Collusive Capacity*

Daehyun Kim† Rustin Partow‡

November 11, 2019

Abstract

It is widely believed that cartels with too many members are destined to fail. The standard argument is that as the number of cartel members increases, shares of collusive profit diminish relative to deviation profits. We show that this argument is built on unreasonable assumptions about plant capacity. We add plant capacity choices to an otherwise standard dynamic oligopoly game. We consider the unsophisticated and easily enforced strategy in which each firm simply chooses plant capacity equal to static Nash output. Our main result is that as the number of firms goes to infinity, the critical discount factor required to sustain collusion converges to less than 0.63. Thus, collusion is (quite) robust to the number of firms. This result applies to a broad class of demand functions and to both Cournot and Bertrand competition.

Keywords: Capacity constraints; Collusion; Repeated games.

JEL Classification Numbers: C72, C73, L13, L41.

*We thank Moshe Buchinsky, Maurizio Mazzocco, and Ichiro Obara for their support. For useful discussions and comments, we thank Simon Board, Elhadi Caoui, and Vitaly Titov. All errors are ours.

†Department of Mathematical Economics and Finance, Wuhan University. E-mail address: daehyunkim@whu.edu.cn

‡Department of Economics, UCLA. E-mail address: rustinpartow@gmail.com

1 Introduction

Economists widely believe that cartels with too many firms are doomed to fail. We argue that the standard intuition underlying this consensus is flawed because it ignores capacity constraints. In this paper, we consider the behavior of a cartel with arbitrarily many members, who endogenously choose plant capacity in addition to the usual quantity or price choices. We show that a relatively unsophisticated and easily enforced strategy concerning plant capacity will tend to ensure successful collusion regardless of the number of members.

In particular, our paper challenges what might be called the *splitting-of-the-pie* intuition, which is well explained in the following excerpt from a European Commission report (Ivaldi et al., 2007):

> Since firms must share the collusive profit, as the number of firms increases each firm gets a lower share of the pie. This has two implications. First, the gain from deviating increases for each firm since, by undercutting the collusive price, a firm can steal market shares from all its competitors; that is, having a smaller share each firm would gain more from capturing the entire market. Second, for each firm the long-term benefit of maintaining collusion is reduced, precisely because it gets a smaller share of the collusive profit. Thus the short-run gain from deviation increases, while at the same time the long-run benefit of maintaining collusion is reduced. It is thus more difficult to prevent firms from deviating.

This pie-splitting reasoning is commonly taught in undergraduate economics courses and appears to be the main reason why cartels with many members are expected to fail.[1] This same intuition (applied in reverse) suggests that a possible benefit of antitrust activity is to prevent collusion by reducing concentration. Theory plays an unusually large role in determining the factors that facilitate collusion because tacit (or sufficiently discrete) cartels are necessarily unobserved. The available evidence from sanctioned or criminally investigated cartels fails to establish a connection between market concentration and collusion, as discussed in Levenstein and Suslow (2006).[2]

The implicit assumption in the above excerpt from Ivaldi et al. (2007) is that capacity can be changed spontaneously. In the physical products industries that are typically modeled under Cournot competition, output can be easily varied at the margin, but expansions in plant capacity are slow and conspicuous. Any realistic theoretical rendition

[1]Another possible reason, which we will not challenge in this paper, pertains to the difficulty in monitoring.

[2]See table 4 on page 59 for a list of cartels and numbers of participants.

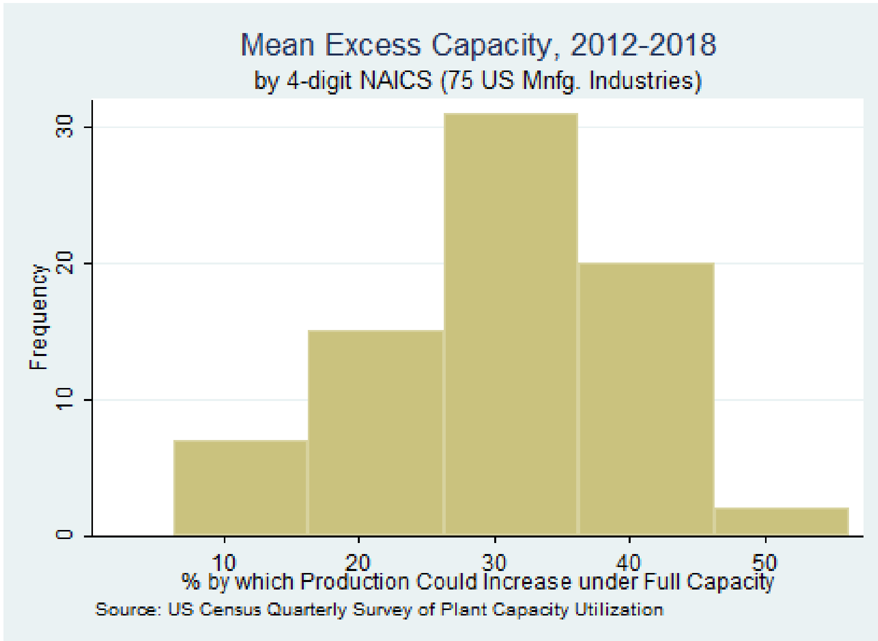


Figure 1

of a cartel defection ought not require installing large amounts of new capacity. In this paper, we claim to show that this is precisely what would needs to happen to force the unraveling of a cartel with many members.

Our basic argument can be easily understood through the mathematics of large cartels. Suppose that a small member of a large cartel deviates from a collusive regime for a one-time increase in output, causing the industry to permanently revert to a price war with zero profit.[3] Then letting r denote the interest rate, the cartel member's percent increase in output must be at least $\frac{1}{r}$. For reasonable interest rates, this means a 1,000% increase or more to make defection attraction.

Can a 1,000% increase in output occur without a conspicuous increase in capacity? The data on excess plant capacity suggest not. In Figure 1, we show the excess capacities for 75 different industries taken from the quarterly Survey of Plant Capacity Utilization conducted by the US Census Bureau. In the survey, individual plants report their current quarterly production, and the maximum output that they could feasible have produced in that quarter. Both numbers are aggregated across industries using the NAICS classification system, and the ratio of the two numbers gives an industry's capacity utilization.

According to Figure 1, excess capacity seldom averages more than 40% across an industry. Combining our brief mathematical example with the data gives some prima facie

[3]Here we assume that the firm's expansion in output by several multiples would have negligible effects on price-cost margins, due to its being small and having constant marginal costs.

evidence to suspect that by realistically modeling capacity constriants, the apparent feasibility of collusion, and its dependence on the number of cartel members, may change.

In a recent paper, Kühn (2012) considers a setting where a fixed industry capacity becomes increasingly fragmented after entry. He finds that collusion may become easier after the increase, because "Very fragmented markets allow for very severe punishments, while capacity constraints limit the incentives for deviation from a collusive price." Meanwhile, Brock and Scheinkman (1985) consider the case where *per-capita* capacity is constant, and find a non-monotonic relationship between the number of firms and the ease of collusion. In their setting, adding firms initially aids collusion by increasing the severity of punishments, but this effect eventually diminishes and is dominated by the dilution of collusive profits.

Our main contribution, relative to these papers, is to properly endogenize the choice of capacity as part of a collusive equilibrium. We present a two-stage game. In stage one, firms simultaneously choose costless and permanent capacity levels. In stage two, they play an infinitely repeated oligopoly game, and each firm's production is constrained by their previous capacity choice. We analyze a simple collusive regime in which the firms choose modest capacity levels and then repeatedly produce equal shares of the monopoly quantity. Defection in both stages is punished with infinite Nash reversion. The important feature of the stage-one capacity choice, which is simply the static stage-Nash level of output, is that it remains incentive compatible regardless of any stage one uncertainty about whether a cooperative or non-cooperative outcome will be played in stage two. We initially assume that capacity is free, and we later show that relaxing this assumption doesn't qualitatively change the results. We find support for Kuhn's assumption of fixed industry capacity. We then derive some practical results on cartel robustness by calculting asymptotic and uniform bounds on the critical discount factor necessary for collusion.

We proceed with a discount factor analysis that shows that collusive capacity makes collusion on output much easier than is suggested by models that abstract from capacity, and this holds uniformly with respect to the number of firms. As the number of firms goes to infinity, industry capacity converges to the constant perfectly competitive output level and monopolistic collusion can be achieved with a discount factor above 0.63. For any finite number of firms, we find that a discount factor of 0.732 is sufficient for a wide range of demand functions, 2/3 for concave demand functions, and 1/2 for linear demand functions.

These results suggest that capacity constraints can completely mitigate the number-of-firms effect discussed in Ivaldi et al. (2007). This is because a model with unlimited capacity constraints will eventually, as N becomes large, imply deviation payoffs that are unrealistically large when compared to, say, the capacity these firms would need to produce competitive output levels. It is still possible that collusion becomes harder as the number of firms increases, but it is probably because of monitoring difficulties and not because of the incentives effect that is typically emphasized by economists. There are a large number of unconcentrated industries with relatively strong monitoring devices such as trade associations and industry newsletters.[4]

Before Kühn (2012), several other papers also considered capacity investment in an oligopoly model. Davidson and Deneckere (1986) imagine the specific case where two cartel participants merge and insist that there shares of cartel profit be unchanged post-merger. If the cartel relies on a Nash reversion punishment, then the severity of the Nash punishment will decline, while some of the participants will still get their original shares, increasing their incentives to defect. Benoit and Krishna (1987) study a case similar to ours where firms in a duopoly make capacity decisions and compete a-la Bertrand (also see Kreps and Scheinkman (1983)). Their focus is to characterize shared properties of equilibria with above-Nash prices, and they show that excess capacity in every period is a necessity.[5]

Our paper essentially extends the Benoit and Krishna (1987) setting to incorporate arbitrarily many firms. The thrust of their results is that excess capacity plays a role as a "sheathed but visible sword". Our paper emphasizes that this is its *only* role. Even if capacity were free, the firms in a minimally sophisticated cartel would carry at most enough to inflict punishments–anything more than this would threaten to unravel collusion without offering any private benefits. To emphasize this we initially assume that capacity is costless. On the other hand, when capacity is costly, it effectively becomes a public good and the cartel's punishment protocol must disincentivize free-riding in the form of capacity underinvestment. We discuss costly capacity as an extension towards the end of the paper.

The rest of this paper is organized as follows. Section 2 describes the basic model. Section 3 establishes some preliminary results on the robustness of our collusive equilibrium to the number of firms. We then introduce the concept of ρ-concavity (see Anderson and Renault (2003)) and use it to establish stronger results on cartel robustness. Section 4 discusses several extensions, including Bertrand rather than Cournot competition, capacity

[4] Levenstein and Suslow (2006) find that trade associations are important to the duration of large cartels.

[5] There is also relatively large literature concerning the use of capacity constraints as a pre-commitment device (e.g., Spence (1977), Fudenberg and Tirole (1984), Bulow et al. (1985), Herk (1993), Spencer and Brander (1992)). In these settings, capacity can be used by an incumbent to deter entry, or by an entrant hoping that by setting aside enough residual demand for the incumbent, it will be induced to maintain high prices (the Judo economics technique).

adjustment costs, and entry. Section 5 concludes.

