plus this deviation gives the expected cost of adding capacity conditional on investing. Similarly, ϕ_w plus the properly scaled second integral gives the expected benefit of withdrawing capacity conditional on disinvesting.

Equilibrium. Our model of product market competition gives rise to symmetric price and profit functions, i.e., $p_1(i, j) = p_2(j, i)$ and $\pi_1(i, j) = \pi_2(j, i)$. We therefore restrict attention to symmetric Markov-perfect equilibria (MPE) with $V_1(i, j) = V_2(j, i)$, $e_1(i, j) = e_2(j, i)$, and $w_1(i, j) = w_2(j, i)$. Existence follows from the arguments in Doraszelski and Satterthwaite (2010). From here on, we use $p(i, j)$ and $\pi(i, j)$ to denote firm 1's price and profit functions, $V(i, j)$ to denote its value function, and $e(i, j)$ and $w(i, j)$ to denote its investment/disinvestment probabilities (or policy functions).

Computation. We use the Pakes and McGuire (1994) algorithm to compute an MPE. A known problem of this iterative method is that it is not guaranteed to converge. Moreover, if there are multiple MPE, some of them cannot be computed by this method (Besanko et al. 2010). To obtain a more complete characterization of the set of MPE, we apply the homotopy method as introduced by Besanko et al. (2010) to the analysis of dynamic stochastic games (see also Borkovsky et al. 2010). Whereas the Pakes and McGuire (1994) algorithm aims to compute an MPE for a particular parameterization of the model, the homotopy method traces out an entire path through the set of MPE by varying one or more parameters of the model. We relegate the details of the homotopy method to the online appendix.

Parameterization. Throughout, we assume $M = 10$ and $\Delta = 5$. Thus capacity is fairly lumpy. The discount factor is $\beta = 0.9524$, corresponding to a yearly interest rate of 5%. We set $\nu = 10$ to approximate hard capacity constraints. The demand parameters are $a = 40$ and $b = 10$. These parameters ensure that demand is small relative to the range of capacities. In other words, it is possible for a firm to acquire more than enough capacity to supply the entire market.

We let the degree of product differentiation γ range from 0 to close to 1 with a focus on three values, namely 0.1, 0.7, and 0.99. These values correspond to almost independent goods, reasonably differentiated goods, and almost homogenous goods, respectively. The resulting intensity of product market competition can be seen from the cross-price elasticities: When both firms have 3Δ units of capacity, the cross-price elasticities are 0.161, 1.424, and 33.815, respectively.³

We set $\phi_e = 72$, $\epsilon_e = 36$, $\phi_w = 24$ or $-1,000$, and $\epsilon_w = 12$. When $\phi_w = 24$, the expected scrap value is one-third of the expected cost of adding capacity, and investment is partially sunk. However, when $\phi_w = -1,000$, withdrawing capacity is too costly to be economically viable, and investment is fully sunk. We use these two values of ϕ_w to study the role of investment sunkness.

Finally, we let the rate of depreciation δ range from 0 to 1 with a focus on three values, namely 0, 0.1, and 0.2. For the latter two values, there is a significant probability that a firm involuntarily sheds capacity.

2.1. Capacity Cartel

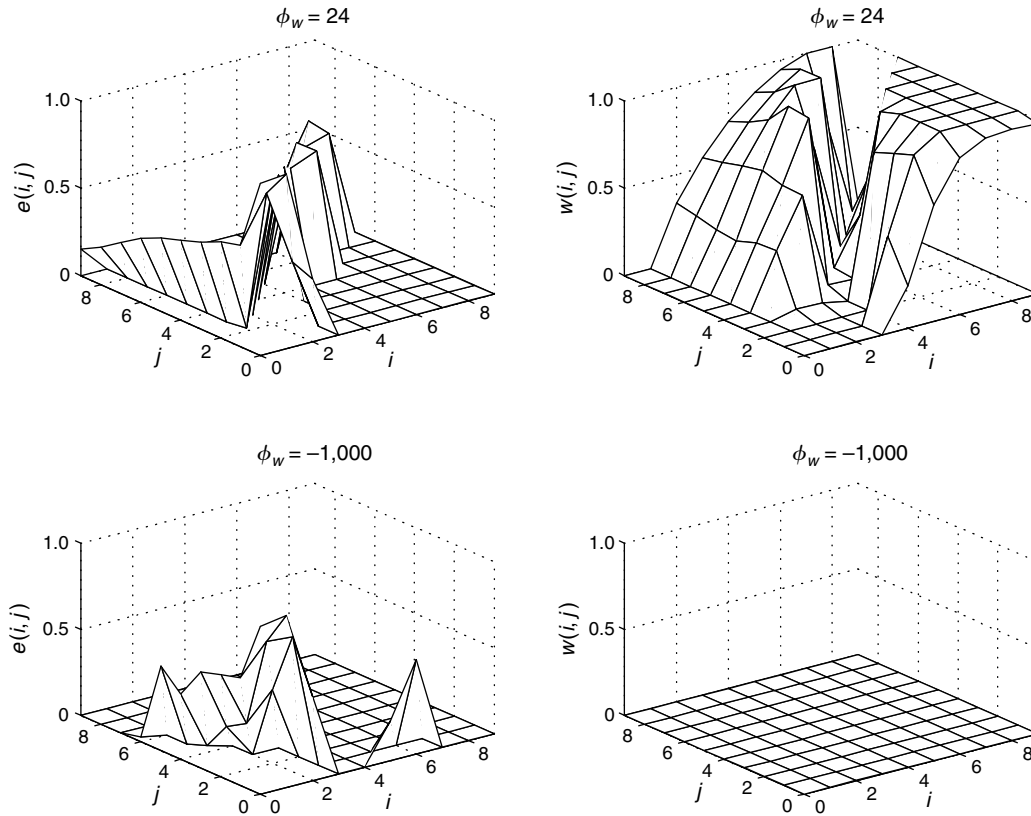
To assess the extent of capacity coordination in the industry, we contrast the behavior of the two firms in the MPE with the optimal forward-looking behavior of a capacity cartel. The cartel produces two products in two plants, indexed by 1 and 2, that have price autonomy. Thus, the single-period profit function of the cartel $\Pi(i, j) = \pi(i, j) + \pi(j, i)$ is identical to the total profits of the two firms in our baseline model. This formulation isolates the effects of capacity coordination from those of price coordination in a monopoly. The two plants have independent realizations of their cost of capacity expansion/withdrawal that are known to the cartel. The cartel has to solve a single-agent dynamic programming problem to arrive at its investment/disinvestment decisions. This problem has a unique solution that can be computed by value function iteration. We relegate the details to the online appendix.

3. Results

We organize our results to answer two questions: First, what economic factors facilitate preemption races? Second, what economic factors facilitate capacity coordination?

