THE JOURNAL OF INDUSTRIAL ECONOMICS 0022-1821
Volume LVIII March 2010 No. 1

OUTPUT-EXPANDING COLLUSION IN THE PRESENCE OF A COMPETITIVE FRINGE*

JUAN-PABLO MONTERO†
JUAN IGNACIO GUZMAN‡

Following the structure of many commodity markets, we consider a few large firms and a competitive fringe of many small suppliers choosing quantities in an infinite-horizon setting subject to demand shocks. We show that a collusive agreement among the large firms may not only bring an output contraction but also an output expansion (relative to the non-collusive output level). The latter occurs during booms and is due to the strategic substitutability of quantities. We also find that the time at which maximal collusion is most difficult to sustain can be either at booms or recessions. The international copper cartel of 1935–39 is used to illustrate some of our results.

I. INTRODUCTION

IN TABLE I WE REPRODUCE Orris C. Herfindahl's Table 3 [1959, p. 115] with a summary of the evolution of the so-called international copper cartel that consisted of the five largest firms and operated during the four years preceding the Second World War. Herfindahl argues that the cartel was successful in restricting production during the periods of low demand (denoted as Quota status and associated with lower spot prices in the London Metal Exchange) but failed to extend such restrictions to the periods of high demand when the cartel and non-cartel firms returned to their non-collusive output levels.[1]

Herfindahl's description appears consistent with some existing collusion theories; in particular, with Rotemberg and Saloner's [1986] prediction for

*We thank Claudio Agostini, Jeremy Bulow, the Editor, Joe Harrington, Bill Hogan, Jason Lepore, Salvador Valdés, Felipe Zurita, two anonymous referees and seminar participants for helpful discussions and comments. Part of this work was carried out while Montero was visiting Harvard's Kennedy School of Government (KSG) under a Repsol YPF-KSG Research Fellowship. Montero also gratefully acknowledges financial support from Fondecyt (1051008) and Instituto Milenio SCI (P05-004F), and Guzmán from PUC's Mineral Economics Program.

†Authors' affiliations: Department of Economics, Pontificia Universidad Católica de Chile, Santiago, Chile and Center for Energy and Environmental Policy Research, MIT.
e-mail:jmontero@faceapuc.cl

‡Program in Mineral Economics, Pontificia Universidad Católica de Chile, Santiago, Chile.
e-mail:jguzmanb@puc.cl

[1] Walters [1944] also comments on the satisfactory operation of the cartel in that there is no indication that sanctions for non-compliance were ever invoked.

TABLE I
EVOLUTION OF THE 1935–39 COPPER CARTEL

Period	Quota status	Cartel production: Production (annual rate)	Cartel production: Per cent change from preceding period	Noncartel production: Production (annual rate)	Noncartel production: Per cent change from preceding period	London spot price: Cents per pound	London spot price: Per cent change from preceding period
1. July–Dec. 1935	Quotas	559		408		8.2	
2. Jan.–Dec. 1936	Quotas	598	+7.0%	368	−9.8%	9.5	+16%
3. Jan.–Nov. 1937	No Quotas	917	+53.4	439	+19.3	13.4	+41
4. Dec. 1937–Sept. 1938	Quotas	762	−16.9	481	+9.6	9.7	−28
5. Oct. 1938–Dec. 1938	No Quotas	948	+24.4	504	+4.8	10.6	+9
6. Jan. 1939–July 1939	Quotas	736	−22.4	511	+1.4	10.0	−6

Source: Table 3 of Herfindahl [1959, p. 115]

the evolution of a cartel under conditions of demand fluctuations in that collusive firms have more difficulties in sustaining collusion during booms (i.e., periods of high demand).[2] We advance a different behavioral hypothesis in this paper. We posit that the large output expansions undertaken by cartel members during the two booms (Jan.–Nov. 1937 and Oct.–Dec. 1938) may not necessarily reflect a return to the non-collusive (i.e., Nash-Cournot) equilibrium but rather a continuation with the collusive agreement in the form of a coordinated output expansion of cartel members above their Nash-Cournot levels.[3]

The objective of this paper is to explore the conditions under which a collusive agreement, if sustained, can take an output-expanding form at least during some part of the business cycle. Although we do not run any empirical test, we will see that the international copper cartel of 1935–39 as well as many of today's commodity markets appear to be good candidates to which such collusive characterization might apply.[4] There are basically two reasons for this. First, in these markets a firm' strategic variable is its level of production while prices are cleared, say, in a metal exchange. Second,

[2] Rotemberg and Saloner's [1986] prediction can change if we introduce imperfect monitoring (Green and Porter [1984]), a less than fully random demand evolution (Haltiwanger and Harrington [1991] and Bagwell and Staiger [1997]), and capacity constraints (Staiger and Wolak [1992]).

[3] From news reports of the time it appears that output expansions were indeed not totally left to each cartel member's unilateral actions but were somehow also orchestrated by the cartel (e.g., *New York Times* [1938], Oct. 11, pg. 37 and Oct. 18, pg. 37).

[4] Note that unlike other commodity cartels that organize around export sales quotas, the 1935–39 copper cartel was organized exclusively around production quotas. The international tin cartel of 1931–1941, for example, combined control of production with export sales quotas. For more, see Walters [1944].

collusive efforts, if any, are likely to be carried out by a fraction of the industry (typically, the largest firms) leaving an important fraction of the industry (consisting mostly of a large number of small firms) outside the collusive agreement but nevertheless enjoying any eventual price increase brought forward by the collusive agreement (we will refer to the group of non-cartel firms as a competitive fringe and to the group of potential cartel firms as strategic or large firms).

Our model consists of a few large and identical firms and a fringe of competitive suppliers simultaneously deciding production in each period over an infinite horizon.[5] Consistent with practical observation, we assume that the entry or exit of large firms is a rare event. Fringe firms are assumed to be infinitesimally small, so they play along their static reaction (best-response) function.[6] The problem of the cartel of large firms is then to find the optimal collusive agreement, taking as given that fringe firms play a static best response in each period. Analogous to the long-run player of Fudenberg and Levine [1989], the cartel's optimal strategy is to play its Stackelberg quantity in each period. Since cartel profits under the Stackelberg play are by definition higher than in the Nash-Cournot equilibrium, the Stackelberg outcome is feasible to sustain by the threat of Nash reversion whenever large firms are sufficiently patient. The main result of the paper is that as demand increases, the cartel's Stackelberg quantity can be greater than the sum of the Nash-Cournot (i.e., non-collusive) quantities of the large firms.[7]

The possibility of an output-expanding collusion arises from a trade off that large firms face when deciding on their optimal collusive agreement. On the one hand, large firms understand that non-cooperative (i.e., Nash-Cournot) play typically results in too much output (as they perceive it), so they want to correct this by restricting production to levels that rest below their respective static reaction curves. This can be seen as the 'corrective' objective of the cartel. On the other hand, the firms in the cartel are aware of the strategic substitutability of quantities (Bulow *et al.* [1985] and Fudenberg and Tirole [1984]). They understand that if they can credibly commit to produce a larger output (i.e., production levels that rest above their respective reaction curves), then the fringe firms will produce less. This

[5] Although the collusion literature usually assumes a structure of identical firms, this heterogeneous structure in which a few large firms compete with many smaller firms has long been recognized (e.g., Arant [1956] and Pindyck [1979]).