2 Dynamic Oligopoly with Irreversible Capacity Investment

Our model is a standard dynamic oligopoly (Cournot) model except that in period $t = 0$, each firm $i \in \mathcal{I} \equiv \{1, 2, \ldots, N\}$ simultaneously chooses a *capacity* $x_i \in \mathbb{R}_+$. An identical game is studied by Benoit and Krishna (1987). Capacity investment is costless and permanent. For each subsequent period $t \in \{1, 2, \ldots\}$, each firm i simultaneously chooses $q_i \in [0, x_i)$.[6]

Goods are perfectly substitutable and $P : \mathbb{R}_+ \to \mathbb{R}_+$ is the industry's inverse demand function. We make the following assumptions on the cost and inverse demand functions:

Assumption 1. *Marginal cost is a constant, $c \geq 0$.*

Assumption 2. *$P(\cdot)$ is twice continuously differentiable, with $P'(Q) < 0$ and $P''(Q)Q + P'(Q) < 0$ for all $Q \in \mathbb{R}_+$.*

Assumption 3. *$P(\bar{Q}) < c$ for some finite quantity $\bar{Q}$.*

For example, any strictly decreasing, strictly log-concave, and twice continuously differentiable inverse demand function would satisfy our assumptions. Assumption 2 is simply a regularity condition which is shown in Lemma 2 to be necessary to ensure that firm profit functions are strictly concave and submodular, and thus rule out the existence of multiple equillibria. Assumption 2 is weaker than strict concavity of P. The other two assumptions are standard assumptions made for tractability.

Each firm's choice of capacity and following outputs are perfectly publicly observed by the other firms. A firm's strategy is defined in the usual way: in period $t = 0$, each firm chooses the capacity; and in each period $t \geq 1$, each firm chooses output within its capacity level depending on the history, i.e., the capacity choice and past outputs by the all the firms up to the period. Each firm's payoff is discounted by a common discounting factor $\delta \in [0, 1)$. We use subgame perfection as our equilibrium concept, which requires that in every possible history, the continuation strategy profile constitutes a Nash equilibrium in the induced subgame.

[6]Note that firms cannot produce output in period 0, eliminating the need to analyze simultaneous quantity-capacity deviations. The assumption is purely cosmetic, as our equilibrium could easily incorporate a "feeling-out" stage where firms produce non-collusive output before reaching the collusive capacity level.

3 Cartel Robustness to the Number of Firms

Our main question is whether collusive profits can be achieved even as the number of firms goes to infinity if firms are allowed to collude in their capacity, and our analysis will specifically focus on the industry's ability to achieve *monopoly* profits.[7] In this section, we show that holding the industry price at the monopoly level (p^m) is robust for a large range of discount factors. Moreover, it can be achieved with a very simple equilibrium, which we will denote as Collusive Capacity.

Note that the definition of Nash reversion in our model accommodates the fact that feasible output levels are determined by the capacity decisions in period 0. The Nash equilibrium of the classic Cournot oligopoly game without capacity, which we will simply call classic Nash output, may not be feasible if some firms have chosen small capacity levels.[8]

We will now present results from a special equilibrium of the model where firms build up capacity levels equal to classic Cournot-Nash output levels, and subsequently produce collusive output. Both components of the equililbrium are supported with the threat of infinite Nash reversion following a defection. The ensuing results will show that this sequence of play is an equilibrium for a broad set of possible discount factors, demand functions, and numbers of firms, and that as the number of firms goes to infinity, there can be no other sequence of play that achieves monopoly profit for a lower discount factor. Our results are easily extended to Bertrand competition, although the emphasis is on Cournot where capacity constraints are typically more relevant.

Let $Q^m \in \mathbb{R}_+$ be the monopolist industry quantity, let Q_N^{ne} be the static Nash industry quantity with N firms, let Q^{pc} be the competitive or choke-quantity at which $P(Q^{pc}) = c$. Given N, let $\underline{\delta}(N)$ denote the critical level of δ for which the grim trigger strategy with Nash reversion achieves Q^m.

Proposition 1. *Under our assumptions,*

(a)

$$\underline{\delta} \equiv \lim_{N \to \infty} \underline{\delta}(N) = 1 - \frac{Q^m}{Q^{pc}} \leq 1 - \frac{1}{e} \approx 0.632. \tag{1}$$

In particular, if $P(Q)$ is concave, then $1 - \frac{Q^m}{Q^{pc}} \leq \frac{1}{2}$.

[7] This is in line with many papers on collusion, such as Fershtman and Pakes (2000), where firms are assumed only to collude when it is feasible to achieve monopoly profits.

[8] For example, on any path where all firms have chosen (weakly) less than the equilibrium capacity level, it is a dominant strategy Nash equilibrium of the ensuing repeated game for each firm to set output equal to capacity.

(b) *Uniform upper bound for* $\underline{\delta}(N)$*: Let* $z_N \equiv \frac{\pi_N^{ne}}{\pi^m}$*, and suppose it is bounded by some sequence* $\bar{z}_N \leq \bar{z}$*. For any* $N \geq 2$*,*

$$\underline{\delta}(N) \leq 1 - \frac{N}{N-1}\frac{1-\bar{z}_N}{e^{\frac{N-1}{N}} - \bar{z}_N} \leq 1 - \frac{1-\bar{z}}{e-\bar{z}}.$$

(c) *Finally, if demand is linear, i.e.,* $P(Q) = a - bQ$ *for some* $a, b > 0$*, then* $\underline{\delta}(N) = \frac{1}{2}$ *for all* $N \geq 2$*.*

Proof. See Appendix. □

Proposition 1 illustrates that Nash capacity makes collusion dramatically resilient to the number of firms. For example, according to our second result, if collusive profits are twice as high as Nash profits, then a discount factor of 0.775 is sufficient to sustain collusion for *any* number of firms. This corresponds to a maximum interest rate of about 29%–much higher than what is typically assumed in the literature. The underlying intuition for our results is simple. For large N, deviations are significantly constrained by collusive capacity, while punishments are unchanged relative to classic grim trigger collusive equilibria with infinite capacity. In the classic case, collusive profits go to zero at rate N while deviation profits remain strictly positive. In our equilibrium, deviation profits also go to 0 at *approximately* rate N.[9]

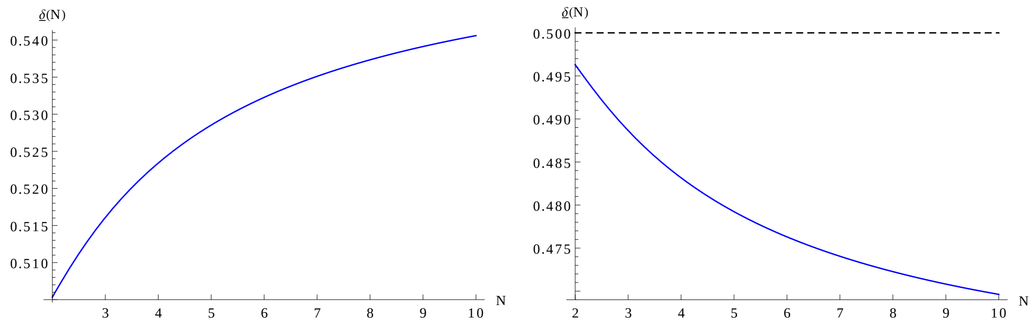
Our limiting result is quite simple to understand–as N tends to infinity, Nash profits as a ratio of collusive per-firm profits (which we denote as $\frac{1}{N}\Pi^m$) converge to 0, and deviation profits as a ratio of collusive profits converge to Q^{pc}/Q^m. The reason is that, in the limit, a firm's capacity becomes arbitrarily small so that, upon deviating, they have a negligible effect on the industry's price. Hence, the ratio of deviation to collusive payoffs is Q^{pc}/Q^m plus a term that is shrinking at rate N. We can write $\underline{\delta}(N) = \frac{\pi^d/\Pi^m - 1}{\pi^d/\Pi^m - \pi_N^{ne}/\Pi^m}$, and it should now be clear why the right-hand side converges to $\frac{Q^{pc}/Q^m - 1}{Q^{pc}/Q^m} = 1 - \frac{Q^m}{Q^{pc}}$. Lemma 4 establishes that this limit must be below $1 - \frac{1}{e} = .632$ or, under the stronger assumption that inverse demand is concave, below $\frac{1}{2}$.

It is particularly convenient that in the most widely studied case of linear demand, we get an exact result: a discount factor above $\frac{1}{2}$ is necessary and sufficient for collusion, regardless of the number of firms.

Example 1. Consider the following inverse demand function:

$$P(Q) = 1 - Q^{\alpha}$$

[9] This intuition is also owed to Kühn (2012).



(a) When $\alpha = 1/2$ **(b)** When $\alpha = 3/2$

and let $c = 0$. We will show that, depending on α, the critical discount factor required for our equilibrium can either increase or decrease with the number of firms. When $\alpha > 1$, P is strictly concave; while it is strictly convex when $\alpha < 1$. For any $\alpha \geq 0$, it satisfies Assumption 2.

One can show $Q_N^{ne} = \left(\frac{N}{\alpha+N}\right)^{\frac{1}{\alpha}}$ and $Q^m = \left(\frac{1}{\alpha+1}\right)^{\frac{1}{\alpha}}$; and $Q^d = \frac{N-1}{N}Q^m + \frac{1}{N}Q^{ne}$. Using these and from

$$\underline{\delta}(N) = \frac{\pi^d(N) - \frac{1}{N}\Pi^m}{\pi^d(N) - \frac{1}{N}\Pi^{ne}(N)}$$

we obtain different direction of δ as N increases depending on the size of α (see Figure 2a and Figure 2b).

□

Proposition 2. *Let $\underline{\underline{\delta}}(N)$ denote the minimum discount factor for which collusion can be achieved in any symmetric subgame perfect equilibrium, and let $\underline{\delta}(N)$ denote the analagous discount factor for our equilibrium.*

$$\lim_{N\to\infty} \underline{\delta}(N) = \lim_{N\to\infty} \underline{\underline{\delta}}(N) \tag{2}$$

Proof. See Appendix. □

Proposition 2 establishes a sense in which our proposed collusive capacity equilibrium is close-to-optimal. The reason this is not obvious is that in classic repeated Cournot games, *stick-and-carrot* strategies of the type considered in Abreu (1986) will generally support collusion for a strictly larger set of feasible discount rates than Nash reversion strategies. This suggests that they may also provide advantages in our setting. In the classic repeated Cournot oligopoly, a stick-and-carrot strategy's punishment entails a single period of very low or even negative profits, followed by a return to collusion. But to achieve these low initial profits, the stick-and-carrot strategy requires larger capacity than an infinite reversion strategy. If the industry had enough capacity to impose this stick-and-carrot punishment, then while the industry's punishment could be made more severe, deviation payoffs would also increase. We believe there are cases where this tradeoff favors Nash reversion, and other cases where it does not.

However, our proposition establishes that for large N, no symmetric SPE can sustain collusion for a larger set of discount factors. The intuition is simple. As N goes to infinity, the industry capacity converges to Q^{pc}, the perfectly competitive quantity, and the Nash reversion punishment payoff converges to 0, the most severe possible punishment. Suppose there exists another sequence of equilibria with a convergent industry quantity, $\bar{Q}$, with a strictly smaller limiting critical discount factor. The key to our proof is recognizing that $\bar{Q} \leq Q^{pc}$. The reason is that if $\bar{Q} > Q^{pc}$, the punishment payoff will still converge to a number no less than 0, as with our equilibrium, but the deviation payoff will be strictly larger than in our equilibrium, causing the limiting critical discount factor to be larger.

Given that $\bar{Q} \leq Q^{pc}$, as N goes to infinity, Nash reversion becomes an arbitrarily close-to-optimal punishment strategy, and producing at full capacity becomes an arbitrarily close-to-optimal deviation strategy. This allows for a convenient closed-form analysis of the asymptotic tradeoffs of capacity. The benefit of more capacity is a more severe punishment, and the cost is a larger deviation incentive. Asymptotically, this tradeoff always favors adding capacity (until capacity equals Q^{pc}).[10]

The downside of our analysis thus far is that it requires N going to infinity. We would like to establish claims that are independent of N. In the next sub-section, we will introduce the concept of ρ-concavity as a means of restricting and analyzing demand in order to obtain stronger results.