3.1. Preemption Races and Capacity Coordination

Conventional wisdom in investment theory holds that irreversibility lends commitment value to investment and makes preemption more credible and thus more likely to be pursued (Tirole 1988, p. 345). In our model, two parameters are tied to investment reversibility. First, the difference between ϕ_e and ϕ_w measures the sunkness of investment. In particular, decreasing the expected benefit of withdrawing capacity by reducing the scrap value ϕ_w (or even by assuming that withdrawing capacity requires a costly payment) while holding ϕ_e fixed makes investment more sunk and thus less reversible. Second, increasing the rate of depreciation makes previously installed capacity more vulnerable to depreciation and thus investment more reversible. One difference between investment sunkness and depreciation is, of course, that a firm controls capacity withdrawal through its disinvestment decision, whereas the firm does not control depreciation. Another difference is that depreciation entails neither a direct benefit nor a direct cost. We first set $\delta = 0$ to focus on the role of investment sunkness. In §3.2 we then come back to the role of depreciation.

Figure 2. Investment/disinvestment probabilities $e(i, j)$ and $w(i, j)$.

Note. $\gamma = 0.99$, $\phi_w = 24$ (upper panels) and $\phi_w = -1,000$ (lower panels), and $\delta = 0$.

Investment Sunkness. We set $\gamma = 0.99$ to model almost homogenous goods and contrast the case of partially sunk investment ($\phi_w = 24$) with the case of fully sunk investment ($\phi_w = -1,000$). Figure 2 illustrates the investment/disinvestment probabilities $e(i, j)$ (left panels) and $w(i, j)$ (right panels) for $\phi_w = 24$ (upper panels) and $\phi_w = -1,000$ (lower panels). In each panel, the x - and y -axes are the capacities of firms 1 and 2 as indexed by i and j , respectively, and the z -axis is the investment/disinvestment probabilities of firm 1 in state (i, j) . The corresponding probabilities of firm 2 can be found by looking at the mirror image reflected through the 45-degree line (because $e_2(i, j) = e(j, i)$ and $w_2(i, j) = w(j, i)$). In the online appendix we further tabulate the investment/disinvestment probabilities $e(i, j)$ and $w(i, j)$.

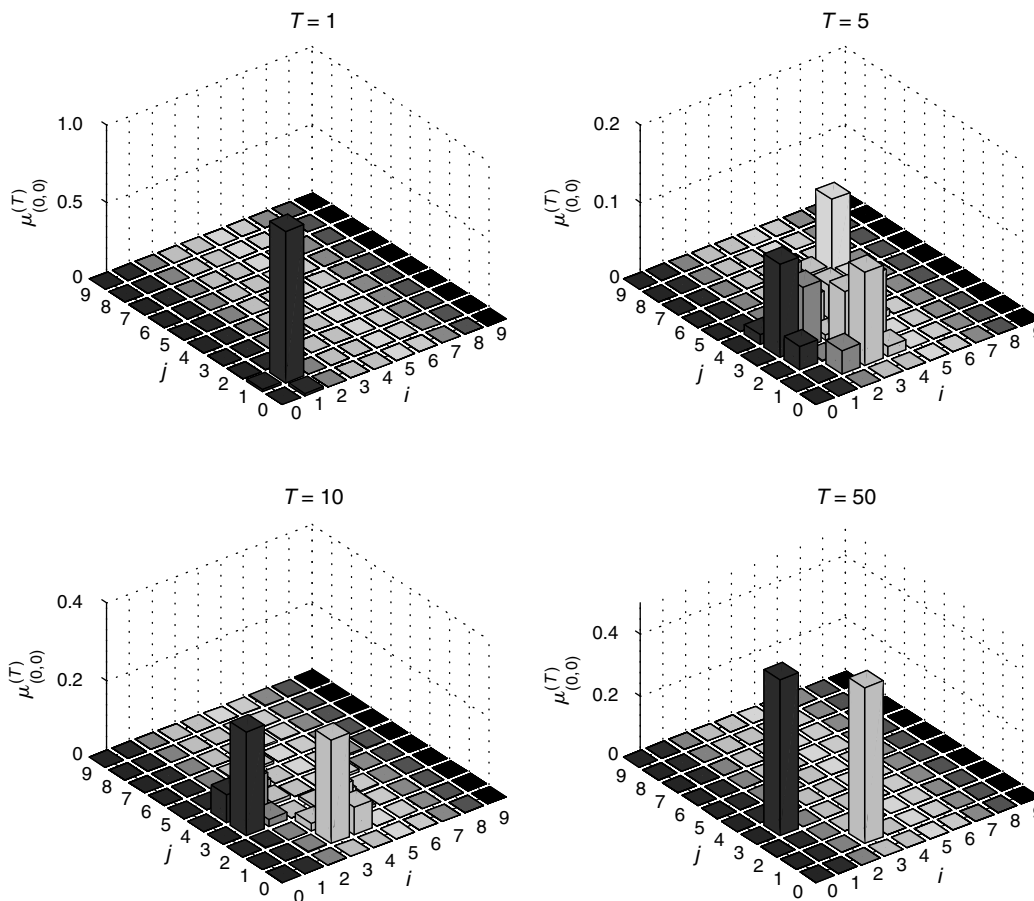
With partially sunk investment, the investment probability is high if a firm's capacity is zero (in states (i, j) with $i = 0$) because its profit from product market competition is zero if its capacity is zero, but positive if its capacity is positive; the investment probability is smaller if the rival's capacity is larger. The investment probability is also high along the diagonal of the state space (in states (i, j) with $i = j$), indicating that a firm invests aggressively when it and its rival have equal or at least similar capacities. In contrast, the firm does not disinvest under these circumstances. Noticeable disinvestment occurs when a firm is either a

fairly large leader (in states (i, j) with $i > j$ and $i > 3$) or a not-too-small follower (in states (i, j) with $i < j$ and $i > 1$).

The “diagonal” investment/disinvestment pattern indicates strong rivalry in adding capacity. This rivalry resembles a preemption race. Firms start off the race by investing, and they continue investing as long as their capacities are similar. The race comes to an end when one of the firms gains the upper hand. At this point, the investment process stops and a process of disinvestment starts. During the disinvestment process, some of the excess capacity that has been built up during the race is removed.

Figure 3 illustrates the preemption race and the resulting asymmetric industry structure. It displays $\mu_{(0,0)}^{(T)} = (\mu_{(0,0)}^{(T)}(i, j))$, the distribution over states in period $T = 1, 5, 10, 50$ starting from state $(0, 0)$, as implied by firms' investment/disinvestment strategies in the MPE.⁴ In each panel, the x - and y -axes are the capacities of firms 1 and 2 as indexed by i and j , respectively, and the z -axis is the probability that the industry is in state (i, j) in period T . By period 5, both firms have most likely (with a probability of 0.125) built up 5Δ units of capacity, indicating that they must have relentlessly invested in every period. The second most likely states are $(3, 1)$ and $(1, 3)$ (each with a probability of 0.120), where the winner of the preemption race has already been decided. By period 10, the preemption race has most likely ended. By period 50, one of the firms has surely acquired

Figure 3. Distribution over states $\mu_{(0,0)}^{(T)}$ in period $T = 1, 5, 10, 50$ with initial state $(0, 0)$.



Note. $\gamma = 0.99$, $\phi_w = 24$, and $\delta = 0$.

industry leadership with 3Δ units of capacity and marginalized its rival with Δ units of capacity. Thus, the total capacity of the industry is likely to fall over time after period 5.