[6] In that sense fringe firms are not different from consumers who also play along their static reaction function. We could also use fringe firms of small but measurable size 'unconcerned' about the future, very much like the short-run players of Fudenberg *et al.* [1990]. For more see Mailath and Samuelson [2006].

[7] It is not new in the literature that the introduction of a competitive fringe can alter existing results. Riordan [1998], for example, shows how and when the presence of a (downstream) fringe reverses the well known procompetitive result of backward vertical integration by a downstream monopolist.

can be seen as the 'strategic' objective of the cartel, which is, of course, closely related to the Stackelberg logic of the first-mover advantage. But in our model, large firms do not move first; instead, they build credibility from their repeated interaction and the threat of Nash-Cournot reversion.

Thus, when a cartel interacts with a fringe, the cartel faces a trade-off between the corrective benefit of restricting output and the strategic benefit of increasing output. When the fringe is negligible, the only objective that matters for the cartel is the corrective one, leading to the traditional output-contracting result (i.e., cartel output below Nash-Cournot levels). Conversely, when there is only one firm in the cartel, the only objective that matters is the strategic one, leading to the output-expanding result (i.e., cartel output above its Nash-Cournot level). For the remaining cases, the resolution of the trade-off will ultimately depend on the number of large firms (the corrective benefit increases with the number of large firms) and on the market share enjoyed by the fringe (the strategic benefit increases with fringe size). This latter will in turn depend on the cost differences between large and fringe firms and on the magnitude of the demand shocks. One can always find the fringe's costs sufficiently low (high) that it is optimal for strategic firms to implement an output expanding (contracting) collusion for all possible realizations of demand. As we speculate for the copper cartel of 1935–39, however, the more interesting case is one in which fringe costs generate both output-contracting collusion during recessions and output-expanding collusion during booms, which is when the fringe market share is larger.

Another way to appreciate our results is by contrasting them with those obtained in a price-setting game with some product differentiation.[8] Because of the strategic complementarity of prices, the basic trade-off between corrective and strategic objectives does not arise when firms set prices. Cartel members prefer price increases for both corrective and strategic reasons.

The presence of an important fraction of non-cartel firms also has implications for cartel firms' ability to sustain the collusive outcome throughout the business cycle. In standard collusion analysis, deviation of a cartel firm is punished by the remaining cartel firms with increases of production to Nash-Cournot levels. The same logic applies here when the deviation occurs at periods of output contraction. The punishment logic is somehow different when the deviation occurs at periods of output expansion. Since a deviation entails cutting back production toward Nash-Cournot levels, the remaining cartel firms as well as fringe firms also profit from the deviation in the short run. Then, in moments of output-expanding collusion the cartel discipline does not come from the threat that

[8] If products are perfectly homogeneous it is immediate that we cannot have an output-expanding collusion (i.e., prices below Nash-Bertrand levels) because that would require large firms to price below their marginal costs.

remaining cartel firms will flood the market after the deviation, but from the fear that cartel firms will lose credibility to effectively sustain the Stackelberg outcome. If strategic firms fail to produce according to Stackelberg principles, this causes fringe firms to believe that cartel firms have no credibility to act in such a way in the future, leading fringe firms to produce according to Nash-Cournot forever in the future.[9] Note that this is a natural enforcement mechanism because the whole idea of expanding output above Nash-Cournot levels hinges around the contracting response of the fringe.

Using the same i.i.d. demand shocks of Rotemberg and Saloner [1986] we also study the sustainability of collusion over the business cycle. Unlike Rotemberg and Saloner [1986], we show cases in which it is more difficult for firms to sustain maximal collusion (i.e., the monopoly outcome) during recessions than during booms. In addition, we illustrate how the optimal collusive agreement departs from maximal collusion when firms are not sufficiently patient.

Because many commodity markets are characterized by the presence of a relatively large fraction of small suppliers that will never enter into a collusive agreement, our results can have important welfare implications. We cannot rule out, on theoretical grounds, that collusion efforts by a group of large firms may be welfare enhancing when periods of output-contracting collusion are followed by periods of output-expanding collusion.[10] Based on the aggregate data of Table I, we illustrate this possibility in a numerical exercise for the copper cartel of 1935–39.

The rest of the paper is organized as follows. In Section II we present the model and derive the (non-collusive) Nash-Cournot equilibrium. In Section III we present the maximal collusive equilibrium and demonstrate the possibility of an output-expanding collusion. In Section IV, we study the cartel stability over the business cycle. The numerical exercise based on the copper cartel data is in Section V. Concluding remarks follow. All proofs are relegated to the Appendix.

II. OLIGOPOLY–FRINGE MODEL

II(i). *Notation*

A group of n identical (strategic) firms ($i = 1, \ldots, n$) and a competitive fringe consisting of a continuum of firms (indexed by $j \in [0, J]$) produce some

[9] Note that this enforcement mechanism is no different, say, from the one used by the durable-good monopolist of Ausubel and Deneckere [1989].

[10] Bulow *et al.* [1985] make a closely related point when they ask whether 'little competition is a good thing.' The answer depends on whether the entry of a *small* fringe will cause the incumbent monopolist to expand or contract output: entry is a good thing with strategic complements and is welfare reducing with strategic substitutes.

commodity in an infinite-horizon setting. At the beginning of each period, firms simultaneously choose their production levels and the price clears according to the inverse demand curve $P(Q;\theta)$, where Q is total production, $\theta \in [\underline{\theta}, \overline{\theta}]$ is a demand shock observed by all firms before they engage in production, $\partial P(Q;\theta)/\partial Q \equiv P_Q(Q;\theta) < 0$, and $\partial P(Q;\theta)/\partial Q \equiv P_\theta(Q;\theta) > 0$ for all Q.[11]

Strictly speaking, only the n strategic firms have the possibility of choosing among different production levels; a fringe firm's decision is simply whether or not to bring its unit of output to the market.[12] The production cost of each strategic firm is $C_s(q_s)$ with $C_s'(q_s) > 0$ and $C_s''(q_s) \geq 0$ ('s' stands for strategic firm). The unit cost of fringe firm j is c_j. The c_j's, which vary across firms, can be cost-effectively arranged along a marginal cost curve $C_f'(Q_f)$ with $C_f''(Q_f) > 0$, where $Q_f = \int_0^J q_{fj}\, \mathrm{d}j$ is fringe output (we will use capital letters to denote group production and small letters to denote individual production, so strategic firms' output is $Q_s = \sum_{i=1}^n q_{si}$ and total output is $Q = Q_s + Q_f$).

In some passages of the paper we will introduce, with little loss of generality, some simplifying assumptions to the model that will allow us to better illustrate some of our results. In particular, we will assume that $P(Q;\theta) = \theta(a - bQ)$, that strategic firms have no production costs and that the fringe's aggregate marginal cost curve is $C_f'(Q_f) = cQ_f$, where a, b and c are strictly positive parameters.[13]

II(ii). *The Nash-Cournot Equilibrium*

A commonly used approach for finding the (static) Nash-Cournot equilibrium in the presence of a competitive fringe is to first subtract the fringe's supply function from the market demand to obtain the 'residual demand' faced by the large firms and then solve the non-cooperative game among the large firms. This residual-demand approach, however, violates the simultaneous-move assumption. It implicitly assumes a sequential timing within any given period: first, large firms announce or choose their quantities; then and after observing large firms' output decisions, fringe firms choose their quantities. In the absence of technical reasons, this sequential timing can only be supported by some degree of cooperation

[11] Below we address the specific cases of multiplicative shock (i.e. $P(Q;\theta) = \theta P(Q)$) and additive shock (i.e., $P(Q;\theta) = P(Q) + \theta$).