3.1 Demand Curvature and Robustness

We need to place modest restrictions on the curvature of demand in order to extend our results to broad classes of demand or numbers of firms. If demand is allowed to be highly concave in one region, and highly convex in another, then it is theoretically possible for deviation payoffs to be very large, and Nash reversion punishments very mild, for any finite number of firms. To tackle this issue formally, we use concept of ρ-concavity (convexity) that was first applied to the Cournot setting by Anderson and Renault (2003). Our preferred interpretation of ρ-concavity (convexity) is that it bounds the rate at which the inverse-elasticity of demand changes in the industry quantity.

[10] The reason is quite intuitive. For a very large number of firms, each with arbitrarily small capacities, a single firm's deviation from collusive output to producing at full capacity has a negligible effect on price. Thus, a one-percent increase in capacity raises deviation profits asymptotically by one-percent. Meanwhile, the elasticity of industry profits with respect to quantity must be strictly less than -1 for any quantity above the collusive level. Thus, a one-percent increase in capacity reduces punishment payoffs asymptotically by *more* than one-percent.

Definition 1. Consider a strictly positive function D with a convex domain $B \subseteq \mathbb{R}_+$. For $\rho \neq 0$, D is *ρ-concave* if for any $p, p' \in B$,

$$(D(\lambda p + (1-\lambda)p'))^\rho \geq \lambda D(p)^\rho + (1-\lambda)D(p')^\rho, \forall \lambda \in [0,1].$$

For $\rho = 0$, D is *0-concave* if

$$\ln D(\lambda p + (1-\lambda)p') \geq \lambda \ln D(p) + (1-\lambda)\ln D(p'), \forall \lambda \in [0,1].$$

A strictly positive function D with a convex domain is said to be *ρ-convex* if the direction of the inequality is opposite.

When $\rho = 0$, ρ-concave function is log concave. Also note that a linear function is 1-concave and 1-convex.

Lemma 1 (Anderson and Renault (2003)[11]). *The following hold:*

(a) D is ρ-concave if and only if

$$\psi(Q) \leq (1-\rho)$$

where $\psi(Q) \equiv -Q\frac{P''(Q)}{P'(Q)}$. Similarly, D is ρ-convex if and only if

$$(1-\rho) \leq \psi(Q).$$

(b) Let D be a strictly positive and decreasing function with a convex domain in $\mathbb{R}_+$. Then, there exists a pair of values in the extended real line, ρ' and ρ'' such that D is ρ'-concave and ρ''-convex. If D is ρ'-concave and ρ''-convex, then $\rho' \leq \rho''$.

We call $\psi(Q) := -Q\frac{P''(Q)}{P'(Q)}$ the *curvature* of the inverse demand. This is the elasticity of the slope of the inverse demand. The lemma says that ρ-concavity (resp. ρ-convexity) of a demand function provides an upper (resp. lower) bound of the curvature of its inverse demand.

The demand function in Example 1 has the constant curvature of $\psi(Q) = 1 - \alpha$ for all $Q \in [0,1]$. It is therefore both α-concave and α-convex.

Proposition 3 (Asymptotic bound). *If D is ρ'-concave and ρ''-convex, where $\rho'' \geq \rho' \geq -1$, then*

$$\xi(\rho'') \leq \lim_{N\to\infty} \delta(N) \leq \xi(\rho')$$

[11] This lemma is Proposition 1 and Claim 1 of Anderson and Renault (2003, p.255). For the proof, refer to the paper.

where

$$\xi(\rho) \equiv 1 - \left(\frac{1}{\rho+1}\right)^{\frac{1}{\rho}}. \tag{3}$$

In addition, $\xi(\rho)$ is decreasing in ρ and when $\rho \to \infty$, $\xi(\rho) \to 0$.

Proof. See Appendix A.2. □

Theorem 1 (Uniform Bound)**.** *For any $\rho' \geq -1$, there exists an upper bound for the critical discount factor that is strictly below 1, uniformly for all N.*

Proof. Let $\rho' \geq 1$ and choose $\epsilon > 0$ so that $\xi(\rho') + \epsilon < 1$ where $\xi(\rho')$ is the expression in (3). From Proposition 3, we know that there exists $\bar{N} \in \mathbb{N}$ such that for any $N \geq \bar{N}$,

$$\underline{\delta}(N) \leq \xi(\rho') + \epsilon.$$

Then, define

$$\bar{\delta} := \max\{\underline{\delta}(2), \cdots, \underline{\delta}(\bar{N}-1), \xi(\rho') + \epsilon\},$$

which is strictly smaller than 1. Then, by definition of the critical discount factor, if $\delta \geq \bar{\delta}$, for any number of firms we can sustain collusion. □

Also, if a demand function is ρ' concave, then it is also $\tilde{\rho}'$-concave for $\tilde{\rho}' \leq \rho'$; and similarly, if a demand function is ρ''-convex, then it is also $\tilde{\rho}''$-convex for any $\tilde{\rho}'' \geq \rho''$, the tight uniform upper bound for (ρ', ρ'') should be less $(\tilde{\rho}', \tilde{\rho}'')$.

Our bound is dependent on ρ' and ρ''. We have calculated it via numerical optimization, and illustrated the bound below. A large value of ρ' corresponds to making a strong lower-bound restriction on the degree of concavity of the demand curve, while a smaller value of ρ'' corresponds to a stronger upper-bound restriction on the degree of concavity. The stronger the restrictions, the smaller is our uniform bound.

Figure 3 shows our uniform bound as a function of ρ_1 and $\rho_2 - \rho_1$ (we choose the axes in this way to ensure a rectangular domain, since ρ_2 must be larger than ρ_1) when $\rho' = \rho_1$ and $\rho'' = \rho_2$.

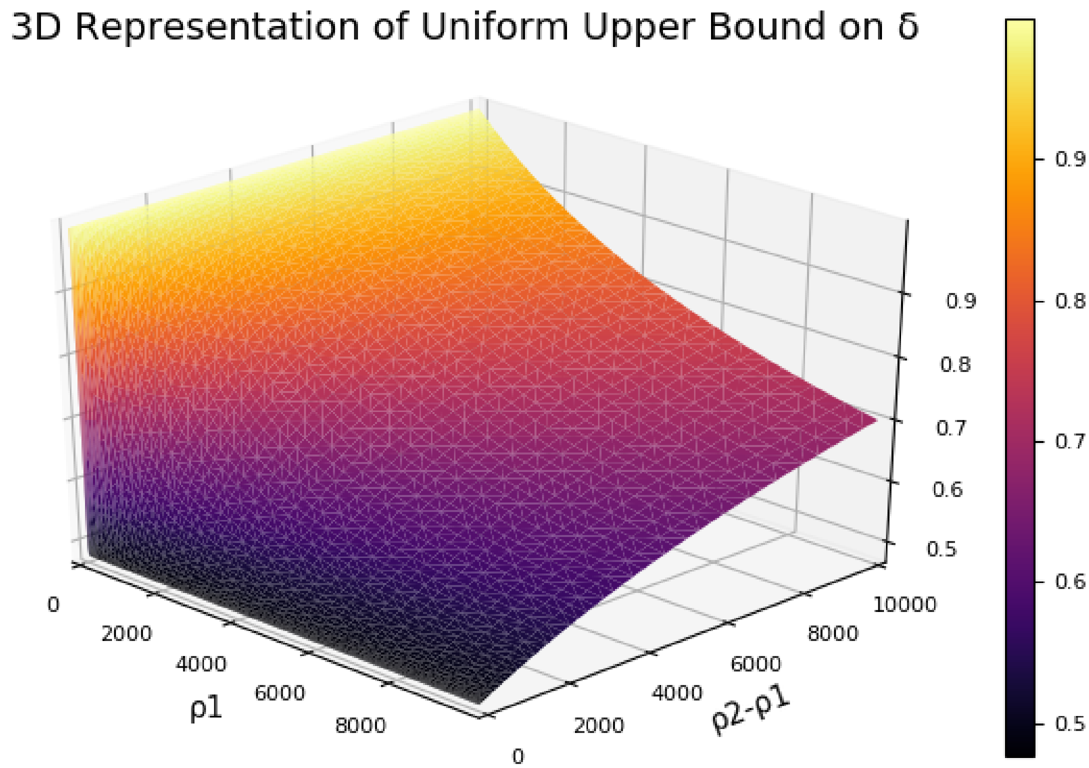
Corollary 1. *The ratio of excess (unused) capacity to total capacity under our equilibrium is*

$$\frac{\frac{1}{N}Q_N^{ne}}{\frac{1}{N}Q^m} = \frac{Q_N^{ne}}{Q^m}$$

and it increases in N. Also, for a demand function D which is ρ'-concave and ρ''-convex,

$$\left(\frac{N(\rho''+1)}{\rho''+N}\right)^{\frac{1}{\rho''}} \leq \frac{Q_N^{ne}}{Q^m} \leq \left(\frac{N(\rho'+1)}{\rho'+N}\right)^{\frac{1}{\rho'}}.$$

Figure 3



Hence, we provide a testable prediction that excess capacity rates should increase with the number of firms in an industry that is colluding on capacity. Having established these results for the basic Cournot model, the next section will consider several interesting extensions.

4 Extensions

4.1 Capacity Costs

Our job so far has been made relatively easy by the assumption of costless capacity. What happens when building capacity requires firms to incur sunk costs? In this section we will show that when the costs of building capacity are sufficiently small, our results are unchanged. We assume that at $t = 0$, when firms choose capacity x, they commit to paying periodic costs $(1-\delta)b(x)$, for some increasing function $b(\cdot)$. Since these costs are sunk in future periods, adding costs only affects the incentive compatibility of the first period of our strategy profile. Our concepts of monopoly and Nash output will still be based on the one-shot Cournot game without capacity.[12]

Proposition 4. *Suppose that $b(\cdot)$ and N satisfies $b\left(\frac{1}{N}Q_N^{ne}\right) \leq \frac{1}{N}(\Pi^m - \Pi_N^{ne})$. Then the Nash reversion strategy profile remains subgame perfect when there are N firms.*

Proof. See Appendix. □

Hence, as long as the per-period costs of each firm's capacity is less than per-period gains from collusion, the incentive compatibility of our equilibrium will not be affected. There is one desirable property of our equilibrium that we have not preserved in the above proof. Previously, our collusive capacity level was the same as what firms would have chosen, had they instead anticipated producing Nash output forever. Hence, it required minimal coordination. This is no longer the case in the equilibrium that we use in our proof. It is simple to understand why. Suppose that a firm expects its competitors to each produce $\frac{1}{N}Q_N^{ne}$. It will choose capacity x and then produce x forever. Its anticipated periodic profits are $x\left(P(\frac{N-1}{N}Q_N^{ne}x - c\right) - b(x)$. Hence, the new capacity cost term, $b(x)$, now enters into its stage one decision, and creates an incentive to produce less than Nash output. Hence, collusive capacity is no longer identical to non-cooperative capacity.

This property can be easily salvaged, at the cost of exposition, by changing the level of collusive capacity to equal the new level of non-cooperative capacity. The new level of non-cooperative capacity, $x^\star$, is the best-response capacity choice in stage one when each firm anticipates that the industry will subsequently produce at full capacity forever. This is exactly the same as the Nash output level corresponding to the static Cournot game when each firm's costs of producing q are $cq + b(x)$.This new level of capacity will be smaller than the one we consider above. In general, collusion via a grim trigger punishment, using this capacity level, will require a strictly higher discount factor than the capacity level that we use in our proof. However, when capacity costs are small, the differences in these capacity levels, and hence critical discount factors, will be negligible. Thus, we conclude that even with capacity costs, collusion can be sustained with arbitrarily many firms, and still requires minimal coordination on capacity.