With fully sunk investment, the disinvestment probability is zero by construction (see the lower-right panel of Figure 2). Removing the option to disinvest also has a dramatic impact on firms' investment strategies and the implied industry dynamics. As can be seen in the lower-left panel of Figure 2, the diagonal investment pattern disappears, so that a preemption race does not occur. Investment activities mostly occur when the total capacity of the industry is lower than 6Δ units.⁵ A firm seems to gradually build up its capacity to a certain target level and then recede forever into inactivity. The industry evolves toward a symmetric structure as illustrated in Figure 4. By period 5, the two firms have most likely built up a total capacity of 5Δ units that is split slightly unevenly between them: the most likely states are $(3, 2)$ and $(2, 3)$ (each with a probability of 0.314). By period 50, the industry has settled at one of three states: $(3, 2)$ or $(2, 3)$ (each with a probability of 0.414) or $(3, 3)$ (with a probability of 0.178). As time passes, a symmetric industry structure becomes more likely. Indeed, state $(3, 3)$ has probability one in the limiting distribution $\mu^{(\infty)} = (\mu^{(\infty)}(i, j))$.

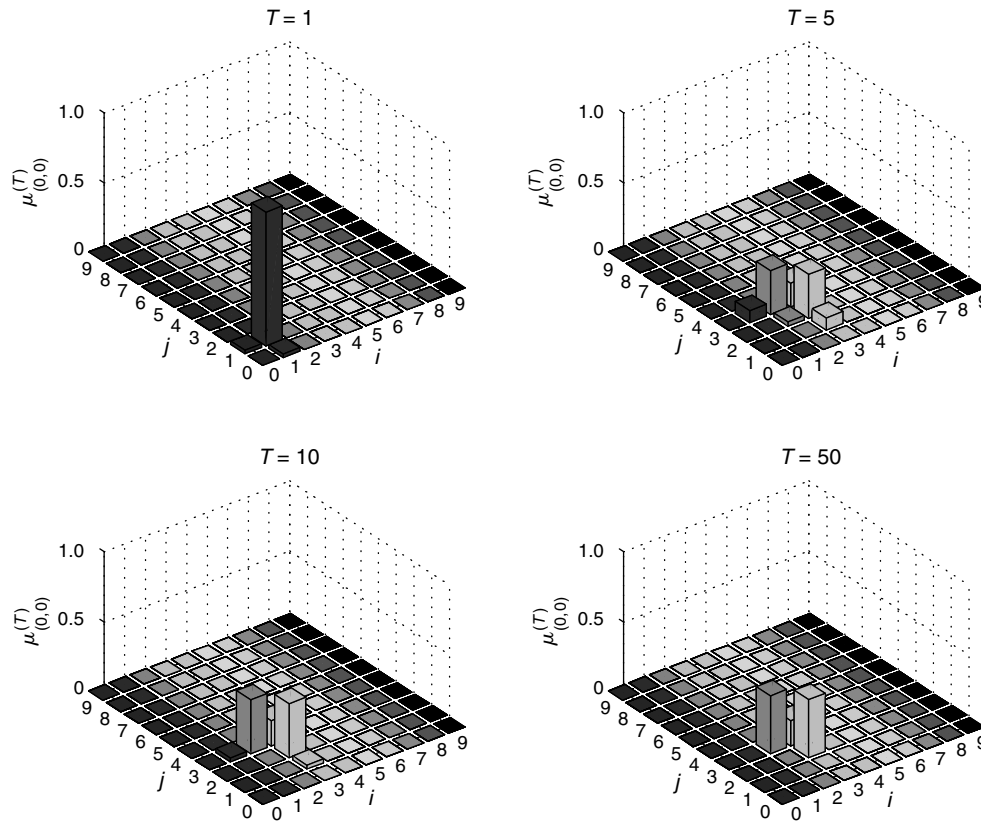
In sum, with partially sunk investment, firms invest aggressively in a bid for industry leadership. This naturally leads to a preemption race and a concomitant build-up of excess capacity in the short run. With fully sunk investment, on the other hand, firms adopt a more timid investment strategy. The question therefore is: Does a preemption race mean that the industry becomes stuck forever in a state of excess capacity and low profitability? Quite to the contrary, as we argue below, the phase of excess capacity might be transitory, and in the long run capacity coordination might be achieved through the disinvestment process that follows a preemption race.

To make this point, we compute the time path of the total capacity of the industry implied by the MPE as

$$\bar{q}_{(0,0)}^{(T)} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} (\bar{q}_i + \bar{q}_j) \mu_{(0,0)}^{(T)}(i, j). \quad (1)$$

The upper-left panel of Figure 5 displays the result as a dashed line for the case of partially sunk investment. The x -axis is the period T , the y -axis is the total capacity in the duopoly in period T . The total capacity peaks slightly above 6Δ units at around period 5. In the long run, it settles at around 4Δ units. Thus, the industry suffers excess capacity as high as 50% above its long-run level. Yet, the excess

Figure 4. Distribution over states $\mu_{(0,0)}^{(T)}$ in period $T = 1, 5, 10, 50$ with initial state $(0, 0)$.



Note. $\gamma = 0.99$, $\phi_w = -1,000$ and $\delta = 0$.

capacity that has been built up during the preemption race is quickly removed after the race comes to an end. In fact, in the long run, capacity coordination obtains, and there is little excess capacity relative to our benchmark of a capacity cartel. The upper-left panel of Figure 5 displays the time path of the total capacity in the cartel as a solid line. As can be seen, when investment is partially sunk, the total capacity in the duopoly (dashed line) converges towards the total capacity in the cartel (solid line), an instance of capacity coordination.

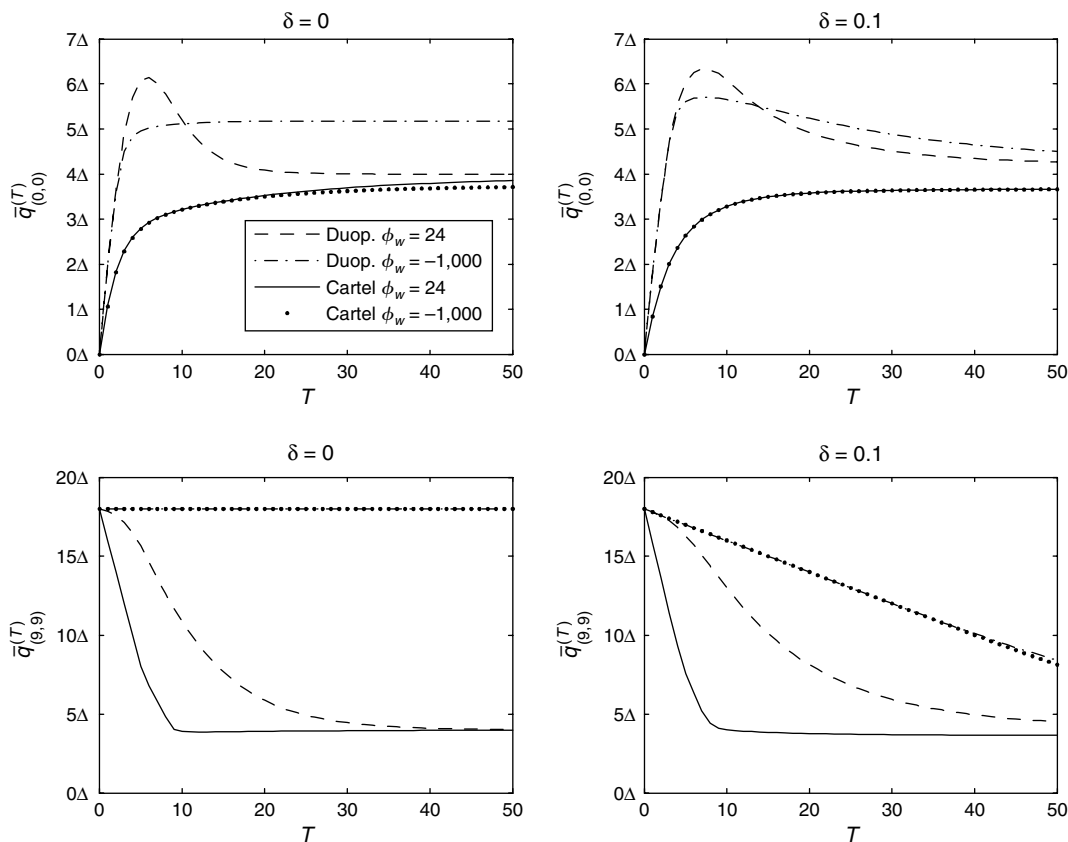
The upper-left panel of Figure 5 also displays the time path of the total capacity in the duopoly when investment is fully sunk as a dash-dotted line. The total capacity does not peak; instead, it is built up gradually and plateaus slightly above 5Δ units. Comparing this to the time path of the total capacity in the cartel (dotted line), it is obvious that, when investment is fully sunk, the industry becomes stuck forever in a state of excess capacity.