[12] Including some fringe firms with output flexibility complicates the algebra with no implications in the results. It would be interesting, however, to study more formally the process of cartel formation when there are heterogenous firms.

[13] That the costs of large firms are, on average, lower than the costs of smaller firms is not a bad assumption – at least for mineral markets (Crowson [2003]). However, we do not need such assumption for our results; we could have just worked with $C_s(q_s) = c_s q_s$ at the expense of mathematical tractability.

(i.e., collusion) among the large firms that ensures that large firms will stick to their announcements or that they will not move again together with the fringe.[14] Without such cooperation, the static game necessarily collapses into a simultaneous move game.[15]

Thus, the Nash-Cournot equilibrium of the one-period (simultaneous-move) game, i.e., the equilibrium in the absence of any collusion efforts, is found by solving each firm's problem as follows[16]

$$\max_{q_{si}} P\left(q_{si} + \sum_{j \neq i} q_{sj} + Q_f; \theta\right) q_{si} - C_s(q_{si}) \quad \text{for all } i = 1, \ldots, n \tag{1}$$

$$q_{fj} = \begin{cases} 1 & \text{if } c_j \leq P(Q_s + Q_f; \theta) \\ 0 & \text{if } c_j > P(Q_s + Q_f; \theta) \end{cases} \quad \text{for all } j \tag{2}$$

The n first-order conditions associated to (1) give us the best response of each strategic firm to the play of all remaining firms. Similarly, (2) summarizes the best response of each fringe firm, which is to produce as long as its unit cost is equal to or below the clearing price.

Given the symmetry of the problem, the equilibrium outcome of the one-period game is given by

$$P_Q(Q^{nc}; \theta) \frac{Q_s^{nc}}{n} + P(Q^{nc}; \theta) - C_s'(Q_s^{nc}/n) = 0 \tag{3}$$

$$P(Q^{nc}; \theta) = C_f'(Q_f^{nc}) \tag{4}$$

where 'nc' stands for Nash-Cournot or non-collusive equilibrium and $Q^{nc} = Q_s^{nc} + Q_f^{nc}$. Solving, we obtain $Q_s^{nc}(\theta)$ and $Q_f^{nc}(\theta)$.[17]

[14] Suppose first that the sequential timing is the result of early production announcements by strategic firms. At the production stage, however, a strategic firm would like to deviate from its original announcement by producing less. Suppose instead that the sequential timing is the result of early and observable production by strategic firms (together with a commitment of no additional production). At the time fringe firms are called to produce, however, a strategic firm will find it profitable to deviate from its commitment by bringing additional output to the market.

[15] Besides, the sequential timing assumption can lead to non-collusive outcomes that are far from realistic, as we shall see in the next section.

[16] One of the first oligopoly-fringe models that explicitily adopts this simultaneous-move assumption is Salant's [1976] model for the oil market.

[17] The corresponding quantities for the simplified model are

$$Q_s^{nc} = \frac{nac}{b[c(n+1) + \theta b]} \quad \text{and} \quad Q_f^{nc} = \frac{\theta a}{c(n+1) + \theta b}$$

Note that under this particular formulation, $dQ_s^{nc}(\theta)/d\theta < 0$ and $dQ_f^{nc}(\theta)/d\theta > 0$; but we do not require these properties in any of our Propositions.

III. COLLUSION

It is well known that in an infinite-horizon setting, strategic firms may be able to sustain outcomes in subgame perfect equilibrium that generate higher profits than the outcome in the corresponding one-period game. Leaving for later discussion how difficult it is for firms to sustain these collusive outcomes in equilibrium, or alternatively, assuming for the moment that the discount factor δ (of strategic firms) is close enough to one,[18] in this section we are interested in finding the maximal collusive agreement for the strategic firms, that is, the agreement that implements the cartel's Stackelberg outcome.

III(i). *Maximal Collusion*

Assuming that all players acting at date t have observed the history of play up to date t,[19] large firms can implement their collusive agreement by following history-dependent equilibrium strategies. Let Q_s^m and Q_f^m denote the (aggregate) quantities corresponding to the maximal collusive agreement, which, for example, can be implemented by the following set of (symmetric) trigger strategies (which depend on the realization of the demand shock θ): In period 0, strategic firm $i = 1,\ldots, n$ plays Q_s^m/n and fringe firms play, on aggregate, Q_f^m. In period t, firm i plays Q_s^m/n if in every period preceding t all strategic firms have played Q_s^m/n; otherwise it plays Q_s^{nc}/n. Fringe firms, on the other hand, play Q_f^m in t if in every period preceding t all strategic firms have played Q_s^m/n; otherwise they play Q_f^{nc}.

Since strategic firms are symmetric and there are no economies of scale, it is optimal for each strategic firm to produce $q_{si}^m = Q_s^m/n$, hence

$$(5)\qquad Q_s^m = \arg\max_{Q_s}\{P(Q_s + Q_f(Q_s); \theta)Q_s - nC_s(Q_s/n)\}$$

where $Q_f(Q_s)$ is the fringe's static best-response to Q_s, which is implicitly given by

$$(6)\qquad P(Q_f + Q_s; \theta) = C_f'(Q_f)$$

Replacing $Q_f(Q_s)$ from (6) into (5), the strategic firms' maximal collusive outcome solves

$$(7)\qquad \frac{C_f''(Q_f^m)P_Q(Q^m;\theta)}{C_f''(Q_f^m) - P_Q(Q^m;\theta)}Q_s^m + P(Q^m;\theta) - C_s'(Q_s^m/n) = 0$$

[18] The discount factor of fringe firms is irrelevant since they always operate along their static best-response function.

[19] As explained by Crowson [1999], it is common in mineral markets to see smaller firms staying around for as long as larger firms. It is also the case that smaller firms can learn about previous play without being physically present, either through word-of-mouth or more likely from written sources.

$$P(Q^m;\theta) = C'_f(Q^m_f) \tag{8}$$

where $Q^m = Q^m_s + Q^m_f$. Solving, we obtain the collusive equilibrium strategies, $Q^m_s(\theta)$ and $Q^m_f(\theta)$, as a function of the observable demand shock θ.[20]

Note that having Q_f as a function of Q_s in (5) resembles a Stackelberg (sequential) game but it is only because in this repeated game, large firms anticipate and use fringe firms' equilibrium response in constructing their optimal actions and not because of some exogenous first-mover advantage.

III(ii). *Output-Expanding Collusion*

Having characterized the collusive and non-collusive equilibrium solutions, we are ready to present our main proposition. By comparing equilibrium conditions (3)–(4) with (7)–(8), it can be established

Proposition 1. Suppose that there is a level of demand $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$ at which Nash-Cournot play satisfies

$$-P_Q(Q^{nc}(\hat{\theta});\hat{\theta}) = (n-1)C''_f(Q^{nc}_f(\hat{\theta})) \tag{9}$$

Then, maximal collusion and Nash-Cournot coincide at $\hat{\theta}$. Suppose further that

$$-\frac{dP_Q(Q^m(\hat{\theta});\hat{\theta})}{d\theta} > (n-1)C'''_f(Q_f(\hat{\theta}))\frac{dQ^m_f(\hat{\theta})}{d\theta} \tag{10}$$

whenever (9) holds. Then, maximal collusion prescribes strategic firms to produce less than Nash-Cournot if $\theta < \hat{\theta}$ and more than Nash-Cournot if $\theta > \hat{\theta}$.