4.2 Deterring Entry

We have argued that the purpose of excess capacity is to inflict punishments on defecting cartel members. Therefore, a colluding industry should have *only* enough excess capac-

[12] In principal, what we are referring to as monopoly output might no longer correspond to the output decisions of a forward-looking monopoly, who would consider both capacity costs and variable costs when determining its joint capacity and output strategy. However, if capacity costs are small, then so too would be the differences in these levels of output and profits.

ity for this purpose, which should correspond to the static Nash output level when the standard grim trigger punishment is being used. This overlooks another potential use for excess capacity: deterring potential entrants. From the perspective of the industry incumbents, it may be more profitable to compete and deter entry, than to accomodate entry and subsequently collude. Hence, it is important to check whether the capacity constraints prescribed in our equilibrium would make an industry more susceptible to entry. Thus, we extend the model to consider endogenous entry.

Suppose that, in addition to the original N firms in the market at stage one, there are infinitely more firms who can pay an entry cost $\kappa \in \mathbb{R}_+$ in order to enter the market at any period during the game, at which point they are in the market permanently. All firms in the market can make capacity and quantity decisions. However, capacity choices must be made, and observed, one period before the capacity becomes available to use.[13] Assume that firms compete a-la Cournot and that capacity costs are 0, as in the original model.

It is straightforward to devise a new equilibrium, similar in spirit to our original one, where collusion remains resilient to the number of firms. Let $r_t \in \mathbb{N}$ denote the total number of entrants observed up to period t. Our revised strategy hinges around a critical number of entrants, $R \in \mathbb{N}$. If strictly more than R firms choose to enter, the firms will produce the corresponding level of Nash output. In equilibrium, exactly R firms will enter, all in the first period, and the industry will collude at monopoly output. We will focus on the equilibrium of this type that delivers the largest profits to the incumbents–the one with the smallest possible number of entrants.

The equilibrium strategy is as follows: In any period where there are fewer than R total entrants, collude at monopoly output. If there are more than R entrants, or if anyone has defected in a previous stage, proceed to the punishment phase. The punishment phase involves producing the stage Nash equilibrium that corresponds to the total number of firms currently in the market, and their respective capacities. Let r_t be the number of entrants in period t. On the equilibrium path, each firm that is in the market chooses capacity equal to $\frac{1}{N+R} Q^{ne}_{N+R}$, and produces output equal to $\frac{1}{N+R} Q^m$. In the first period, R firms enter the market. In all subsequent periods, there is no entry. We call this collusive capacity with entry deterrence.

Proposition 5. *Suppose that in an industry with $N \in \mathbb{N}$ firms and a given demand function, collusive capacity without entry is an equilibrium for some discount factor δ. Then there exists some $R < \infty$ for which collusive capacity with entry deterrence is also an equilibrium (for the*

[13] Imagine, for example, the time it takes to build a new manufacturing plant. Notice that this implies that firms immediately make capacity decisions after entering, but their competitors observe their new capacity next period before they are able to start producing.

same δ). Moreover, the minimum possible value of R, $\underline{R}$, satisfies

$$\underline{R} = \min \left\{ R \in \mathbb{N} : \frac{1}{R+N+1} \Pi^{ne}_{R+N+1} \right\} \leq \kappa. \tag{4}$$

Moreover, $\underline{R}$ is exactly the same as the number of entrants in the equilibrium where all firms choose infinite capacity and subsequently produce static Nash output. Hence, we may conclude that collusive capacity does not come at the expense of entry deterrence.

Proof. See appendix. □

Our proposition shows that entry will occur up to the point at which entry costs are just below the Nash profits. This is exactly the same amount of entry that would be expected in a "competitive" market with unrestricted capacity. Hence, there seems to be little evidence that the capacity constraints we propose would undermine the industry's ability to deter entry. More importantly, we show that any such entry, if it occured, would *not* undermine the ability of the cartel to subsequently collude (with the new entrants as members).

4.3 Competition in Prices rather than Quantities

In this section, we modify the game so that firms compete choosing prices rather than quantitities. We assume the efficient rationing rule, which states that aggregate demand is always served first by firms offering the lowest prices, and is then split evenly in the case of ties.[14]

Proposition 6. *Suppose $N \geq 3$ and let $\underline{\delta}(N)$ denote the critical discount factor for which collusion can be sustained and let $\underline{\delta}$ denote its limit as N goes to infinity. Then there exists a strategy profile such that*

(a) $\underline{\delta}(N) = 1 - \frac{N-2}{N} \frac{Q^m}{Q^{pc}} \leq 0.877$

(b) $\underline{\delta} = 1 - \frac{Q^m}{Q^{pc}} \leq 1 - \frac{1}{e}$.

One such strategy profile is:

- *In period 0, each firm i chooses capacity* $x_i = \frac{Q^{pc}}{N-2}$.
- *In all periods $t \geq 1$, if no firm has deviated, each firm i sets* $p_i = p^m$.

[14] Specifying a rationing rule is necessary in settings of price competition with capactiy constraints because, in the event that capacity exceeds demand, payoffs will in general depend on how the scarce demand is allocated to different firms charging potentially different prices.

- *If there has been exactly one deviation, all firms set price equal to c.*

- *If there has been more than one deviation, play any stage Nash equilibrium (potentially in mixed strategies).*

Proof. See Appendix A.3. □

Hence, in a Bertrand setting, our results on capacity constraints are quite similar to our Cournot results.

Unlike the Cournot case, we require industry capacity to be $\frac{N}{N-2}$ times larger than the Nash output level (in this case, Q^{pc}). If any firm's capacity is "pivotal" to the industry's ability to produce the Nash punishment quantity, then that firm could profitably deviate from the punishment phase by raising price above marginal cost and serving a positive residual demand. This implies that at the end of the capacity selection stage, the industry capacity must exceed $\frac{N}{N-1}Q^{pc}$. To ensure that this holds even if one of the firms deviates on capacity selection, each firm must choose $\frac{1}{N-2}Q^{pc}$. These assumptions are similar to those needed in Kühn (2012) (see page 1,123).

This equilibrium is no longer robust to miscoordination where firms incorrectly anticipate competition at stage one. In such a case, a firm would choose to accumulate $\frac{1}{N}Q^{pc}$ instead of the prescribed level. However, this property can be easily recovered with a minor modification at the expense of exposition. Instead of reverting to Nash output following *any* deviation, we merely need to introduce some forgiveness with respect to deviations on capacity. Suppose that the industry ignores deviations on capacity when $\frac{1}{N}Q^{pc}$ is selected instead of $\frac{1}{N-2}Q^{pc}$. After such a deviation, the industry is still capable of implementing the Nash punishment, so the same proof regarding incentive compatibility applies.

5 Conclusion

In this paper, we provide theoretical evidence that collusion is more resilient to the number of firms than has been previously thought. In industries where capacity is public and costly to change, a large number of firms does not necessarily make it difficult to collude. On the other hand, if collusion really is harder with more firms, we believe that monitoring problems have more theoretical appeal than the incentive problems typically emphasized in the literature.

We see the most fruitful avenue for extending our work to be one that incorporates collusive capacity into an empirical framework such as that in Fershtman and Pakes

(2000). Their paper considers the consequences of collusion on technological investment, entry and exit, and welfare in a differentiated product Bertrand setting. Collusion in their model is dictates that, conditional on the industry state, firms produce monopoly output whenever doing so is supported by a grim-trigger punishment and otherwise play a stage Nash equilibrium. But a potential shortcoming of their paper is that investment decisions are made unilaterally. They do not consider how a (tacit) cartel would potentially benefit from collusive investment, also supported by grim-trigger punishment. Our paper provides some theoretical evidence that collusive investment strategies can qualitatively increase the scope for collusion while remaining very tractable.

A Appendix

A.1 Omitted Proofs

A.1.1 Useful Lemmas

Let $\pi(Q_i, q_i)$ denote the profits of firm i given the sum of its competitors quantities, $Q_i \in \mathbb{R}_+$, and its own quantity, $q_i \in \mathbb{R}_+$.

Lemma 2. *Under our Assumptions 1,2 and 3,*

1. *The profit-maximizing choice of q_i is submodular in Q_i*
2. *$\pi(Q_i, q_i)$ is a strictly concave and twice continuously differentiable function*
3. *For all $Q, q \geq 0$*

$$P''(Q+q)q + P'(Q) < 0.$$

Proof. The FOC of firm i's profit with respect to its own output is

$$\pi_2(Q_{-i}, q_i) = P'(Q_{-i} + q_i)q_i + P(Q_{-i} + q_i) - c = 0 \tag{5}$$

By differentiating the equation with respect to Q_{-i} yields

$$\pi_{21} + \pi_{22}\frac{dq_i}{dQ_{-i}} = 0. \tag{6}$$

Submodularity requires that $\frac{dq_i}{dQ_{-i}} < 0$, and concavity requires that $\pi_{22}(Q_{-i}, q_i) < 0$. Hence, (6) implies $\pi_{21}(Q_{-i}, q_i) < 0$. Derivation of (5) with respect to Q_i yields

$$\pi_{21} = P''(Q_{-i} + q_i)q_i + P'(Q_{-i} + q_i).$$

Since above we showed that $\pi_{21} < 0$, the first result follows. The second result follows immediately from the assumption that demand is decreasing in Q. $\square$

The remaining lemmas make the assumption that

$$P''(Q+q)q + P'(Q) < 0, \forall Q, q \geq 0$$

and $P'(Q) < 0, \forall x \geq 0$.

Lemma 3. *The Nash first-order condition (5) implies*

$$P(Q_N^{ne}) - c + \frac{1}{N} Q_N^{ne} P'(Q_N^{ne}) = 0, \forall N \in \mathbb{N}. \tag{7}$$

By totally differentiating (18) w.r.t. N,

$$P'(Q_N^{ne})\frac{dQ_N^{ne}}{dN} - \frac{1}{N^2}Q_N^{ne}P'(Q_N^{ne}) + \frac{1}{N}\frac{dQ_N^{ne}}{dN}P'(Q_N^{ne}) + \frac{1}{N}Q_N^{ne}P''(Q_N^{ne})\frac{dQ_N^{ne}}{dN} = 0$$

or

$$\frac{dQ_N^{ne}}{dN} = \frac{Q_N^{ne}P'(Q_N^{ne})}{N\left(NP'(Q_N^{ne}) + (P'(Q_N^{ne}) + Q_N^{ne}P''(Q_N^{ne})\right)}$$

Because we assume that $P' < 0$ and $P''(x)x + P'(x) < 0$, the above equation implies the following inequality:

$$0 < \frac{dQ_N^{ne}}{dN} < \frac{1}{N^2}Q_N^{ne}. \tag{8}$$

Lemma 4. *1. The industry Nash output Q_N^{ne} is increasing in N*

2. For all $K, N \in \mathbb{N}, K < N$

$$1 < \frac{Q_N^{ne}}{Q_K^{ne}} < e^{\frac{N-K}{NK}} < e,$$

and

$$\lim_{N\to\infty} P(Q_N^{ne}) = c$$

thus

$$\lim_{N\to\infty} \pi_N^{ne} \to 0.$$

Moreover, if $P(Q)$ is concave, then

$$1 < \frac{Q_N^{ne}}{Q_K^{ne}} \leq \frac{N}{N+1}\frac{K+1}{K}, \forall K, N \in \mathbb{N}, K < N.$$

Proof. The Nash first-order condition (5) implies

$$P(Q_N^{ne}) - c + \frac{1}{N} Q_N^{ne} P'(Q_N^{ne}) = 0, \forall N \in \mathbb{N}. \tag{9}$$