The lower-left panel of Figure 5 is analogous to the upper-left panel but replaces $\mu_{(0,0)}^{(T)}$ by $\mu_{(9,9)}^{(T)}$ in Equation (1). Starting the industry in state $(9, 9)$ models a situation of massive excess capacity. This can be thought of as an industry that has established long-run equilibrium capacities of 9Δ unit per firm, but then faces an unanticipated collapse of demand. When investment is partially sunk, both the duopoly (dashed line) and the cartel (solid line) remove excess capacity. Not

surprisingly, the cartel sheds capacity at a faster rate than the duopoly. The duopolists disinvest much more reluctantly than the cartel because they fight for industry leadership along the way. Indeed, in state $(9, 9)$, a war of attrition takes place, and the disinvestment probability $w(9, 9) = 0.068$ is very low. Eventually, however, one of the firms draws a high scrap value and disinvests. This breaks the deadlock: the disinvestment probability of the smaller firm jumps to one ($w(8, 9) = 1$) and that of the larger firm to zero ($w(9, 8) = 0$), ensuring that in the next period the gap in capacities between the smaller and larger firm widens to 2Δ units. This gap is large enough for the larger firm to follow behind the smaller firm in withdrawing capacity while avoiding triggering a preemption race between equal-sized firms. In the long run, the total capacity in the duopoly (dashed line) converges toward the total capacity in the cartel (solid line), another instance of capacity coordination. This clearly indicates that the option to disinvest resolves the problem of excess capacity. When investment is fully sunk, neither the duopoly (dash-dotted line) nor the cartel (dotted line) can remove any excess capacity, and the industry remains stuck forever in the initial state.

In sum, capacity coordination succeeds when firms fight a preemption race but fails otherwise. Put differently, preemption races in the short run go hand-in-hand with capacity coordination in the long run. The question therefore is: Why

Figure 5. Total capacity $\bar{q}^{(T)}$ in period T with initial state (0, 0) (upper panels) and (9, 9) (lower panels).



Notes. $\gamma = 0.99$, $\delta = 0$ (left panels), and $\delta = 0.1$ (right panels). Duopoly with $\phi_w = 24$ (dashed line) and $\phi_w = -1,000$ (dash-dotted line) and cartel with $\phi_w = 24$ (solid line) and $\phi_w = -1,000$ (dotted line). In the upper panels the solid line overlaps with the dotted line. In the lower panels the dash-dotted line overlaps with the dotted line.

do firms build up excess capacity initially only to remove it later on? The sunkness of investment plays a key role in promoting both preemption races and capacity coordination because it affects whether or not firms can remove capacity. When investment is partially sunk, firms know they can remove capacity and recover part of the investment cost. Because the option to disinvest is economically viable, firms are emboldened to fight a preemption race in a bid for industry leadership. However, because a firm's cost of adding capacity varies from period to period, one of the firms is bound to get unlucky with a high draw and thus lose the race at some point. The loser of the race gives up by removing capacity. Note from the upper-right panel of Figure 2 that firm 1's disinvestment probabilities are high in states (3, 4), (4, 5), (5, 6), and (6, 7), where firm 1 is the follower. The leader also removes capacity. However, to preserve its leadership and avoid restarting the preemption race, the leader does not disinvest as aggressively as the follower until the gap between the two firms is at least 2Δ units of capacity: firm 1's disinvestment probabilities in states (4, 1), (5, 3), (6, 4), and (7, 5) are much higher than those in states (4, 3), (5, 4), (6, 5), and (7, 6).

Because removing capacity reduces the pressure on prices and enhances the profitability of the industry, it is in

fact in the self-interest of both the leader and the follower to start a disinvestment process at the end of a preemption race. To more clearly see why, consider the profit function in the right panel of Figure 1. As we discussed in §2, with low product differentiation (i.e., high γ), a firm's profit peaks in its capacity. Hence, it is often better for the smaller firm to be considerably smaller than the larger firm rather than slightly smaller. For example, if the smaller firm has 4Δ units of capacity and the larger firm has 5Δ units, then the smaller firm earns a profit of 3.50. On the other hand, if the smaller firm were to scale back to 2Δ units of capacity, then it earns profit of 5.96. Of course, the smaller firm would prefer that the larger firm removes capacity from the industry, but the larger firm has fought for the leadership position, won the preemption race, and will remove capacity only if the smaller firm does so first. Thus there is a benefit for the smaller firm in assuming a "puppy dog" posture while allowing its rival to be a "top dog." Once the follower has scaled back to 2Δ units of capacity, the leader has an incentive to remove capacity because its profit stands to increase from 14.55 to 20.13 if it scales back from 5Δ to 3Δ units. It is thus also in the self-interest of the leader to remove capacity. This disinvestment process, in which

the smaller firm leads and the larger firm follows, continues until the total capacity in the duopoly approaches that in a cartel. At that point, withdrawing capacity no longer increases profitability, and the industry settles into an asymmetric structure where the leader has 3Δ units of capacity and the follower has Δ units.