The first part of the proposition identifies the demand level – provided that it exists, i.e., $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$ – at which the non-collusive and collusive solutions coincide. Depending on the functional forms of the demand and cost functions, this demand level may not be unique. More importantly, condition (9) opens up the possibility for an output-expanding collusion whenever $-P_Q(Q(\theta);\theta) > (n-1)C''_f(Q_f(\theta))$, which will always be the case for $C''_f(\cdot)$ sufficiently small. The second part of the proposition establishes the condition for the collusive agreement to involve output expansions (above Nash-Cournot levels) during periods of higher demand (i.e., $\theta > \hat{\theta}$) and output contractions during periods of lower demand (i.e., $\theta < \hat{\theta}$).

[20] For the simplified model the equilibrium quantities are

$$Q^m_s = \frac{a}{2b} \quad \text{and} \quad Q^m_f = \frac{\theta a}{2(c+\theta b)}$$

Note that $\partial Q^m_f(\theta)/\partial\theta > 0$.

In providing intuition for these results, it is useful to see first how Proposition 1 applies to our simplified model: $P(Q;\theta) = \theta(a - bQ)$, $C_s(q_s) = 0$ and $C_f'(Q_f) = cQ_f$. From (9), we obtain that the demand level at which the collusive outcome coincides with the Nash-Cournot outcome is $\hat{\theta} = (n-1)c/b$. And since it is clear that (10) holds, the collusive agreement among the strategic firms is to expand output (and, hence, to lower prices) whenever $\theta > (n-1)c/b$; a condition that can be conveniently expressed in terms of Cournot market shares as follows (see footnote 17 for the values of Q_f^{nc} and Q_s^{nc})

$$(11) \qquad \theta > \frac{(n-1)c}{b} \Leftrightarrow \frac{Q_f^{nc}}{Q^{nc}} > \frac{n-1}{2n-1}$$

If $n = 2$, for example, the collusive agreement will be output-expanding during those periods in which the fringe's Cournot share would be above 1/3. This market share 'threshold' increases with the number of large firms to the limiting value of 1/2. Thus, in this simplified model, whenever the fringe's Cournot share is above 1/2, it is collusive optima for large firms to expand output above Cournot levels regardless of their number.

The simplified model illustrates neatly the trade-off that large firms face when deciding on their collusive agreement. On the one hand, large firms understand that a Nash-Cournot play typically results in too much output (as they perceive it), so they want to correct this by restricting production to levels that rest below their respective static reaction curves. This can be seen as the 'corrective' objective of the cartel. On the other hand, the firms in the cartel are aware of the strategic substitutability of quantities in that they understand that if they can credibly commit to produce a larger output (i.e., production levels that rest above their respective reaction curves), then the fringe firms wil! produce less (note from (6) that $Q_f'(Q_s) = P_Q(Q)/[C_f''(Q_f) - P_Q(Q)] < 0$). This can be called the 'strategic' objective of the cartel, which is, of course, closely related to the Stackelberg logic of the first-mover advantage. But in our model, large firms do not move first; instead, they build credibility from their repeated interaction and the threat of Nash-Cournot reversion.

Thus, when a cartel interacts with a fringe, the cartel faces a trade-off between the corrective benefit of restricting output and the strategic benefit of increasing output. At $\theta = \hat{\theta}$, the two objectives exactly balance out. For demand levels different than $\hat{\theta}$ the resolution of the trade-off can go in either direction depending on the number of large firms, cost differences and demand levels. At one extreme, when there is only one firm in the cartel ($n = 1$), the only objective that matters is the strategic one, leading the one-firm cartel to produce above its Nash-Cournot level for all levels of demand, i.e., $-P_Q(Q(\theta);\theta)/C_f''(Q_f(\theta) > 0$ for all θ. As n increases, the corrective objective becomes more important and may eventually dominate the strategic one, i.e., $n-1>-P_Q(Q(\theta);\theta)/C_f''(Q_f(\theta))$.

At the other extreme, when the fringe is negligible ($C_f''(Q_f)$ very large for all Q_f), the only objective that matters for the cartel is the corrective one, leading to the traditional output-contracting result, i.e., $-P_Q(Q(\theta);\theta)<(n-1)C_f''(Q_f(\theta))$ for all θ. As we increase the relative size of the fringe, the strategic objective assumes more weight and eventually may become more important than the corrective one. It is clear from (11) that for the simple model, the relative size of the fringe is larger the lower its cost (i.e., lower c) or higher the demand, or both. For the more general model, the effect of a demand increase on the likelihood of an output expansion is less obvious. Expression (10) in Proposition 1 establishes the exact conditions for this to be the case. In fact, when the fringe supply is convex in prices (i.e., $C_f'''(Q_f)<0$),[21] condition (10) holds for any particular shape of the demand shock.[22]

Finally, another way to appreciate the results of Proposition 1 is by contrasting them with those obtained in a price-setting game with some product differentiation. In the Appendix we develop a (general) model where we formally show that it is never optimal for the strategic firms to jointly price below their (non-collusive) Nash-Bertrand price levels. Because of the strategic complementarity of prices, the basic trade-off between corrective and strategic objectives does not arise when firms set prices.[23] Cartel members are attracted to price increases for both corrective and strategic reasons.

IV. COLLUSION OVER THE BUSINESS CYCLE

We have characterized the maximal collusive agreement but have said nothing about how difficult is for the strategic firms to sustain such an agreement under varying demand conditions. The question of whether it is more difficult for firms to sustain collusion during booms than during recessions (or vice versa) has received a great deal of attention in the

[21] At least for mineral markets, Crowson [1999] and [2003] explains that total supply from smaller firms (including that from non-conventional sources such as recycling) tend to be much more elastic at higher prices.

[22] The right hand side of (10) is always negative (recall that $dQ_f'''(\theta)/d\theta>0$ for all θ) while the left hand side is positive for a multiplicative shock (i.e., $P(Q;\theta)=\theta P(Q)$) and zero for an additive shock (i.e., $P(Q;\theta)=P(Q)+\theta$). Note also that for an additive shock, the condition that separates output-contracting collusion from output-expanding collusion reduces to whether $-(n-1)C_f''/P_Q$ is greater or lower than the unity. So, if both the fringe's supply function $C_f'(Q_f)$ and the demand function $P(Q)$ are linear, the collusive agreement would exhibit either output contractions or expansions over the entire business cycle, which appears not very realistic. But if $C_f'''(Q_f)<0$, we could then have a mixed regime with output contraction for $\theta<\hat{\theta}$ and output expansion for $\theta>\hat{\theta}$, where $\hat{\theta}$ solves $-(n-1)C_f''(Q_f(\hat{\theta}))/P_Q(Q(\hat{\theta}))=1$. It is possible that in reality, shocks are best modeled as a combination of both specifications.