By totally differentiating (18) w.r.t. N,

$$P'(Q_N^{ne})\frac{dQ_N^{ne}}{dN} - \frac{1}{N^2} Q_N^{ne} P'(Q_N^{ne}) + \frac{1}{N}\frac{dQ_N^{ne}}{dN} P'(Q_N^{ne}) + \frac{1}{N} Q_N^{ne} P''(Q_N^{ne})\frac{dQ_N^{ne}}{dN} = 0$$

or

$$\frac{dQ_N^{ne}}{dN} = \frac{Q_N^{ne} P'(Q_N^{ne})}{N\left(NP'(Q_N^{ne}) + (P'(Q_N^{ne}) + Q_N^{ne} P''(Q_N^{ne})\right)}$$

Because we assume that $P' < 0$ and $P''(x)x + P'(x) < 0$, the above equation implies the following inequality:

$$0 < \frac{dQ_N^{ne}}{dN} < \frac{1}{N^2} Q_N^{ne}. \tag{10}$$

We can express the Nash quantity with the following integral, for $K < N$:

$$Q_N^{ne} = \exp\left(\ln Q_K^{ne} + \int_K^N \frac{d \ln Q_M^{ne}}{dM} dN\right) \tag{11}$$

$$= \exp\left(\ln Q_K^{ne} + \int_K^N \frac{d \ln Q_M^{ne}}{dM} dM\right) \tag{12}$$

$$< \exp\left(\ln Q_K^{ne} + \int_K^N \frac{1}{M^2} dM\right) \tag{13}$$

where the last inequality comes from (10) and solving this integral yields

$$1 < \frac{Q_N^{ne}}{Q_K^{ne}} < e^{\frac{N-K}{NK}} < e, \forall K, N \in \mathbb{N}, K < N.$$

Next, we will prove the limit result. Since Q_N^{ne} is increasing in N, as shown above, and is bounded above by $\bar{Q}$ (see Assumption 3), it must converge. Taking the limit of the Nash FOC in (18) as N goes to infinity, and noting that $P'(Q_N^{ne})$ must be bounded because $P(Q)$ was assumed to be twice continuously differentiable, we have

$$\lim_{N\to\infty} P(Q_N^{ne}) = c$$

Finally, we will derive the tighter bound on the growth of the Nash quantity with respect to N under the assumption that P is concave. From (10), we have

$$\frac{dQ_N^{ne}}{dN} = \frac{Q_N^{ne}}{N\left(\frac{P''(Q_N^{ne})}{P'(Q_N^{ne})}Q_N^{ne} + (N+1)\right)}$$
$$\leq \frac{Q_N^{ne}}{N(N+1)}$$

where the second inequality comes from $P''(Q) \leq 0$. Thus,

$$\frac{d \ln Q_N^{ne}}{dN} \leq \frac{1}{N(N+1)}.$$

Applying this to (11) and solving the integral yields

$$1 < \frac{Q_N^{ne}}{Q_K^{ne}} \leq \frac{N}{N+1}\frac{K+1}{K}.$$

□

Lemma 5. *When firms have classic Nash output as capacity, an optimal one-shot deviation from a collusive output level is to produce at capacity.*

Proof. When other firms produce $\frac{N-1}{N}Q_N^{ne}$ (equal shares of the industry Nash quantity), the optimal response is to produce $\frac{1}{N}Q_N^{ne}$ (this is the definition of a symmetric Nash equilibrium). Under collusion, firms produce strictly less than equal shares of Nash output, so the unconstrained best response is to produce strictly more than $\frac{1}{N}Q_N^{ne}$. (2) also showed that our assumptions guarantee that profits are concave in a firm's own quantity, which implies that when the unconstrained best response would involve producing above capacity, the constrained best response must be to produce at capacity, which equals $\frac{1}{N}Q_N^{ne}$. □

Lemma 6. *Suppose $Q > Q^m$. Then $P(Q) - c \leq (P^m - c)\frac{Q}{Q^m}$.*

Proof. The elasticity of price-cost margins with respect to quantity must be larger in absolute value than 1 for all quantities above the monopoly quantity (otherwise we do not have a unique monopoly quantity, and we require demand functions that pose unique Nash equilibria for all numbers of players $N = 1, ..., \infty$).

Hence,

$$\ln(P(Q) - c) \leq \ln(P^m - c) - (\ln Q - \ln Q^m)$$

Thus,

$$\ln(P(Q) - c) \leq (P^m - c)\frac{Q^m}{Q}$$

□

A.1.2 Proof of Proposition 1

Proof. By Lemma 2,

$$P''(Q+q)q + P'(Q) < 0, \forall Q, q \geq 0$$

and $P'(Q) < 0, \forall x \geq 0$. Under collusion with Nash capacity, the minimal value of δ, $\underline{\delta}$, for which this is an equilibrium satisfies

$$\underline{\delta}(N) = \frac{\pi_N^d - \frac{1}{N}\Pi^c}{\pi_N^d - \frac{1}{N}\Pi_N^{ne}} = \frac{N\pi_N^d - \Pi^c}{N\pi_N^d - \Pi_N^{ne}}$$

where $\pi^d(N)$ denotes an optimal deviation when $N-1$ other players are producing equal shares of collusive output, Π^c denotes industry collusive profits, and Π^{ne} denotes industry Nash profits. The collusive output level is assumed to remain fixed as N varies. Hence, Q^m must be uniformly smaller than Q_N^{ne} for all N. In other words, the maximal possible value of Q^m for this exercise is Q^{pc} (we will show that Q_N^{ne} increases in N and converges to Q^{pc}) and the minimal possible value is Q^m.

First, we will derive the limiting value of $\underline{\delta}(N)$, as N goes to infinity.

$$\lim_{N\to\infty} \underline{\delta}(N) = \frac{\lim_{N\to\infty} N\pi_N^d - \Pi^c}{\lim_{N\to\infty} N\pi_N^d - \lim_{N\to\infty}\Pi_N^{ne}}.$$

By Lemma 5, an optimal deviation is to produce $\frac{1}{N}Q_N^{ne}$. Hence,

$$\lim_{N\to\infty} N\pi^d(N) = \lim_{N\to\infty} N\frac{Q_N^{ne}}{N}\left(P\left(\frac{N-1}{N}Q^c + \frac{1}{N}Q_N^{ne}\right) - c\right). \tag{14}$$

By Lemma 4,

$$\begin{aligned}\lim_{N\to\infty} N\pi^d(N) &= \left(\lim_{N\to\infty} Q_N^{ne}\right)(P(Q^c) - c)\\ &= Q^{pc}\left(P(Q^c) - c\right).\end{aligned}$$

Also, industry Nash profits converge to 0,[15]

$$\lim_{N\to\infty} N\pi_N^{ne} = \lim_{N\to\infty} Q_N^{ne}(p(Q_N^{ne}) - c) = Q^{pc}(p(Q^{pc}) - c) = 0$$

[15]Note the importance here of constant marginal cost.

Finally, note that $N\Pi^m = Q^m\left(P(Q^m) - c\right)$. Substituting in these results, we have

$$\lim_{N\to\infty} \delta(N) = \frac{Q^{pc}\left(P(Q^c) - c\right) - Q^m\left(P(Q^m) - c\right)}{Q^{pc}\left(P(Q^c) - c\right)} = 1 - \frac{Q^m}{Q^{pc}} \leq 1 - \frac{1}{e}$$

where the last inequality comes from Lemma 4.

Second, we will prove the uniform bound result.

$$\underline{\delta}(N) = \frac{\pi_N^d/\pi^m - 1}{\pi_N^d/\pi^m - \pi_N^{ne}/\pi^m}. \tag{15}$$

The numerator is smaller than the denominator. Hence, applying the quotient rule reveals that the expression must be increasing in π_N^d/π^m. The expression is also obviously increasing in π_N^{ne}/π^m. Thus, to obtain an upper-bound for $\underline{\delta}(N)$, it is sufficient to replace π_N^d/π^m and π_N^{ne}/π^m with upper-bounds.

Let $\pi^d = Q_N^{ne}\left(P(Q^d) - c\right)$, and write $Q^d = \lambda Q_N^{ne} + (1-\lambda)Q^m$, where $\lambda = \frac{1}{N}$. Then

$$\begin{aligned}
\pi^d &= Q^d\left(P(Q^d) - c\right) + \left(Q_N^{ne} - Q^d\right)\left(P(Q^d) - c\right) \\
&= \pi(Q^d) + \left(Q_N^{ne} - Q^d\right)\left(P(Q^d) - c\right).
\end{aligned}$$

Because $\pi(Q)$, the industry profit under homogenous output, is strictly concave, we know that $\pi(Q^d) = \pi\left(\lambda Q_N^{ne} + (1-\lambda)Q^m\right) > \lambda\pi(Q_N^{ne}) + (1-\lambda)\pi^m$. Hence,

$$\begin{aligned}
\pi^d &< \lambda\pi_N^{ne} + (1-\lambda)\pi^m + \left(Q_N^{ne} - Q^d\right)\left(P(Q^d) - c\right) \\
&= \lambda\pi_N^{ne} + (1-\lambda)\pi^m + \left(Q_N^{ne} - (\lambda Q_N^{ne} + (1-\lambda)Q^m)\right)\left(P(Q^d) - c\right) \\
&= \lambda\pi_N^{ne} + (1-\lambda)\pi^m + (1-\lambda)\left(Q_N^{ne} - Q^m\right)\left(P(Q^d) - c\right) \\
&< \lambda\pi_N^{ne} + (1-\lambda)\pi^m + (1-\lambda)\left(Q_N^{ne} - Q^m\right)\left(P^m - c\right)
\end{aligned}$$

From Lemma 4, we know that $Q_N^{ne} < e^{\frac{N-1}{N}}Q^m$. Hence,

$$\begin{aligned}
\pi^d &< \lambda\pi_N^{ne} + (1-\lambda)\pi^m + (1-\lambda)\left(e^{\frac{N-1}{N}} - 1\right)Q^m\left(P^m - c\right) \\
&= \lambda\pi_N^{ne} + (1-\lambda)\pi^m + (1-\lambda)\left(e^{\frac{N-1}{N}} - 1\right)\pi^m,
\end{aligned}$$

which upon dividing by π^m implies

$$
\begin{aligned}
\frac{\pi^d}{\pi^m} &< \lambda \frac{\pi_N^{ne}}{\pi^m} + (1-\lambda) + (1-\lambda)\left(e^{\frac{N-1}{N}} - 1\right) \\
&= \lambda \frac{\pi_N^{ne}}{\pi^m} + (1-\lambda)\left(e^{\frac{N-1}{N}}\right).
\end{aligned}
$$

Plugging this into (15) and letting $z \equiv \frac{\pi_N^{ne}}{\pi^m}$, we have

$$
\begin{aligned}
\underline{\delta}(N) = \frac{\frac{\pi_N^d}{\pi^m} - 1}{\frac{\pi_N^d}{\pi^m} - \frac{\pi_N^{ne}}{\pi^m}} &< \frac{\lambda \frac{\pi_N^{ne}}{\pi^m} + (1-\lambda)\left(e^{\frac{N-1}{N}}\right) - 1}{\lambda \frac{\pi_N^{ne}}{\pi^m} + (1-\lambda)\left(e^{\frac{N-1}{N}}\right) - \frac{\pi_N^{ne}}{\pi^m}} \\
&= \frac{\frac{\lambda}{1-\lambda} z + e^{\frac{N-1}{N}} - \frac{1}{1-\lambda}}{e^{\frac{N-1}{N}} - z} \\
&= 1 - \frac{1}{1-\lambda} \frac{1-z}{e^{\frac{N-1}{N}} - z}.
\end{aligned}
$$

Replacing $\lambda = \frac{1}{N}$ and rearranging terms, we have

$$
\underline{\delta}(N) \leq 1 - \frac{N}{N-1} \frac{1-z}{e^{\frac{N-1}{N}} - z} \leq 1 - \frac{1-z}{e-z}. \tag{16}
$$

□

A.1.3 Proof of Proposition 2

Proof. For each $N \in \mathbb{N}$, let $Q(N)$ denote an industry capacity under which $\underline{\underline{\delta}}(N)$ can be achieved. Because the equilibrium is assumed to be symmetric, each firm's capacity is $Q(N)/N$.