Preemption races and capacity coordination hinge on the fact that investment is partially sunk. If investment is fully sunk, then the incentive to preempt vanishes. Anticipating that the industry will be in a state of excess capacity and low profitability after the race, without the option to disinvest firms have no reason to enter the race in the first place. Instead, a firm gradually builds up its capacity to a certain target level (3Δ units in our example) and then stops there. When both firms have 3Δ units of capacity, they each earn a profit of 13.91. Neither would like to further expand to earn a reduced profit of 11.68 because this reduction in profit will be permanent once the capacity has been added. In states (2, 3) and (3, 2), the smaller firm can improve its profit from 13.47 to 13.91 by adding one unit of capacity. Because capacity is costly, however, and the improvement in profit is small, the firm has to wait for a very favorable draw before it pays to add capacity. This explains the slow transition from states (2, 3) and (3, 2) to state (3, 3).⁶

Product Differentiation. When $\gamma = 0.7$ and goods are reasonably differentiated, firms' investment/disinvestment strategies and the implied industry dynamics change drastically even when we continue to set $\phi_w = 24$ to model partially sunk investment. The investment/disinvestment probabilities of a firm are almost completely independent of the capacity of its rival, as Figure 6 illustrates. Clearly, no trace of a preemption race can be seen. As expected, this leads to a symmetric industry structure both in the short run and in the long run. The time path of the total capacity in the duopoly in Figure 7 indicates that firms gradually "invest up" to a certain target level (3Δ units per firm). Moreover, as long as firms start below the target level, it does not matter whether investment is partially or fully sunk: the dashed and dash-dotted lines coincide.⁷ Comparing the time path of the total capacity in the duopoly (dashed and dash-dotted lines) with that in the cartel (solid

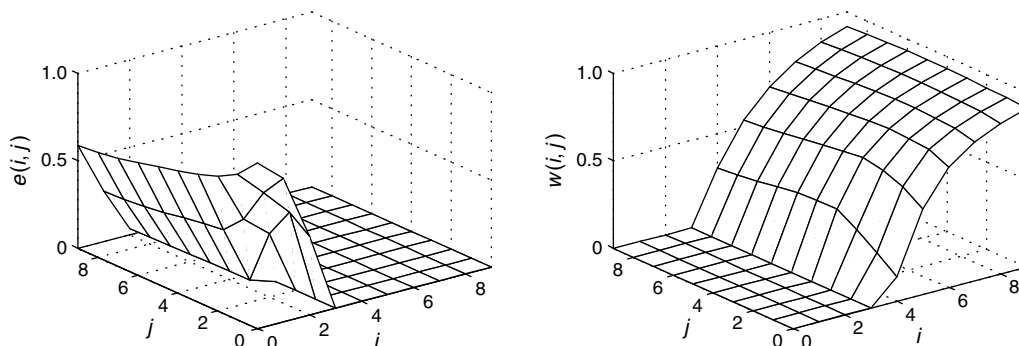
and dotted lines) shows that the industry suffers excess capacity as high as 50%. This suggests that low product differentiation is necessary not only for preemption races but also for capacity coordination.

It might seem puzzling that more product differentiation leads to less capacity coordination in this example. A closer look at the profit function in the left panel of Figure 1 resolves this apparent paradox. With high product differentiation (i.e., low γ), a firm's profit plateaus in its capacity. Except when the rival has zero capacity, the plateau is reached at 4Δ units of capacity, but the improvement in profit from 3Δ to 4Δ units is negligible relative to the cost of adding capacity. Once a firm has made it to the cusp of the plateau, it therefore has no incentive to further increase its capacity. Moreover, the firm has no incentive to decrease its capacity, not even when it is facing a larger firm, because the intense price competition associated with almost homogenous goods does not exist here to incentivize the smaller firm to take a "puppy dog" position. Consequently, the industry becomes stuck forever in a state of excess capacity.

Of course, in the extreme case of independent goods ($\gamma = 0$), the distinction between duopoly and cartel is moot and capacity coordination obtains trivially. Hence, in the vicinity of this case, more product differentiation must lead to more capacity coordination. If we set $\gamma = 0.1$ to model almost independent goods, then a firm's investment and disinvestment decisions become completely independent of its rival's capacity level. Not surprisingly, the time path of the total capacity (not shown) for this case indicates that the gap between the duopoly and the cartel is much smaller than in Figure 7, indicating a greater extent of capacity coordination.

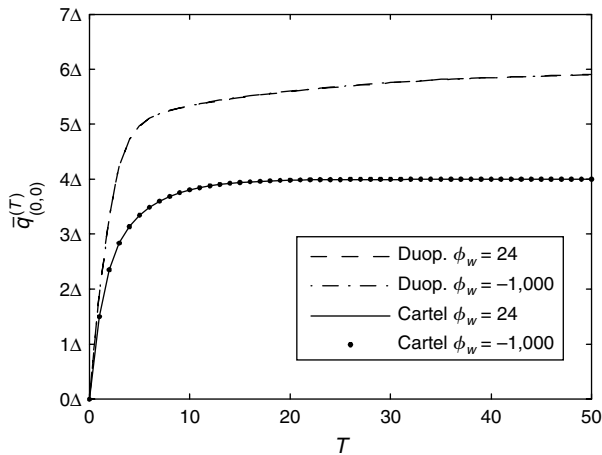
Summary. So far, our examples have shown that low product differentiation (i.e., high γ) and low investment sunkness (i.e., high ϕ_w) might together be sufficient for preemption races and capacity coordination. This suggests that preemption races and excess capacity in the short run often go hand-in-hand with capacity coordination in the long run. Low product differentiation appears to be necessary for preemption races and capacity coordination

Figure 6. Investment/disinvestment probabilities $e(i, j)$ and $w(i, j)$.



Note. $\gamma = 0.7$, $\phi_w = 24$, and $\delta = 0$.

Figure 7. Total capacity $\bar{q}_{(0,0)}^{(T)}$ in period T with initial state $(0, 0)$.



Notes. $\gamma = 0.7$ and $\delta = 0$. Duopoly with $\phi_w = 24$ (dashed line) and $\phi_w = -1,000$ (dash-dotted line) and cartel with $\phi_w = 24$ (solid line) and $\phi_w = -1,000$ (dotted line). The dashed line overlaps with the dash-dotted line. The solid line overlaps with the dotted line.

because it intensifies capacity utilization and price competition and ensures that the leader and the follower have a common interest in starting a disinvestment process at the end of a preemption race. Contrary to conventional wisdom in investment theory, it is actually low sunkness, and thus high reversibility, of investment that promotes preemption races by allowing firms to invoke the option to disinvest and remove some of the excess capacity that has been built up during the race.

3.2. Depreciation

In our model, depreciation is another source of investment reversibility. If investment sunkness promotes preemption races and capacity coordination, does depreciation do the same? Besanko and Doraszelski (2004) have shown that depreciation indeed induces preemption races when firms compete in a homogenous product market by setting prices subject to capacity constraints. We next study the role of depreciation in our model where firms can invoke the option to disinvest. Unless otherwise noted, we set $\delta = 0.1$ so that there is a significant probability that a firm's capacity shrinks if it does not invest.

Starting with the case of almost homogenous goods ($\gamma = 0.99$), our results readily confirm that depreciation induces preemption races when investment is fully sunk (as assumed by Besanko and Doraszelski 2004). In the upper panels of Figure 5, we see a preemption race with depreciation (see the dash-dotted line in the upper-right panel) where there was none without depreciation (see the dash-dotted line in the upper-left panel). The reason that depreciation induces preemption races is the same as the reason that partially sunk investment induces preemption races: both economic factors enable a reduction in some of the excess capacity that has been built up during the race.