[23] Note from eq. (17) in the Appendix that $\partial p_f(p_{s1},\ldots p_{sn})/\partial p_{si}>0$ for all i, where p_f and p_{si} are, respectively, the prices charged by fringe firms and strategic firm i.

literature after the pioneers works of Green and Porter [1984] and Rotemberg and Saloner [1986]. Since our intention is not to provide a discussion of how all existing results could change with the introduction of a (large) fringe, we follow Rotemberg and Saloner [1986] in that demand is subject to (observable) i.i.d. θ shocks. We also assume that all (strategic) firms use the same factor $\delta \in (0,1)$ to discount future profits.

IV(i). *Critical Discount Factor*

For maximal collusion to be sustained throughout the business cycle, it must hold for all θ and for each strategic firm that the profits along the collusive path be equal or greater than the profits from cheating on the collusive agreement and falling, thereafter, into the punishment path, that is

$$\pi^m(\theta) + \delta V^m \geq \pi^d(\theta) + \delta V^p \tag{12}$$

where $V^m = E_\theta[\pi^m(\theta)]/(1-\delta)$ is the firm's expected present value of profits along the collusive path, $\pi^d(\theta)$ is the profit obtained by the deviating firm in the period of deviation and V^p is the firm's expected present value of profits along the punishment path. Although in principle the punishment path can take different forms (which may include return to collusion after some period of time), reversion to Nash-Cournot appears to us as more reasonable, particularly because of the fringe presence. Expression (12) adopts this view,[24] so $V^p = E_\theta[\pi^{nc}(\theta)/(1-\delta)]$.

It is important to notice that the direction of the deviation from the collusive agreement vary along the business cycle. If the deviation occurs sometimes during the output-contracting phase of the collusive agreement (i.e., when $\theta < \hat{\theta}$), the optimal deviation is to increase output towards Nash-Cournot levels. But if deviation occurs sometimes during the output-expanding phase (i.e., when $\theta > \hat{\theta}$), the optimal deviation is to reduce output below the collusive level. Since the remaining cartel firms as well as fringe firms also profit from this type of deviation in the short run, we see that in moments of output-expansion the cartel discipline does not come from the remaining cartel firms flooding the market, but from the loss of credibility that the cartel can effectively sustain its Stackelberg quantity. And if the fringe (correctly) anticipates the cartel deviation incentives from the Stackelberg outcome, it will rather play its Nash-Cournot quantity and so will each of the large firms.

We can now use (12) to obtain the discount factor function $\underline{\delta}(\theta) = (\pi^d(\theta) - \pi^m(\theta))/(V^m - V^p)$ that establishes the minimum discount factor needed to sustain maximal collusion at θ provided that maximal collusion is sustained at all other θ's. Then, the critical demand level θ^c at

[24] We also consider the optimal penal codes of Abreu [1986, 1988] and find no qualitative changes in our results.

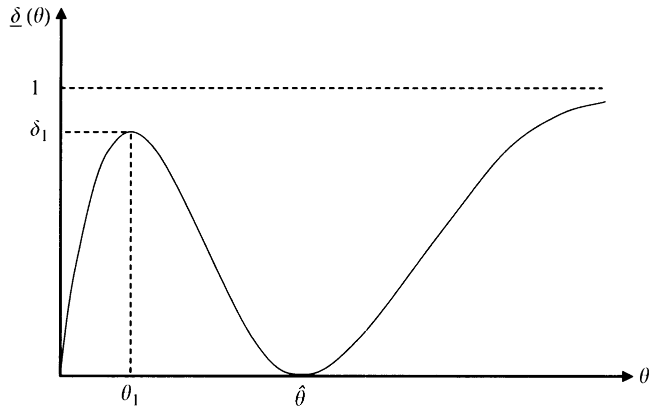

Figure 1
Critical Time for Maximal Collusion

which it becomes most difficult for firms to sustain maximal collusion can be defined as $\theta^c = \arg\max_\theta \underline{\delta}(\theta)$. In other words, firms can sustain maximal collusion throughout the business cycle, i.e., for all levels of demand, only if $\delta \geq \underline{\delta}(\theta^c)$.

To facilitate the exposition, let us adopt, for a moment, the simplifying assumptions of linear demand and costs, which allows us to obtain tractable expressions for $\pi^d(\theta)$ and $\pi^m(\theta)$. Solving, we obtain[25]

$$(13) \qquad \underline{\delta}(\theta) = \frac{\theta[(n-1)c - \theta b]^2}{(c+\theta b)^2} K$$

where $K = a^2/16n^2 b(V^m - V^p)$.

The function $\underline{\delta}(\theta)$ is plotted in Figure 1, which exhibits a local maximum at $0 < \theta_1 < \hat{\theta}$ and a global minimum at $\hat{\theta}$ – when maximal collusion reduces to the Nash-Cournot outcome. Note that $\underline{\delta}(\theta)$ has been drawn without paying attention to the fact that the support of θ is some subset $[\underline{\theta}, \overline{\theta}]$ of $\Re^+$.[26] Although both V^m and V^p depend on the actual support (and distribution) of θ, they enter as constant terms in (13), so changes in the support (and/or distribution) of θ will only scale $\underline{\delta}(\theta)$ up or down with no effects on the discussion that follows (e.g., θ_1 is independent of $\underline{\theta}$ and $\overline{\theta}$). More importantly, depending on the values of $\underline{\theta}$ and $\overline{\theta}$ one can construct cases in which the critical time is either at booms (e.g., $[\underline{\theta} = 0, \overline{\theta} = \theta_1]$, $[\underline{\theta} = \hat{\theta}, \overline{\theta} > \hat{\theta}]$)

[25] $\pi^d(\theta) = \theta b[q^d(\theta)]^2$, where $q^d(\theta) = a[c(n+1) + \theta b]/[4nb(c+\theta b)]$ is the optimal deviation when each of the remaining strategic firms are playing Q_s^m/n.

[26] Note also that $\underline{\delta}(\theta)$ converges to the unity as $\theta \rightarrow \infty$ because the output reduction of the deviating firm causes prices to explode, and with that, $\pi^d(\theta)$ (while $\pi^m(\theta)$ is bounded by the fringe presence).

or at recessions (e.g., $[\underline{\theta} = \theta_1, \overline{\theta} = \hat{\theta}]$). The latter example is most interesting because even if we restrict attention to output-contracting collusion, that is, $\overline{\theta} \leq \hat{\theta}$, we do not need to invoke Green and Porter's [1984] imperfect monitoring to generate procyclical pricing (i.e., prices closer to collusive levels at booms and non-collusive levels at recessions).[27]

The above discussion extends to the general model in that we can establish

Proposition 2. The time at which it is more difficult for large firms to sustain maximal collusion can be either at booms or recessions.

The above result hinges entirely on the fact that there exists a demand level $\hat{\theta}$ at which the collusive and the non-collusive outcomes are indistinguishable, so collusion for any nearby θ is easily sustained. As we increase the fringe's costs, $\hat{\theta}$ moves to the right (see Proposition 1) and eventually falls outside the support of θ. In the limit, when the fringe's costs are so high that its market share goes to zero, we return to Rotemberg and Saloner's [1986] prediction that in the absence of fringe firms, collusion is more difficult to sustain during booms (i.e., at the largest θ).[28]

IV(ii). *Best Collusive Agreement*

In this section we numerically illustrate for the simplified model how the optimal collusive agreement departs from maximal collusion when firms are not sufficiently patient, that is, when the discount factor δ is below the critical discount factor $\underline{\delta}(\theta^c)$ identified above. We also contrast the prediction of our model with that of Rotemberg and Saloner [1986]. Since the latter does not, at least explicitly, consider fringe suppliers, to provide a meaningful comparison with our oligopoly-fringe model we let large firms in the Rotemberg and Saloner formulation face the residual demand $P^{-1}(Q) - C_f'^{-1}(Q_f)$, or alternatively, the residual inverse demand

$$P_r(Q_s) = \theta_r(a - bQ_s)$$

where $\theta_r = \theta c/(c + \theta b)$ corresponds to the multiplicative demand shock in this alternative 'residual-demand' formulation and $Q_s = Q - Q_f$ is, as before, total output from large firms.[29] Note that specifying that the demand under Rotemberg and Saloner be equal to the residual demand implies that the two models produce the exact same maximal-collusion outcome. This does not

[27] The Bagwell and Staiger's [1997] model of serially correlated demand shocks is also able to generate procyclical pricing for some parameter values.