Claim 1. *For any sequence of equilibria whose limiting critical discount rate is less than 1, there exists a finite $\bar{N} \in \mathbb{N}$ such that for $N \geq \bar{N}$,*

$$
\frac{Q(N)}{N} = \underset{q_i \in \left[0, \frac{Q(N)}{N}\right]}{\arg\max}\; q_i \left(P\left(q_i + \frac{N-1}{N} Q^m\right) - c\right).
$$

In addition, $\bar{Q}(N)$ is uniformly bounded over N.

In words, when there are sufficiently large number of firms the optimal deviation involves producing at capacity.

Proof. Such an $\bar{N}$ must exist because without binding capacity constraints, the critical discount rate must go to 1, whereas we proved in Proposition 1 that with binding capacity constraints, a critical discount rate less than 1 is achievable.

Also,

$$\frac{1}{N}\bar{Q}(N) \to 0.$$

Note that

$$\underline{\underline{\delta}}(N) \geq \frac{\bar{Q}(N)(P(\frac{1}{N}\bar{Q}(N) + \frac{N-1}{N}Q^m) - c) - Q^m(P(Q^m) - c)}{\bar{Q}(N)(P(\frac{1}{N}\bar{Q}(N) + \frac{N-1}{N}Q^m) - c) - 0}$$

If $\bar{Q}(N)$ is not uniformly bounded the RHS is arbitrarily close to 1. □

Assume that $Q(N)$ converges to $\bar{Q} < \infty$.[16] We henceforth assume that $N \geq \bar{N}$ where $\bar{N}$ satisfies Claim 1.

We will now show that $\bar{Q} \leq Q^{pc}$. Suppose instead that $\bar{Q} > Q^{pc}$. Note that the limiting punishment payoff (times N) is $\Pi^p \geq 0$. The limiting deviation payoff (times N) is $\bar{Q}\,(P(Q^m) - c)$.[17] However, by reducing capacity to Q^{pc} and playing a Nash reversion strategy, we can always achieve a limiting punishment payoff of 0, but a strictly lower limiting deviation payoff (times N) of $Q^{pc}\,(P(Q^m) - c)$. Thus, $\bar{Q} \leq Q^{pc}$.

Having established that $\bar{Q} \leq Q^{pc}$, and therefore $\bar{Q}(N) \leq Q^{pc} + \kappa(N)$ for some sequence $\kappa(N) \to 0$ and $\kappa(N) > 0$, we will explain why the punishment payoff for any alternative sequence of collusive equilibria achieves payoffs that are *no larger* than what would be achieved if the industry produced at capacity forever ("capacity reversion", for short). This will justify analyzing the alternative equilibria *as if* they used a simple capacity reversion punishment.

Recall that

$$\underline{\underline{\delta}}(N) = \frac{\pi_N^d - \pi^m}{\pi_N^d - \pi_N^p} = \frac{\Pi_N^d - \Pi^m}{\Pi_N^d - \Pi_N^p}$$

where Π_N^p is the lowest strongly symmetric equilibrium payoff.

[16] This is without loss in generality. To achieve $\underline{\underline{\delta}}(N)$, capacity must eventually become bounded. Thus, $Q(N)$ is bounded and by Bolzano-Weirstrass, has a convergent subsequence. Hence, if $Q(N)$ did not converge, we could always consider instead a convergent subsequence. The limiting discount factor of that subsequence must also lie below the limiting discount factor of our equilibrium, otherwise the supposition that the original sequence achieved a critical discount rate below $\underline{\underline{\delta}}(N)$ is wrong. Thus, it is sufficient to prove that no convergent subsequence is asymptotically *better* than our equilibrium sequence.

[17] As long as industry capacity is bounded (by assumption), in the limit, an individual firm's deviation has no effect on price.

Claim 2. *For each $N \in \mathbb{N}$,*

$$\Pi_N^p \equiv N\pi_N^p \geq Q(N)(P(Q(N)) - c).$$

Namely, the industry profit from the worst punishment industry is greater than that from "capacity reversion."

Proof. In any symmetric equilibrium, where each firm produces the same output, in any period involving punishment a firm's profit is $\frac{1}{N}Q(P(Q) - c)$ for some $Q \geq 0$.

First we claim that in any punishment involving the industry output greater than Q^m, the harshest punishment is achieved when the industry output is $Q(N)$. This is simply because for $Q > Q^m$, the industry profit $Q(P(Q) - c)$ is decreasing.

Given this, it is sufficient for the result to show that any punishment that involves the industry output less Q^m in one or more periods cannot be the harshest punishment. This is because, deviation from these punishment periods would be *strictly more* profitable than deviation from collusive output itself. Formally, consider firm i's deviation and suppose $Q_{-i} < \frac{N-1}{N}Q^m$, then,

$$\begin{aligned}\max_{q_i \in \left[0, \frac{1}{N}Q(N)\right]} q_i(P(q_i + Q_{-i}) - c) &\geq \frac{1}{N}Q^m \left(P\left(\frac{1}{N}Q^m + Q_{-i}\right) - c\right) \\ &> \frac{1}{N}Q^m P((Q^m) - c).\end{aligned}$$

□

We should emphasize that "capacity reversion" may not actually be subgame perfect (it is certainly not when $Q(N) > Q_N^{ne}$). Our purpose is merely to provide a lower bound for the worst possible punishment payoff that *could* be subgame perfect.

From the above discussion, we can proceed to consider alternative sequences of equilibria achieving critical discount rate sequence $\underline{\underline{\delta}}(N)$ which (1) have asymptotic industry capacity $\bar{Q} \in [Q^m, Q^{pc}]$, (2) use capacity reversion as punishment, and (3) where an optimal deviation is to produce at capacity.

By Claim 1, the deviation payoff satisfies

$$\Pi^d = \lim_{N\to\infty} N\pi_N^d = \bar{Q}\left(P(Q^m) - c\right).$$

And by Claim 2, the punishment payoff, π_N^p, satisfies

$$\Pi^p = \lim_{N\to\infty} N\pi_N^p \geq \bar{Q}\left(P(\bar{Q}) - c\right).$$

Thus,

$$\underline{\underline{\delta}} \equiv \lim_{N\to\infty} \underline{\underline{\delta}}(N) = \frac{\bar{Q}(P(Q^m)-c) - Q^m(P(Q^m)-c)}{\bar{Q}(P(Q^m)-c) - \Pi^p}$$
$$\geq \delta(\bar{Q}) \equiv \frac{\bar{Q}(P(Q^m)-c) - Q^m(P(Q^m)-c)}{\bar{Q}(P(Q^m)-c) - \bar{Q}(P(\bar{Q})-c)}. \tag{17}$$

Because the critical discount rate is increasing in the punishment payoff, a lower bound is formed by replacing the punishment payoff with its upper bound that was given above.

Since the Nash reversion equilibrium we construct yields the critical discounting factor of $\delta(Q^{pc})$, it is sufficient for the conclusion to show that $\delta(\bar{Q})$ is indeed decreasing over $[Q^m, Q^{pc}]$. Although the direction of $\delta(\bar{Q})$ is *a priori* not obvious (note that there is a trade-off of having a larger capacity; i.e., harsher punishment and larger deviation gain), the following claim shows that this is the case.

Claim 3. *$\delta(\bar{Q})$ in (17) is decreasing over $[Q^m, Q^{pc}]$.*

Proof. Let $\tilde{P}(Q) \equiv P(Q) - c$. Then,

$$\delta'(x) = \frac{\tilde{P}(Q^m)(x\tilde{P}(Q^m) - x\tilde{P}(x)) - (\tilde{P}(Q^m) - \tilde{P}(x) - x\tilde{P}'(x))(x\tilde{P}(Q^m) - Q^m\tilde{P}(Q^m))}{(x\tilde{P}(Q^m) - x\tilde{P}(x))^2}.$$

Let

$$\eta(x) \equiv \tilde{P}(Q^m)(x\tilde{P}(Q^m) - x\tilde{P}(x)) - (\tilde{P}(Q^m) - \tilde{P}(x) - x\tilde{P}'(x))(x\tilde{P}(Q^m) - Q^m\tilde{P}(Q^m))$$

Note that at Q^m, $\eta(x)$ is 0; so is $\delta'(x)$. Thus, to show $\delta''(x) \leq 0$ for all $x \geq Q^m$, it is sufficient that $\eta(x)$ is decreasing.

$$\begin{aligned}
\frac{d\eta(x)}{dx} &= \tilde{P}(Q^m)(\tilde{P}(Q^m) - \tilde{P}(x) - x\tilde{P}'(x)) \\
&\quad - (-\tilde{P}'(x) - x\tilde{P}''(x))(x\tilde{P}(Q^m) - Q^m\tilde{P}(Q^m)) - (\tilde{P}(Q^m) - \tilde{P}(x) - x\tilde{P}'(x))\tilde{P}(Q^m) \\
&= -(-\tilde{P}'(x) - x\tilde{P}''(x))(x\tilde{P}(Q^m) - Q^m\tilde{P}(Q^m)) \\
&\leq 0
\end{aligned}$$

where the last inequality comes from Assumption 2. □

Thus, the derivative is non-positive over the entire interval $[Q^m, Q^{pc}]$, and we can conclude that $\delta(\bar{Q}) < \delta(Q^{pc})$. Hence, no other sequence of equilibria can admit a broader set of discount factors than ours, asymptotically.

□

A.1.4 Proof of Proposition 5

Proof. We want to solve for the minimum value of R that makes this an equilibrium. If the $(R+1)^{\text{th}}$ firm enters the market, then its competitors will revert to the non-cooperative outcome: producing $N+R+1$-firm Nash output, $\frac{1}{N+R+1}Q^{ne}_{N+R+1}$. Anticipating this, a best response is for these entrants is to respond by choosing capacities and subsequent outputs of $\frac{1}{N+R+1}Q^{ne}_{N+R+1}$, obtaining profits of $\frac{1}{N+R+1}\pi^{ne}_{N+R+1} - \kappa$. The ex-ante profits of this non-cooperative outcome, from the perspective of an entrant, is decreasing in R, holding N fixed. Thus, the minimum value of R is the smallest number for which the non-cooperative outcome with $R+1$ entrants is not profitable. I.e., we need to find

$$\min\left\{R \in \mathbb{N} : \frac{1}{R+N+1}\Pi^{ne}_{R+N+1}\right\} \leq \kappa.$$

Because Nash profits converge to 0 with N (Lemma 4), as long as κ is strictly positive, there must be some $R < \infty$ for which this holds. In periods 2 and after, one can simply modify all of our previous results on N firms to account for $R+N+1$ firms. We can use the same uniform bound for δ that was previously obtained in Proposition 1. Note that the resulting bound is even smaller than that using the original number of firms, N.[18]

Finally, we will prove the last statement regarding the number of firms that would enter when all firms are choosing infinite capacity and subsequently competing by producing static Nash output. The proof is trivial. First, assume that firm capacity and output strategies do not depend on the actions of other firms. Then clearly, conditional on the number of entrants in the first period, it is subgame perfect for each individual firm to choose infinite capacity and to produce the static Nash output level (by definition of Nash equilibrium) in each period. In each period, it must be unprofitable for an additional firm to enter the market. Consider the last period of entry, T (readers may anticipate that actually, all entry must occur in period 1, but this is not necessary for the proof). The entering firm will pay period cost κ forever, and earn period profits of $\frac{1}{R+N+1}\pi^{n}_{R+N+1}$ forever. Hence, the smallest possible number of entrants satisfies $\underline{R} = \min\{R : \frac{1}{R_T+N+1}\pi^{n}_{R_T+N+1}\}$. $\square$