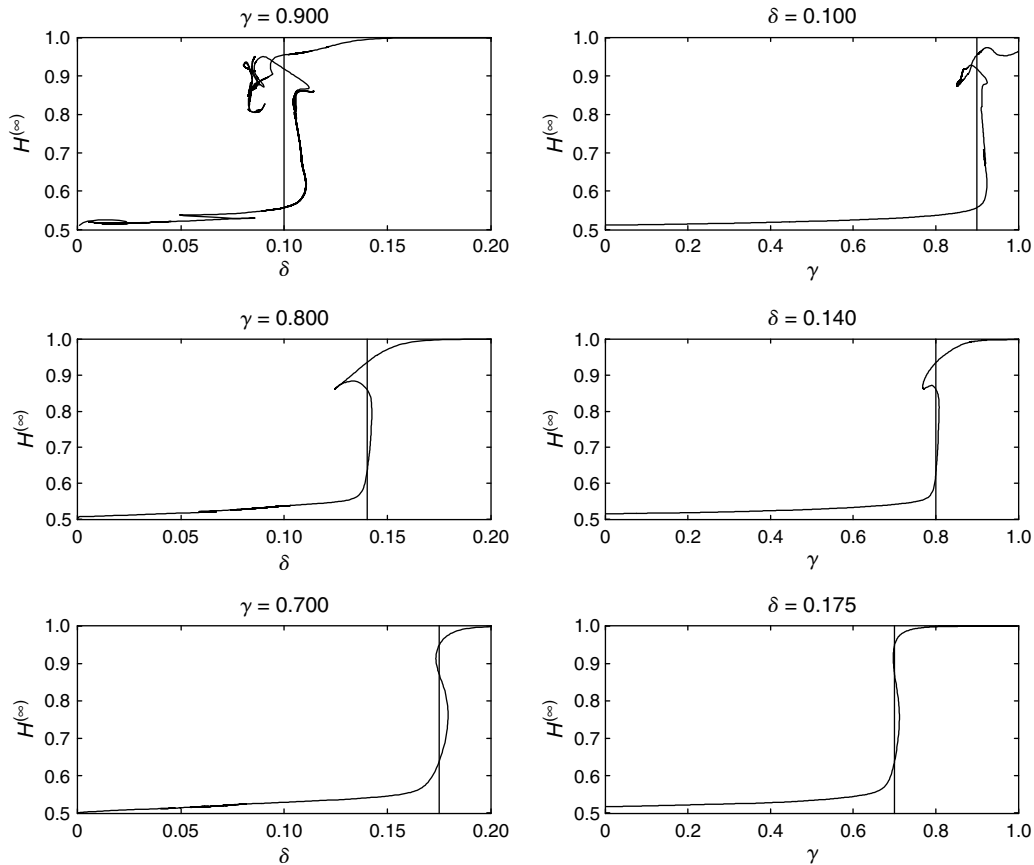
Turning to the case of reasonably differentiated goods ($\gamma = 0.7$), we once more see a preemption race with sufficiently high depreciation ($\delta = 0.2$, not shown) where there was none without depreciation (see the dashed line in Figure 7). In §3.3 we show that for a given degree of product differentiation, there exists a threshold for the rate of depreciation above which a preemption race occurs, and that this threshold increases as the degree of product differentiation increases (i.e., γ decreases).

Perhaps more interestingly, depreciation renders preemption races fiercer when investment is partially sunk. Returning to the case of almost homogenous goods ($\gamma = 0.99$), consider the upper panels of Figure 5. Note that the peak capacity built up during the preemption race with depreciation (see the dashed line in the upper-right panel) exceeds that without depreciation (see the dashed line in the upper-left panel). The key feature of depreciation is that it removes capacity whether or not this is in the interest of firms. Because depreciation is beyond the control of firms, it ensures that any excess capacity will be removed swiftly at the end of the preemption race. Knowing this emboldens firms to fight more fiercely during the preemption race.

Because depreciation removes capacity whether or not this is in the interest of firms, a reasonable conjecture is that depreciation alleviates the problem of excess capacity, and it often does. For example, in the upper panels of Figure 5, we see that when investment is fully sunk, the extent of capacity coordination is greater with depreciation (see the difference between the dash-dotted and dotted lines in the upper-right panel) than without depreciation (see the difference between the dash-dotted and dotted lines in the upper-left panel).

However, depreciation could also impede capacity coordination. For example, in the upper panels of Figure 5, we see that when investment is partially sunk, the extent of capacity coordination is smaller with depreciation (see the difference between the dashed and solid lines in the upper-right panel) than without depreciation (see the difference between the dashed and solid lines in the upper-left panel). In the lower panels of Figure 5, firms shed capacity at a slower rate with depreciation (see the dashed line in the lower-right panel) than without depreciation (see the dashed line in the lower-left panel) when investment is partially sunk. By period 10, for example, the total capacity of the industry is around 13Δ units with depreciation but around 10Δ units without depreciation. In both cases the reason is that, because the leader wants to preserve its position relative to the follower, its disinvestment activities are more cautious. After all, because depreciation is beyond the control of firms, there is always a risk that the industry leadership is lost to depreciation. To counter this risk, the leader pursues a more cautious disinvestment strategy and keeps a “safety stock” of capacity so as to prevent unintended restoration of symmetry between firms. Hence, in this case, depreciation impedes rather than promotes capacity coordination.⁸

Figure 8. Herfindahl index of firms' capacities $H^{(\infty)}$.



Note. $\phi_w = 24$.

In sum, depreciation promotes preemption races but it might promote or impede capacity coordination. This shows that depreciation—the involuntary withdrawal of capacity—and disinvestment—the voluntary withdrawal of capacity—are less than perfect substitutes.

3.3. Multiplicity

Our model has multiple MPE for a range of parameter values. We use the homotopy method to explore the equilibrium correspondence in a systematic fashion by tracing out paths through the set of MPE. Throughout, we focus on the case of partially sunk investment ($\phi_w = 24$). To succinctly describe an MPE on a path, we compute the Herfindahl index of firms' capacities implied by the MPE as

$$H^{(\infty)} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \left(\left(\frac{\bar{q}_i}{\bar{q}_i + \bar{q}_j} \right)^2 + \left(\frac{\bar{q}_j}{\bar{q}_i + \bar{q}_j} \right)^2 \right) \mu^{(\infty)}(i, j),$$

where $\mu^{(\infty)}$ is the limiting distribution over states. The Herfindahl index summarizes expected industry structure and dynamics. It ranges from 0.5 in an industry with two equal-sized firms to 1 in an industry with one firm. A higher Herfindahl index therefore indicates a more asymmetric industry structure in the long run.

Figure 8 visualizes the equilibrium correspondence. In the left panels, we trace out paths through the set of MPE by varying the rate of depreciation δ (x -axis) while holding the remaining parameters fixed and, in the right panels, by varying the degree of product differentiation γ (x -axis). In each panel, the y -axis is $H^{(\infty)}$. As can be seen from Figure 8, higher depreciation tends to lead to more asymmetric industry structures as does lower product differentiation (i.e., higher γ). Strikingly, the Herfindahl index in each panel very suddenly shoots up from close to 0.5 to close to 1. This indicates that the transition from a symmetric to an extremely asymmetric industry structure occurs rapidly around certain critical parameter values. In the vicinity of these critical values, multiplicity is pervasive.