[28] Strictly speaking, Rotemberg and Saloner [1986] show that in a quantity-setting game (with no fringe) it is not always the case that collusion is more difficult to sustain at booms. It is the case though when demand and costs are linear. In fact, in the limiting case of no fringe eq. (13) reduces to $\lim_{c\to\infty} \underline{\delta}(\theta) = \theta(n-1)^2 K$, where $K = a^2/16n^2 b(V^m - V^p)$ and V^m and V^p correspond to the no-fringe values.

[29] Note that $d\theta_r/d\theta = c^2/(c + \theta b)^2 > 0$.

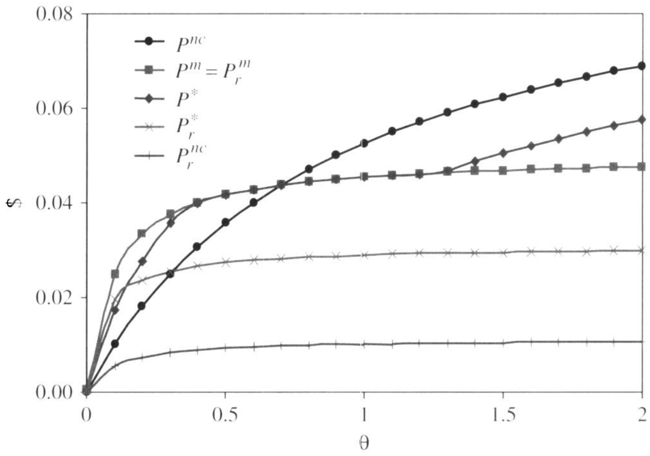

Figure 2
Best Collusive Agreement with Impatient Firms

mean, however, that the best achievable collusive agreements are the same since the Nash-Cournot outputs are not the same.

For the numerical illustration we use the following values: $n = 8$, $a = b = 1$, $\theta \sim U[0,2]$ and $c = 0.1$. Figure 2 depicts how prices vary with the demand shock θ under the two formulations and for the different solution concepts, i.e., maximal collusion (P^m and P^m_r), Nash-Cournot (P^{nc} and P^{nc}_r) and best collusive agreement for $\delta = 0.5$ (P^* and P^*_r), which is lower than the critical discount factor in either model. Consistent with Figure 1, curve P^* shows that impatient firms in the oligopoly-fringe model can sustain maximal collusion only for demand levels not too far from $\hat{\theta} = 0.7$(and obviously when $\theta \approx 0$ because there is nothing to collude on).

While consistent with the Rotemberg and Saloner [1986] prediction that maximal collusion is more difficult to sustain at higher levels of demand, the residual-demand model provides a remarkably different characterization of the best collusive agreement. This is because the residual-demand model imposes an artificial sequential timing (i.e., in each period large firms move before than fringe firms), which produces an unrealistic Nash-Cournot pattern. In fact, Nash-Cournot prices under the residual-demand model (P^{nc}_r) hardly react to demand shocks above $\theta = 0.5$. Furthermore, while in the residual-demand model the Cournot market share of the fringe remains very much constant at 10%, in the oligopoly-fringe model it is increasing in demand from 0 to 70% when $\theta = 2$.

This divergence can also be seen in the payoff that strategic firms can sustain according to either model. We know that the maximal collusion payoff is the same in either model, say, 100 (in present value). Even though the critical discount factor in the oligopoly-fringe model is higher than the critical discount factor in the residual-demand model -0.784 and 0.745, respectively – the oligopoly-fringe model predicts a much higher payoff under the best collusive agreement – 98.8 vs. 87.4. Part of the explanation is that in the oligopoly-fringe model the Nash-Cournot payoff is much closer to the maximal collusion payoff than in the residual-demand model – 90.6 and 39.5, respectively. In that sense, the predictions from the oligopoly-fringe model seem more consistent with Herfindahl's [1959] description of the operation of the 1935–39 copper cartel in that, while successful, it did not bring astronomical increases in profits for participating firms.

V. WELFARE: A NUMERICAL EXERCISE

One of the main implications of our results is that the effect of collusion on welfare is to be signed on a case-by-case basis. If Herfindahl's behavioral hypothesis is correct, the copper cartel of 1935–39 had an unambiguous negative impact on welfare. But this is not necessarily so if one believes the cartel was also able to sustain (output-expanding) collusion during booms. We explore this possibility with a numerical exercise. It is important to emphasize that the exercise does not attempt to uncover the exact behavior of the cartel during the 1935–39 period (we do not have the data to do that). We simply want to examine which of these two competing behavioral hypothesis 'fits' the data better: (i) the cartel was able to sustain maximal collusion in each of the six periods of Table I, and (ii) the cartel was able to sustain maximal collusion in the four 'Quotas' periods of Table I but returned to Nash-Cournot during the 'No Quotas' periods. Finding more support for hypothesis (i) could be interpreted as an indication that the cartel was indeed able to coordinate on output expansions during booms.

The idea of our exercise is to use the (aggregate) price and quantity data of the six periods of Table I to recover cost and demand parameters under the two behavioral hypothesis and then discuss how reasonable the estimates obtained for the different parameters are. In carrying out the exercise, we assume that (a) the five cartel members are identical, (b) the demand in period $t = 1, \ldots, 6$ is $P_t(Q_t;\theta_t) = \theta_t(a - bQ_t)$,[30] (b) the marginal cost function of each of the cartel members is $C_{st}(q_{st}) = c_{st}q_{st}^{\gamma}$, and (d) the fringe's marginal cost function is $C'_{ft}(Q_{ft}) = c_{ft}Q_{ft}^{\eta}$.

Under either behavioral hypothesis, there are 22 parameters to be estimated: a, b, θ_t, c_{st}, γ, c_{ft} and η for $t = 1, \ldots, 6$. Since we have only 12 equilibrium conditions for the estimation –two for each of the six periods–, we impose values upon a subset of the parameters.[31] We let $b = 0.7$ to work with demand elasticity numbers around -0.35; similar to those in Agostini

[30] We also consider the additive-shock model, i.e., $P_t(Q_t;\theta_t) = a - bQ_t + \theta_t$, with no qualitative changes in the results.

[31] We tried different subsets with no changes in the discussion that follows.