A.1.5 Proof of Proposition 4

Proof. First, note that setting deviant capacity, $\tilde{x}$, larger than the prescribed Nash capacity is clearly never profitable since it involves greater spending on capacity and worse profits in the subsequent supergame. Assume that deviations at future stages are unprofitable

[18] This does not necessarily mean that entry actually makes collusion easier, because we haven't identified the exact critical discount factor for arbitrary demand functions.

so that we can focus on how the introduction of costs affects the robustness of our results. A deviation to a capacity of $x < \frac{1}{N}Q_N^{ne}$ produces the following per-period profit forever:

$$\tilde{\pi}(x) \equiv -b(x) + x\left(P\left(\frac{N-1}{N}Q_N^{ne} + x\right) - c\right).$$

If the firm conforms to the equilibrium capacity, it gets

$$-b\left(\frac{1}{N}Q_N^{ne}\right) + \frac{1}{N}Q^m P(Q^m).$$

The gains to deviation are (in per-period terms)

$$\begin{aligned}\tilde{\pi}(x) - \left(\frac{1}{N}\Pi^m - b\left(\frac{1}{N}Q_N^{ne}\right)\right) &= b\left(\frac{1}{N}Q_N^{ne}\right) - b(x) + x\left(P\left(\frac{N-1}{N}Q_N^{ne} + x\right) - c\right) - \frac{1}{N}\Pi^m \\ &\leq b\left(\frac{1}{N}Q_N^{ne}\right) + \frac{1}{N}\Pi_N^{ne} - \frac{1}{N}\Pi^m\end{aligned}$$

where the last inequality comes from the definition of Nash equilibrium, i.e., the term $x\left(P\left(\frac{N-1}{N}Q_N^{ne} + x\right) - c\right)$ is maximized at $x = \frac{1}{N}Q_N^{ne}$, achieving the value of $\frac{1}{N}\Pi_N^{ne}$. A sufficient condition for deviations to be unprofitable is, therefore,

$$b\left(\frac{1}{N}Q_N^{ne}\right) \leq \frac{1}{N}(\Pi^m - \Pi_N^{ne}).$$

□

A.2 Proof of Proposition 3

The proof follows from the sequence of lemmas given below.

Lemma 7. *For any $N \in \mathbb{N}$,*

$$\frac{d\ln\Pi(Q_N^{ne})}{dN} = -(N-1)\frac{d\ln Q_N^{ne}}{dN}.$$

Proof.

$$\ln\Pi(Q_N^{ne}) = \ln Q_N^{ne} + \ln(P(Q_N^{ne}) - c)$$

and

$$\frac{d\ln\Pi(Q_N^{ne})}{dN} = \frac{dQ_N^{ne}}{dN}\left[\frac{1}{Q_N^{ne}} + \frac{1}{P(Q_N^{ne}) - c}P'(Q_N^{ne})\right]$$

From the FOCs for the Nash equilibrium, we have

$$P(Q_N^{ne}) - c + \frac{1}{N} Q_N^{ne} P'(Q_N^{ne}) = 0$$

$$\iff 1 + \frac{1}{N} Q_N^{ne} \frac{P'(Q_N^{ne})}{P(Q_N^{ne}) - c} = 0$$

$$\frac{P'(Q_N^{ne})}{P(Q_N^{ne}) - c} = -\frac{N}{Q_N^{ne}}$$

Thus,

$$\begin{aligned}\frac{d \ln \Pi(Q_N^{ne})}{dN} &= \frac{1}{Q_N^{ne}} \frac{dQ_N^{ne}}{dN}(1 - N) \\ &= -(N-1)\frac{d \ln Q_N^{ne}}{dN}.\end{aligned}$$

□

Lemma 8. *For any* $N \in \mathbb{N}$*,*

$$\frac{d \ln Q_N^{ne}}{dN} = \frac{1}{N\left(\psi(Q_N^{ne}) + N + 1\right)}.$$

Proof. The Nash first-order conditions for each firm imply

$$P(Q_N^{ne}) - c + \frac{1}{N} Q_N^{ne} P'(Q_N^{ne}) = 0, \forall N \in \mathbb{N}. \tag{18}$$

By totally differentiating (18) with respect to N, we get

$$P'(Q_N^{ne})\frac{dQ_N^{ne}}{dN} - \frac{1}{N^2} Q_N^{ne} P'(Q_N^{ne}) + \frac{1}{N}\frac{dQ_N^{ne}}{dN} P'(Q_N^{ne}) + \frac{1}{N} Q_N^{ne} P''(Q_N^{ne})\frac{dQ_N^{ne}}{dN} = 0$$

or

$$\frac{dQ_N^{ne}}{dN} = \frac{Q_N^{ne} P'(Q_N^{ne})}{N\left(NP'(Q_N^{ne}) + (P'(Q_N^{ne}) + Q_N^{ne} P''(Q_N^{ne})\right)}.$$

Dividing both sides by $Q_N^{ne} P'(Q_N^{ne})$ and applying the definition $\psi(Q_N^{ne}) = Q_N^{ne} \frac{P''(Q_N^{ne})}{P'(Q_N^{ne})}$ yields

$$\frac{d \ln Q_N^{ne}}{dN} = \frac{1}{N\left(N + 1 + \psi(Q_N^{ne})\right)}.$$

□

Lemma 9. *For any $N \in \mathbb{N}$,*

$$\frac{d \ln \Pi(Q_N^{ne})}{dN} = -\frac{N-1}{N} \frac{1}{\psi(Q_N^{ne}) + N + 1}.$$

Proof. This comes from straightforward application of Lemma 7 and Lemma 8. □

Lemma 10. *Suppose demand is ρ'-concave. Then*

$$\frac{\Pi(Q)}{\Pi(Q^m)} \leq \frac{Q}{Q^m} \left(1 - \frac{1}{\rho'} \left((\frac{Q}{Q^m})^{\rho'} - 1 \right) \right).$$

Proof. Observe that

$$-P'(Q) = -P'(Q^m) \exp \int_{\ln Q^m}^{\ln Q} \frac{d \ln(-P'(Q))}{d \ln Q} d \ln Q.$$

Using Lemma 1, we have

$$\frac{d \ln(-P'(Q))}{d \ln Q} = \frac{-P''(Q)Q}{-P'(Q)} = -\psi(Q) \geq \rho' - 1,$$

and we also know from the monopoly FOC that

$$P'(Q^m) = -\frac{P^m - c}{Q^m}.$$

Applying the two properties yields

$$\begin{aligned} -P'(Q) &\geq \frac{P^m - c}{Q^m} \exp \left(\int_{\ln Q^m}^{\ln Q} (\rho' - 1) \, d \ln Q \right), \\ &= \frac{P^m - c}{Q^m} \left(\frac{Q}{Q^m} \right)^{\rho' - 1}. \end{aligned}$$

This implies

$$\begin{aligned} P(Q) = P(Q^m) + \int_{Q^m}^{Q} P'(q) dq &\leq P(Q^m) - \int_{Q^m}^{Q} \left(\frac{P^m - c}{Q^m} \left(\frac{q}{Q^m} \right)^{\rho' - 1} \right) dq \\ = P^m - \frac{P^m - c}{(Q^m)^{\rho'}} \left(\frac{Q^{\rho'}}{\rho'} - \frac{(Q^m)^{\rho'}}{\rho'} \right) \end{aligned}$$

From this,

$$
\begin{aligned}
&\implies P(Q)-c \le P^m - c - (P^m - c)\,\frac{1}{\rho'}\left(Q^{\rho'} - (Q^m)^{\rho'}\right)\\
&\implies \frac{P(Q)-c}{P^m-c} \le 1 - \frac{1}{\rho'}\left((\frac{Q}{Q^m})^{\rho'} - 1\right)\\
&\implies \frac{Q\,(P(Q)-c)}{Q^m\,(P^m-c)} \le \frac{Q}{Q^m}\left(1 - \frac{1}{\rho'}\left((\frac{Q}{Q^m})^{\rho'} - 1\right)\right)\\
&\implies \frac{\Pi(Q)}{\Pi(Q^m)} \le \frac{Q}{Q^m}\left(1 - \frac{1}{\rho'}\left((\frac{Q}{Q^m})^{\rho'} - 1\right)\right).
\end{aligned}
$$

□

Lemma 11. *Suppose that D, market demand, is ρ'-concave and ρ''-convex, with $\rho'' \ge \rho'$. Then for any $N > K$,*

$$\left(\frac{N(\rho''+K)}{K(\rho''+N)}\right)^{\frac{1}{\rho''}} \le \frac{Q_N^{ne}}{Q_K^{ne}} \le \left(\frac{N(\rho'+K)}{K(\rho'+N)}\right)^{\frac{1}{\rho'}},$$

and

$$\left(\frac{\rho'+K}{\rho'+N}\right)^{\left(1+\frac{1}{\rho'}\right)}\left(\frac{N}{K}\right)^{\frac{1}{\rho'}} \le \frac{\Pi(Q_N^{ne})}{\Pi(Q_K^{ne})} \le \left(\frac{\rho''+K}{\rho''+N}\right)^{\left(1+\frac{1}{\rho''}\right)}\left(\frac{N}{K}\right)^{\frac{1}{\rho''}}.$$

Proof. From Lemma 9, we have

$$\frac{d\ln\Pi(Q_N^{ne})}{dN} = -\frac{N-1}{N}\frac{1}{\psi(Q_N^{ne})+N+1}.$$

From Lemma 1, we have

$$\frac{1}{\rho''+N} \le \frac{1}{\psi(Q_N^{ne})+N+1} \le \frac{1}{\rho'+N}.$$

Thus,

$$-\frac{N-1}{N}\frac{1}{\rho'+N} \le \frac{d\ln\Pi(Q_N^{ne})}{dN} \le -\frac{N-1}{N}\frac{1}{\rho''+N}.$$

In turn, this implies

$$-\int_K^N \frac{M-1}{M}\frac{1}{\rho'+M}dM \le \ln\frac{\Pi(Q_N^{ne})}{\Pi(Q_K^{ne})} \le -\int_K^N \frac{M-1}{M}\frac{1}{\rho''+M}dM$$

Note that

$$
\begin{aligned}
-\int_K^N \left(1-\frac{1}{M}\right)\frac{1}{\rho+M}dM &= -\int_K^N \left(\frac{1}{\rho+M}-\frac{1}{\rho}\left(\frac{1}{M}-\frac{1}{\rho+M}\right)\right)dM \\
&= -\int_K^N \frac{1}{\rho+M}\left(\left(1+\frac{1}{\rho}\right)-\frac{1}{\rho}\frac{1}{M}\right)dM \\
&= -\left[\left(1+\frac{1}{\rho}\right)\ln(\rho+M)\Big|_K^N - \frac{1}{\rho}\ln M\Big|_K^N\right] \\
&= -\ln\left[\left(\frac{\rho+N}{\rho+K}\right)^{\left(1+\frac{1}{\rho}\right)}\left(\frac{K}{N}\right)^{\frac{1}{\rho}}\right] \\
&= \ln\left[\left(\frac{\rho+K}{\rho+N}\right)^{\left(1+\frac{1}{\rho}\right)}\left(\frac{N}{K}\right)^{\frac{1}{\rho}}\right]
\end{aligned}
$$

Thus,

$$
\left(\frac{\rho'+K}{\rho'+N}\right)^{\left(1+\frac{1}{\rho'}\right)}\left(\frac{N}{K}\right)^{\frac{1}{\rho'}} \le \frac{\Pi(Q_N^{ne})}{\Pi(Q_K^{ne})} \le \left(\frac{\rho''+K}{\rho''+N}\right)^{\left(1+\frac{1}{\rho''}\right)}\left(\frac{N}{K}\right)^{\frac{1}{\rho''}}.
$$