The transition from a symmetric to an extremely asymmetric industry structure occurs around $\delta = 0.1$ in the upper-left panel of Figure 8 and around $\gamma = 0.9$ in the upper-right panel. We mark these critical values by vertical lines in both panels. Note that the upper panels “intersect” each other at the vertical line. That is, the slices through the equilibrium correspondence in the upper panels are centered at the point $(\gamma = 0.9, \delta = 0.1)$. The middle and lower panels are constructed analogously and centered at the point $(\gamma = 0.8, \delta = 0.14)$ and $(\gamma = 0.7, \delta = 0.175)$, respectively. It thus appears that if we increase the rate of depreciation,

then we also have to increase the degree of product differentiation (i.e., decrease γ) in order to continue to obtain a symmetric industry structure.

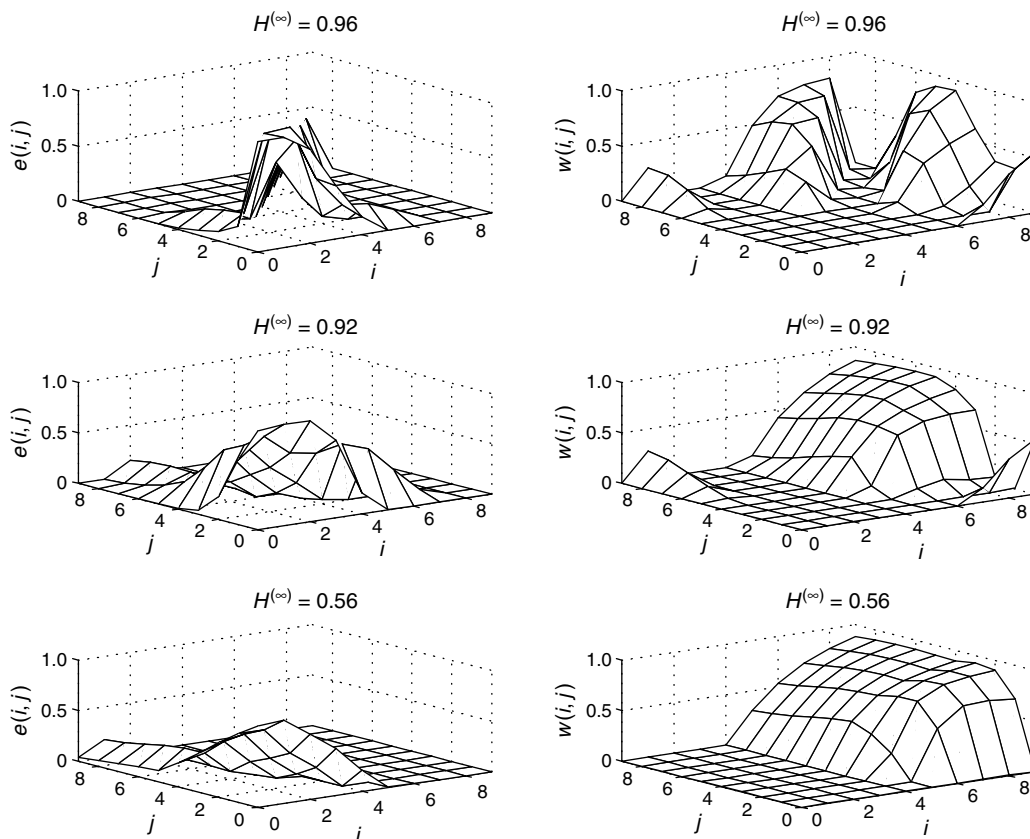
To understand this trade-off, note that a higher rate of depreciation makes it harder and therefore more costly for a firm to acquire and maintain its capacity. On the other hand, a higher degree of product differentiation (i.e., a lower γ) softens product market competition and thus leads to higher profits. Hence, to continue to obtain a symmetric industry structure, we have to offset the higher cost of capacity due to the increase in the rate of depreciation by the higher profit due to the increase in the degree of product differentiation. Otherwise, we obtain an extremely asymmetric industry structure, with one firm being nearly a monopolist.

Multiplicity arises around the critical parameter values, suggesting that firms' expectations about the evolution of the industry matter. If γ and δ are too high relative to the critical values, the market cannot possibly sustain two profitable firms. Knowing this, firms have no choice but to invest aggressively. The result is a preemption race and, ultimately, an extremely asymmetric industry structure, with the winner of the race attaining industry leadership. Conversely, if γ and δ are too low, the market can easily sustain two profitable firms, and firms might as well adopt a more timid investment strategy. In both cases, the

primitives of the model tie down the MPE; but the primitives of the model no longer suffice to tie down the MPE around the critical parameter values. Thus the stage is set for multiplicity, with MPE that might or might not entail preemption races and capacity coordination.

Figure 9 makes this point by displaying the investment/disinvestment probabilities in the three MPE that we have found for $\gamma = 0.9$ and $\delta = 0.1$. The three MPE are displayed in decreasing order of $H^{(\infty)}$ (visually, they correspond to the intersections, from the top down, of the equilibrium correspondence and the vertical lines in the upper panels of Figure 8). The MPE in the upper panels displays the diagonal investment/disinvestment pattern that is characteristic for a preemption race. The MPE in the lower panels exhibits much more cautious behavior. Investment activities are spread out over the low capacity states and disinvestment activities over the high capacity states. Moreover, the investment/disinvestment probabilities of a firm are almost completely independent of the capacity of its rival. Taken together, it is not surprising that this leads to a symmetric industry structure with no or little capacity coordination. The MPE in the middle panels appears to be a "mixture" of the other two MPE and can be thought of as the transition from a market that can sustain two profitable firms to one that cannot.

Figure 9. Investment/disinvestment probabilities $e(i, j)$ and $w(i, j)$.



Note. Multiple MPE. $\gamma = 0.9$, $\phi_w = 24$, and $\delta = 0.1$.

We conjecture that equilibrium selection is driven by firms' expectations. Around the critical parameter values, whether or not the market can sustain two profitable firms depends on how firms behave. Firms' behavior, in turn, depends on their expectations. If firms believe that the market cannot sustain two profitable firms, then they will invest aggressively in a bid for industry leadership. This naturally leads to a preemption race and, in the long run, to an extremely asymmetric industry structure exhibiting capacity coordination.⁹ If, however, firms believe that the market can sustain two profitable firms, then they will avoid a preemption race and the concomitant build-up of excess capacity in the short run. This, in turn, ensures that the market can indeed sustain two profitable firms. In sum, how the industry evolves depends on how firms expect the industry to evolve.

4. Conclusions

In this paper we apply the Markov-perfect equilibrium framework of Ericson and Pakes (1995) together with the purification technique of Doraszelski and Satterthwaite (2010) to study the evolution of an oligopolistic industry with lumpy capacity and lumpy investment/disinvestment. Because firms cannot pinpoint rivals' exact cost/benefit of capacity addition/withdrawal, they face strategic uncertainty about their rivals' investment/disinvestment decisions.