TABLE II
NON-COLLUSIVE PREDICTION FOR THE 1935–39 COPPER CARTEL

t	θ	c_f	c_s (i)	c_s (ii)	Q_s^{nc}	Q_f^{nc}	P^{nc}	P^{m}
1	0.23	2.00	2.19	2.19	679	337	7.6	8.2
2	0.27	2.41	2.31	2.31	739	290	8.6	9.5
3	1.00	3.16	1.44	0.93	847	498	14.0	13.4
4	0.49	2.22	1.75	1.75	798	457	9.5	9.7
5	1.36	2.38	1.12	− 1.47	624	800	12.8	10.6
6	0.51	2.22	2.02	2.02	743	508	9.9	10.0

Note: $a = 92.5$.

[2006] and the studies cited therein. In addition, we set $\gamma = \eta = 0.4$. We do not have a good reason to differentiate between γ and η and these numbers produce less variation among the c_{kt}'s ($k = s,f$) under either behavioral hypothesis, which we think should not vary much in a four-year period. Besides, lower numbers (e.g., $\gamma = \eta = 0.1$) produce the unrealistic scenario of output-expanding collusion at all periods under hypothesis (i) while higher numbers (e.g., $\gamma = \eta = 0.7$) result not only in wide variation among c_{kt}'s but also in some negative c_{ft}'s. We also normalize the demand shocks to the apparently largest shock, that is, $\theta_3 = 1$.

Estimates for the remaining demand and cost parameters (i.e., θ_t, a, c_{st}, c_{ft}) are reported in the first four columns of Table II. Note that the only differences between the estimates obtained under behavioral hypothesis (i) and (ii) are the costs of the large firms in the 'No Quotas' periods.[32] It is clear that hypothesis (i) leads to more reasonable estimates; only unexpectedly low and negative costs of strategic firms could support Nash-Cournot behavior during the 'No Quotas' periods.[33]

Finally, using the cost and demand estimates from assuming that the cartel was able to sustain maximal collusion throughout, the next three columns of Table II present counterfactual non-collusive predictions. As supported by our theory, the non-collusive prices are lower during recessions ($t = 1$, 2, 4 and 6) but higher during booms ($t = 1$ and 5). Furthermore, the average non-collusive price (weighted by the number of months in the period) is almost equal to the average collusive price (10.3 vs. 10.4). Provided that collusion leads to more stable prices and that consumer surplus is likely to increase with a mean-preserving contraction of prices,[34] it may well be that the copper cartel of 1935–39 did not have a negative impact

[32] This can be easily checked from looking at first-order conditions (3)–(4) and (7)–(8).

[33] Notice the variation of the cartel firms' cost parameters (i.e., c_s's), particularly the low numbers in $t = 3$ and 5, even under hypothesis (i). It may be that these low numbers reflect an asymmetric expansion with greater participation of lower cost firms.

[34] Because consumption is larger in booms, the loss in consumer surplus from a price increase during recessions is more than compensated by the gain from an equivalent price decrease during booms.

on welfare but the opposite. Obviously, this is just a reasonable hypothesis that has yet to be tested econometrically.

VI. FINAL REMARKS

Following the structure of many commodity markets, we have studied the properties of a collusive agreement when this is carried out only by the largest firms of the industry. We have found that as the (non-collusive) output of the noncartel firms expands, it may be optimal for the cartel to jointly produce above their non-collusive levels. Consequently, we cannot rule out, at least in theory, the possibility of a collusive agreement in which periods of output-contracting collusion are accompanied by periods of output-expanding collusion.

We also found that due to the presence of a significant fraction of noncartel firms (i.e., fringe firms), we do not need Green and Porter's [1984] imperfect information to generate procyclical pricing. More generally, it may be equally difficult for large firms to sustain maximal collusion during booms as during recessions. It would be, nevertheless, interesting to extend the model to the case of imperfect information.

There are other theoretical extensions worth pursuing. So far we have assumed that large firms have sufficient flexibility to expand production as needed. While this seems to be less of a problem for the international copper cartel of 1935–39 thanks to the excess capacity left by the 1929–33 world contraction,[35] the introduction of capacity constraints is likely to affect the properties of the collusive agreement (Staiger and Wolak [1992]). One can go even further and study altogether collusion in output and capacity (recall that in these markets firms are constantly expanding their capacities to cope with depreciation and new demand). This surely opens up the possibility for a capacity-expanding collusion even when firms set prices in the spot market. In addition to capacity constraints, the opportunity of forward contracting part of future production can also have implications for the collusive agreement (Liski and Montero [2006]).

Finally, it would be most interesting to carry out an empirical analysis of the copper cartel of 1935–39 along the works of Porter [1983] and Ellison [1994] for the JEC railroad cartel and test for periods of output-expanding collusion. One important difference with these previous studies is that we not only need to econometrically distinguish between regimes of (output-contracting) collusion and 'price wars' (i.e., return to Nash-Cournot) but perhaps more difficult between regimes of output-expanding collusion and price wars.

[35] It would also be less of a problem if large firms manage, as in mineral markets, for an in-house inventory to be built up during recessions and withdrawn during booms.

APPENDIX

A. *Proof of Proposition 1*

The first part is straightforward. For $Q^{nc} = Q^m$ we need $P_Q(\cdot)/n = C_f''(\cdot)P_Q(\cdot)/(C_f''(\cdot) - P_Q(\cdot))$, which rearranged leads to $(n-1)C_f''(\cdot) = -P_Q(\cdot)$. For the second part, we need to show that if we let θ to increase by a small amount from $\hat{\theta}$ to, say, θ', the term $C_f''/(C_f'' - P_Q)$ in (7) decreases more (recall that $P_Q < 0$) than the term $1/n$ in (3). If this is so, $Q_s^m(\theta')$ must be larger than $Q_s^{nc}(\theta')$ (and, hence, $Q^m(\theta')$ larger than $Q^{nc}(\theta')$) for both (3) and (7) to continue holding at θ'. Therefore, we need to show

$$(14) \qquad \frac{d}{d\theta}\left(\frac{C_f''(Q_f^m(\theta))}{C_f''(Q_f^m(\theta)) - P_Q(Q^m(\theta);\theta)}\right)\Bigg|_{\theta=\hat{\theta}} < 0$$

Totally differentiating, condition (14) reduces to

$$C_f'''(Q_f(\hat{\theta}))[C_f''(Q_f(\hat{\theta})) - P_Q(Q(\hat{\theta});\hat{\theta})]\frac{dQ_f^m(\hat{\theta})}{d\theta} + \\ - C_f''(Q_f(\hat{\theta}))\left[C_f'''(Q_f(\hat{\theta}))\frac{dQ_f^m(\hat{\theta})}{d\theta} - \frac{dP_Q(Q^m(\hat{\theta});\hat{\theta})}{d\theta}\right] < 0$$

Using $(n-1)C_f''(Q_f(\hat{\theta})) = -P_Q(Q(\hat{\theta});\hat{\theta})$ and rearranging lead to expression (10) in the text. *QED*

B. *Price-Setting Game*

Consider a group of n strategic firms ($i = 1, \ldots, n$) and a competitive fringe consisting of a continuum of firms (indexed by j) engaged in a simultaneous price-setting game of infinite horizon. Strategic firms produce differentiated goods at the same cost $C_s(q_{si})$. Fringe firms produce a homogenous good according to the aggregate marginal cost curve $C_f'(Q_f)$ (as before, a fringe firm's unit-cost is denoted by c_j). Strategic firm i's demand is $q_{si} \equiv D_{si}(p_{si}, \mathbf{p}_{-si}, p_f)$, where $\mathbf{p}_{-si}$ is the vector of prices charged by the remaining strategic firms and p_f is the price charged by all fringe firms (it should be clear that in equilibrium p_f will be equal to the unit-cost of the most expensive fringe firm that entered into production, that no fringe firm with a unit-cost equal or lower than p_f would want in equilibrium to charge anything different from this price, and that no firm with unit-cost higher than p_f would want to charge lower than p_f). Fringe aggregate demand is $Q_f = D_f(p_f, \mathbf{p}_s)$, where $\mathbf{p}_s = (p_{s1}, \ldots, p_{sn})$ is the vector of prices charged by strategic firms. It is also known that $\partial D_k/\partial p_k < 0$, $\partial D_k/\partial p_{\neq k} > 0$, and $|\partial D_k/\partial p_k| > |\partial D_k/\partial p_{\neq k}|$.