From Lemma 8, and using D is ρ'-concave,

$$
\begin{aligned}
\frac{1}{Q_N^{ne}}\frac{dQ_N^{ne}}{dN} &= \frac{1}{N(\psi(Q_N^{ne})+(N+1))} \\
&\le \frac{1}{N((\rho-1)+N+1)} \\
&= \frac{1}{N(\rho'+N)}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
Q_N^{ne} &\le \exp\left(\ln Q_K^{ne} + \int_K^N \frac{1}{M(\rho'+M)}dM\right) \\
&= \exp\left(\ln Q_K^{ne} + \int_K^N \frac{1}{\rho}\left(\frac{1}{M}-\frac{1}{\rho'+M}\right)dM\right)
\end{aligned}
$$

when $\rho \neq 0$. Thus,

$$
\frac{Q_N^{ne}}{Q_K^{ne}} \le e^{\frac{1}{\rho}\ln\left(\frac{N(\rho+K)}{K(\rho+N)}\right)}
$$

In a similar way, since D is ρ''-convex, we can obtain

$$
\frac{1}{N(\rho''+N)} \le \frac{1}{Q_N^{ne}}\frac{dQ_N^{ne}}{dN}
$$

and thus we have

$$\left(\frac{N(\rho''+K)}{K(\rho''+N)}\right)^{\frac{1}{\rho''}} \leq \frac{Q^{ne}_N}{Q^{ne}_K} \leq \left(\frac{N(\rho'+K)}{K(\rho'+N)}\right)^{\frac{1}{\rho'}}$$

In particular, when $K = 1$,

$$\left(\frac{N(\rho''+1)}{\rho''+N}\right)^{\frac{1}{\rho''}} \leq \frac{Q^{ne}_N}{Q^{m}} \leq \left(\frac{N(\rho'+1)}{\rho'+N}\right)^{\frac{1}{\rho'}}$$

Thus,

$$1-\left(\frac{\rho''+N}{N(\rho''+1)}\right)^{\frac{1}{\rho''}} \leq 1-\frac{Q^{m}}{Q^{ne}_N} \leq 1-\left(\frac{\rho'+N}{N(\rho'+1)}\right)^{\frac{1}{\rho'}}, \forall N \in \mathbb{N}.$$

Taking N to infinity for each term, we obtain

$$1-\left(\frac{1}{\rho''+1}\right)^{\frac{1}{\rho''}} \leq 1-\frac{Q^{m}}{Q^{pc}} \leq 1-\left(\frac{1}{\rho'+1}\right)^{\frac{1}{\rho'}},$$

where we remind the reader that Q^{pc} is the perfectly competitive quantity–the limiting value of Nash output. When $\rho' \to 0$, the upper bound goes to $1-\frac{1}{e}$. When $\rho' = \rho'' = 1$ (i.e., when D is concave), both the lower and upper bounds converge to $\frac{1}{2}$. □

A.3 Proof of Proposition 6

Proof. We only need to check whether there is a profitable one-shot deviation on the equilibrium paths, and the off-the-equilibrium paths in which only one firm has ever deviated (because for the other off-the equilibrium paths, firms play a static Nash equilibrium).

Unilaterally deviating at the capacity stage generates 0 profits forever and is unprofitable compared to the collusive outcome.

Consider a deviation from an off-the equilibrium path in which there has been only one deviation. First, consider the case in which there was a unilateral deviation in capacity choice. Consider a deviation by a firm i to set a lower price than p^m. Since the other firms set price 0, and the sum of their capacity is at least Q^{pc} (this is because, in the capacity choice, even in the case the deviated firm k (assume $k \neq i$) chose 0 capacity,

$$\sum_{j\neq i} x_j \geq 0 + \sum_{j\neq i,k} x_j = (N-2)\frac{Q^{pc}}{N-2} = Q^{pc}$$

where x_j is firm j's capacity. If $k = i$, then the sum is even bigger as $(N-1)\frac{Q^{pc}}{N-2}$. In words, their is no residual demand for firm i, so setting price c is incentive compatible. An immediate implication is that in such off-the-equilibrium paths, the continuation payoff is 0.[19]

Now consider a one-shot unilateral deviation from a history on the equilibrium path at $t \geq 1$. The profits under deviation are maximized by slightly undercutting p^m and selling at maximum capacity. In the previous paragraph, we have shown that the continuation payoff following this deviation is 0. Thus, the deviation is unprofitable if and only if

$$\frac{1}{N} Q^m p^m \geq (1-\delta) \frac{1}{N-2} Q^{pc} p^m$$

Thus, the critical discount rate is given by

$$\underline{\delta} = 1 - \frac{N-2}{N} \frac{Q^m}{Q^{pc}}.$$

Finally, we can use ρ-concavity and ρ-convexity to bound the right-hand-side.

Lemma 12.

$$Q_K^{ne} \left(\frac{N}{K} \frac{K+\rho''}{N+\rho''} \right)^{\frac{1}{\rho''}} \leq Q_N^{ne} \leq Q_K^{ne} \left(\frac{N}{K} \frac{K+\rho'}{N+\rho'} \right)^{\frac{1}{\rho'}},$$

and

$$Q_K^{ne} \left(\frac{K+\rho''}{K} \right)^{\frac{1}{\rho''}} \leq Q^{pc} \leq Q_K^{ne} \left(\frac{K+\rho'}{K} \right)^{\frac{1}{\rho'}}.$$

Proof. Suppose that demand is ρ'-concave and ρ''-convex, Lemma 8 states that $\frac{d \ln Q_N^{ne}}{dN} = \frac{1}{N\left(N+1+\psi(Q_N^{ne})\right)}$. By Lemma 1, we therefore know that $\frac{1}{N(N+\rho'')} \leq \frac{d \ln Q_N^{ne}}{dN} \leq \frac{1}{N(N+\rho')}$.

For $N > K$, we have

$$Q_N^{ne} = Q_K^{ne} \exp\left(\int_K^N \frac{d \ln Q_M^{ne}}{dM} dM \right). \tag{19}$$

We will use the aforementioned bounds to bound Q_N^{ne} in terms of Q_K^{ne}.

$$\int_K^N \frac{1}{M\left(M+\rho''\right)} dM \leq \int_K^N \frac{d \ln Q_M^{ne}}{dM} dM \leq \int_K^N \frac{1}{M\left(M+\rho''\right)} dM,$$

which, by integrating, implies

$$\frac{\ln M - \ln(M+\rho'')|_K^N}{\rho''} \leq \int_K^N \frac{d \ln Q_M^{ne}}{dM} dM \leq \frac{\ln M - \ln(M+\rho')|_K^N}{\rho'},$$

[19] For this argument, we need $N \geq 3$. We do not know whether we can dispense this assumption.

or equivalently,

$$\frac{1}{\rho''}\ln\left(\frac{N}{K}\frac{K+\rho''}{N+\rho''}\right)\leq\int_K^N\frac{d\ln Q_M^{ne}}{dM}dM\leq\frac{1}{\rho'}\ln\left(\frac{N}{K}\frac{K+\rho'}{N+\rho'}\right).$$

Applying this to Equation 19 yields

$$Q_K^{ne}\exp\left(\frac{1}{\rho''}\ln\left(\frac{N}{K}\frac{K+\rho''}{N+\rho''}\right)\right)\leq Q_N^{ne}\leq Q_K^{ne}\exp\left(\frac{1}{\rho'}\ln\left(\frac{N}{K}\frac{K+\rho'}{N+\rho'}\right)\right),$$

which simplifies to

$$Q_K^{ne}\left(\frac{N}{K}\frac{K+\rho''}{N+\rho''}\right)^{\frac{1}{\rho''}}\leq Q_N^{ne}\leq Q_K^{ne}\left(\frac{N}{K}\frac{K+\rho'}{N+\rho'}\right)^{\frac{1}{\rho'}},$$

Finally, taking the limit as N goes to infinity, and writing $Q^{pc}=\lim_{N\to\infty}Q_N^{ne}$, we have

$$Q_K^{ne}\left(\frac{K+\rho''}{K}\right)^{\frac{1}{\rho''}}\leq Q^{pc}\leq Q_K^{ne}\left(\frac{K+\rho'}{K}\right)^{\frac{1}{\rho'}}.$$

□

Using Lemma 12, with $K=1$, we conclude that $(1+\rho'')^{\frac{1}{\rho''}}\leq\frac{Q^{pc}}{Q^m}\leq(1+\rho')^{\frac{1}{\rho'}}$. Applying this to our earlier expression for the critical discount rate yields

$$1-\frac{N-2}{N}(1+\rho')^{\frac{1}{\rho'}}\leq\underline{\delta}_N\leq 1-\frac{N-2}{N}(1+\rho'')^{\frac{1}{\rho''}}.$$

Thus, taking the limit as $N\to\infty$, we get bounds for the asymptotic critical discount rate, $\underline{\delta}$:

$$1-\left(\frac{1}{1+\rho'}\right)^{\frac{1}{\rho'}}\leq\underline{\delta}\leq 1-\left(\frac{1}{1+\rho''}\right)^{\frac{1}{\rho''}}.$$

Finally, a set of uniform bounds can be derived by letting N go to infinity on the left-hand-side, and letting $N=3$ in the right-hand-side.

$$0\leq\underline{\delta}\leq 1-\frac{1}{3}\left(\frac{1}{1+\rho''}\right)^{\frac{1}{\rho''}}.$$

□

References

Abreu, Dilip (1986), "Extremal equilibria of oligopolistic supergames." *Journal of Economic Theory*, 39, 191–225.

Anderson, Simon P. and Régis Renault (2003), "Efficiency and surplus bounds in cournot competition." *Journal of Economic Theory*, 113, 253–264.

Benoit, Jean-Pierre and Vijay Krishna (1987), "Dynamic duopoly: Prices and quantities." *The review of economic studies*, 54, 23–35.

Brock, William A and Jose A Scheinkman (1985), "Price setting supergames with capacity constraints." *The Review of Economic Studies*, 52, 371–382.

Bulow, Jeremy, John Geanakoplos, and Paul Klemperer (1985), "Holding idle capacity to deter entry." *The Economic Journal*, 95, 178–182.

Davidson, Carl and Raymond Deneckere (1986), "Long-run competition in capacity, short-run competition in price, and the cournot model." *The Rand Journal of Economics*, 404–415.

Fershtman, Chaim and Ariel Pakes (2000), "A dynamic oligopoly with collusion and price wars." *The RAND Journal of Economics*, 207–236.

Fudenberg, Drew and Jean Tirole (1984), "The fat-cat effect, the puppy-dog ploy, and the lean and hungry look." *The American Economic Review*, 74, 361–366.

Herk, Leonard F (1993), "Consumer choice and cournot behavior in capacity-constrained duopoly competition." *The RAND Journal of Economics*, 399–417.

Ivaldi, Marc, Bruno Jullien, Patrick Rey, Paul Seabright, and Jean Tirole (2007), *The Political Economy of Antitrust (Contributions to Economic Analysis, Volume 282)*, chapter The Economics of Tacit Collusion:Implications for Merger Control, 217–239. URL https://www.emeraldinsight.com/doi/abs/10.1016/S0573-8555%2806%2982008-0.

Kreps, David M and Jose A Scheinkman (1983), "Quantity precommitment and bertrand competition yield cournot outcomes." *The Bell Journal of Economics*, 326–337.

Kühn, Kai-Uwe (2012), "How market fragmentation can facilitate collusion." *Journal of the European Economic Association*, 10, 1116–1140.

Levenstein, Margaret C and Valerie Y Suslow (2006), "What determines cartel success?" *Journal of economic literature*, 44, 43–95.

Spence, A Michael (1977), "Entry, capacity, investment and oligopolistic pricing." *The Bell Journal of Economics*, 534–544.

Spencer, Barbara J and James A Brander (1992), "Pre-commitment and flexibility: Applications to oligopoly theory." *European Economic Review*, 36, 1601–1626.