We use our model to construct a series of examples that suggest answers to two questions: First, what economic factors facilitate preemption races? Second, what economic factors facilitate capacity coordination? Preemption races entail building up excess capacity in the short run. In contrast, capacity coordination means that there is little (if any) excess capacity relative to our benchmark of a capacity cartel. Yet our results suggest that preemption races and excess capacity in the short run often go hand-in-hand with capacity coordination in the long run. In particular, we show that low product differentiation and low investment sunkness tend to promote preemption races and capacity coordination. Although depreciation promotes preemption races, it might impede capacity coordination. Finally, we show that multiple equilibria arise over at least some range of parameter values, which suggests that firms' expectations play a key role in determining whether or not industry dynamics are characterized by preemption races and capacity coordination.

5. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://or.journal.informs.org/>.

Endnotes

1. Throughout we focus on firm 1; the derivations for firm 2 are analogous.

2. $\theta_1 = \underline{\theta}_1(i, j) = \bar{\theta}_1(i, j)$ implies $e_1(i, j, \theta_1) = w_1(i, j, \theta_1) = 1$, but this event has probability zero.
3. Because they depend on firms' capacities, the cross-price elasticities vary across states. Nevertheless, because firms' long-run capacities are close to 3Δ units in the MPE we study later on, the reported numbers give a sufficiently accurate picture of the intensity of competition.
4. Let P be the $M^2 \times M^2$ transition matrix of the Markov process of industry dynamics constructed from the investment/disinvestment probabilities $e(i, j)$ and $w(i, j)$. The distribution over states in period T is given by $\mu^{(T)} = \mu^{(0)}P^T$, where $\mu^{(0)}$ is the $1 \times M^2$ initial distribution. We focus on two initial distributions, one that puts probability one on state $(0, 0)$ and one that puts probability one on state $(9, 9)$. We denote the resulting distributions in period T by $\mu_{(0,0)}^{(T)}$ and $\mu_{(9,9)}^{(T)}$, respectively. The limiting distribution over states solves $\mu^{(\infty)} = \mu^{(\infty)}P$.
5. The spike in the investment probability of firm 1 when $i = 5$ or 6 , and $j = 0$ occurs because firm 2 ceases to invest, and *de facto* does not enter the industry if firm 1 succeeds in building up at least 6Δ units of capacity: $e(0, j)$ is close to zero when $j \geq 6$.
6. Recall that state $(3, 3)$ has probability one in the limiting distribution $\mu^{(\infty)}$.
7. If firms start above the target level, then it clearly matters whether or not they can invoke the option to disinvest. In particular, if investment is fully sunk, then the industry remains stuck forever in the initial state.
8. Of course, in the extreme case of $\delta = 1$, the industry never takes off and capacity coordination succeeds trivially.
9. Because an industry with an extremely asymmetric structure is close to a monopoly, it is not surprising that capacity coordination obtains.

Acknowledgments

The authors thank Ron Borkovsky, Jeff Campbell, Jiawei Chen, Steve Kryukov, Ken Judd, and an anonymous referee for helpful comments and discussions. Besanko and Doraszelski gratefully acknowledge financial support from the National Science Foundation under grant 0615615. Satterthwaite acknowledges gratefully that this material is based upon work supported by the National Science Foundation under grant 0121541.

References

- Allen, B., R. Deneckere, T. Faith, D. Kovenock. 2000. Capacity precommitment as a barrier to entry: A Bertrand-Edgeworth approach. *Econom. Theory* 15(3) 501–530.
- Bernstein, J. 1992. Price margins and capital adjustment: Canadian mill products and pulp and paper industries. *Internat. J. Indust. Organ.* 10 491–510.
- Besanko, D., U. Doraszelski. 2004. Capacity dynamics and endogenous asymmetries in firm size. *RAND J. Econom.* 35(1) 23–49.
- Besanko, D., U. Doraszelski, Y. Kryukov, M. Satterthwaite. 2010. Learning-by-doing, organizational forgetting, and industry dynamics. *Econometrica* 78(2) 453–508.

- Booth, D., V. Kanetkar, D. Vertinsky, D. Whistler. 1991. An empirical model of capacity expansion and pricing in an oligopoly with barometric price leadership: A case study of the newsprint industry in North America. *Internat. Econom. Rev.* **39** 255–276.
- Borkovsky, R., U. Doraszelski, Y. Kryukov. 2010. A user's guide to solving dynamic stochastic games using the homotopy method. *Oper. Res.* **58**(4, Part 2 of 2) 1116–1132.
- Brock, W., J. Scheinkman. 1985. Price-setting supergames with capacity constraints. *Rev. Econom. Stud.* **52**(3) 371–382.
- Caballero, R., E. Engel. 1999. Explaining investment dynamics in U.S. manufacturing: A generalized (S, s) approach. *Econometrica* **67**(4) 783–826.
- Christensen, L., R. Caves. 1997. Cheap talk and investment rivalry in the pulp and paper industry. *J. Indust. Econom.* **45** 47–73.
- Compte, O., F. Jenny, P. Rey. 2002. Capacity constraints, mergers and collusion. *Eur. Econom. Rev.* **46**(1) 1–29.
- Deneckere, R., D. Kovenock. 1996. Bertrand-Edgeworth duopoly with unit cost asymmetry. *Econom. Theory* **8**(1) 1–25.
- Dixit, A. 1980. The role of investment in entry-deterrence. *Econom. J.* **90** 227–243.
- Doms, M., T. Dunne. 1998. Capital adjustment patterns in manufacturing plants. *Rev. Econom. Dynam.* **1**(2) 409–429.
- Doraszelski, U., M. Satterthwaite. 2010. Computable Markov-perfect industry dynamics. *RAND J. Econom.* **41**(2) 215–243.
- Ericson, R., A. Pakes. 1995. Markov-perfect industry dynamics: A framework for empirical work. *Rev. Econom. Stud.* **62** 53–82.
- Ghemawat, P. 1984. Capacity expansion in the titanium dioxide industry. *J. Indust. Econom.* **32** 145–163.
- Gilbert, R., M. Lieberman. 1987. Investment and coordination in oligopolistic industries. *RAND J. Econom.* **18**(1) 17–33.
- Hall, E. 1990. An analysis of preemptive behavior in the titanium dioxide industry. *Internat. J. Indust. Organ.* **8** 469–484.
- Kreps, D., J. Scheinkman. 1983. Quantity precommitment and Bertrand competition yield Cournot outcomes. *Bell J. Econom.* **14** 326–337.
- Lieberman, M. 1987. Strategies for capacity expansion. *Sloan Management Rev.* **28**(4) 19–27.
- Maggi, G. 1996. Strategic trade policies with endogenous mode of competition. *Amer. Econom. Rev.* **86**(1) 237–258.
- Pakes, A., P. McGuire. 1994. Computing Markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model. *RAND J. Econom.* **25**(4) 555–589.
- Spence, M. 1977. Entry, capacity, investment and oligopoly pricing. *Bell J. Econom.* **8**(2) 534–544.
- Staiger, R., F. Wolak. 1992. Collusive pricing with capacity constraints in the presence of demand uncertainty. *RAND J. Econom.* **23**(2) 203–220.
- Tirole, J. 1988. *The Theory of Industrial Organization*. MIT Press, Cambridge, MA.
- Wood, A. 2005. Investment interdependence and the coordination of lumpy investments: Evidence from the British brick industry. *Appl. Econom.* **37** 37–49.