The (non-collusive) Nash-Bertrand equilibrium of the one-period game is obtained by simultaneously solving each firm's problem

$$\max_{p_i} D_{si}(p_{si}, \mathbf{p}_{-si}, p_f)p_i - C_s(D_{si}(p_{si}, \mathbf{p}_{-si}, p_f)) \quad \text{for all } i = 1, \ldots, n$$

$$p_{fj} = \begin{cases} p_f & \text{if } c_j \leq p_f \\ > p_f & \text{if } c_j > p_f \end{cases} \quad \text{for all } j$$

Then, the Nash-Bertrand equilibrium outcome is given by

$$(15)\qquad D_{si}(p_{si}^{nb}, p_{-si}^{nb}, p_f^{nb}) + [p_{si}^{nb} - C_s'(D_{si})]\frac{\partial D_{si}}{\partial p_{si}} = 0$$

$$C_f'(D_f(p_f^{nb}, \mathbf{p}_s^{nb})) = p_f^{nb}$$

On the other hand, and following Section III(i), the maximal collusive outcome for the strategic firms is obtained by solving

$$(16)\qquad \max_{p_{s1},\ldots,p_{sn}} \sum_{i=1}^{n} \{D_{si}(p_{si}, \mathbf{p}_{-si}, p_f(\mathbf{p}_s))p_i - C_s(D_{si}(p_{si}, \mathbf{p}_{-si}, p_f(\mathbf{p}_s)))\}$$

where $p_f(\mathbf{p}_s)$ is implicitly given by the fringe's equilibrium response

$$(17)\qquad C_f'(D_f(p_f, \mathbf{p}_s)) = p_f$$

Using the latter, the first-order conditions associated to the optimal collusive outcome can, after rearranging some terms, be written as

$$(18)\qquad \begin{array}{l} D_{si}(p_{si}^{m}, p_{-si}^{m}, p_f^{m}) + \displaystyle\sum_{k=1}^{n}[p_{si}^{m} - C_s'(D_{sk})]\left(\frac{\partial D_{sk}}{\partial p_{si}} + \frac{\partial D_{sk}}{\partial p_f}\frac{\partial p_f}{\partial p_{si}}\right) = 0 \\ \text{for all } i = 1, \ldots, n \end{array}$$

and $C_f'(D_f(p_f^m, \mathbf{p}_s^m)) = p_f^m$.

Given that $\partial p_f/\partial p_{si} > 0$, the difference between (18) and (15) is a stream of various positive terms (those price effects internalized in the collusive agreement); therefore, it is immediate that $D_{si}(p_{si}^m, p_{-si}^m, p_f^m) < D_{si}(p_{si}^{nb}, p_{-si}^{nb}, p_f^{nb})$ for all $i = 1, \ldots, n$, and with that, $D_f(p_f^m, \mathbf{p}_s^m) < D_f(p_f^{nb}, \mathbf{p}_s^{nb})$.

C. *Proof of Proposition 2*

We simply need to reproduce the relevant characteristics of Figure 1 for the general model. Using the no-deviation condition (12) and the definitions for $V^m = E_\theta[\pi^m(\theta)]/(1-\delta)$ and $V^p = E_\theta[\pi^{pc}(\theta)]/(1-\delta)$, the function $\underline{\delta}(\theta)$ becomes

$$\underline{\delta}(\theta) = \frac{\pi^d(\theta) - \pi^m(\theta)}{\pi^d(\theta) - \pi^m(\theta) + E_\theta[\pi^m(\theta)] - E_\theta[\pi^{nc}(\theta)]} = \frac{1}{1 + \Delta(\theta)}$$

where $\Delta(\theta) \equiv E_\theta[\pi^m(\theta) - \pi^{nc}(\theta)]/[\pi^d(\theta) - \pi^m(\theta)] > 0$. Notice again that the support of θ, i.e., $[\underline{\theta}, \overline{\theta}]$, will only scale $\Delta(\theta)$ up or down. For $\theta = 0$, we have that $\pi^d(\theta) = \pi^m(\theta) = 0$, so $\Delta(\theta) \to \infty$ and $\underline{\delta}(\theta) = 0$. For $\theta = \hat{\theta}$, we have that $\pi^d(\theta) = \pi^m(\theta) > 0$, so $\Delta(\theta) \to \infty$ and $\underline{\delta}(\theta) = 0$. For $0 < \theta < \hat{\theta}$, $\Delta(\theta) > 0$ and $0 < \underline{\delta}(\theta) < 1$; and there will be a demand shock $0 < \theta_1 < \theta$ associated to the local maximum δ_1. For $\theta > \hat{\theta}$, $\Delta(\theta) > 0$ and $0 < \underline{\delta}(\theta) < 1$. Finally, note that it is irrelevant for our Proposition whether $\underline{\delta}(\theta)$ converges to the unity as $\theta \to \infty$, as it does in the simplified model. *QED*

REFERENCES

Abreu, D., 1986, 'Extremal Equilibria of Oligopolistic Supergames,' *Journal of Economic Theory*, 39, pp. 191–225.

Abreu, D., 1988, 'On the Theory of Infinitely Repeated Games with Discounting,' *Econometrica*, 56, pp. 383–396.

Agostini, C., 2006, 'Estimating Market Power in the U.S. Copper Industry,' *Review of Industrial Organization*, 28, pp. 17–39.

Arant, W., 1956, 'Competition of the Few among the Many,' *Quarterly Journal of Economics*, 70, 327–345.

Ausubel, L. M. and Deneckere, R. J., 1989, 'Reputation in Bargaining and Durable Goods Monopoly,' *Econometrica*, 57, pp. 511–531.

Bagwell, K. and Staiger, R., 1997, 'Collusion over the Business Cycle,' *RAND Journal of Economics*, 28, 82–106.

Bulow, J. I.; Geanakoplos, J. D. and Klemperer, P. D., 1985, 'Multimarket Oligopoly: Strategic Substitutes and Complements,' *Journal of Political Economy*, 93, pp. 488–511.

Crowson, P., 1999, *Inside Mining* (Mining Journal Books Limited, London, U.K.).

Crowson, P., 2003, 'Mine Size and the Structure of Costs,' *Resources Policy*, 29, pp. 15–36.

Ellison, G., 1994, 'Theories of Cartel Stability and the Joint Executive Committee,' *RAND Journal of Economics*, 25, pp. 37–57.

Fudenberg, D. and Tirole, J., 1984, 'The Fat Cat Effect, the Puppy Dog Ploy and the Lean and Hungry Look,' *American Economic Review*, 74, pp. 361–368.

Fudenberg, D. and Levine, D., 1989, 'Reputation and Equilibrium Selection in Games with a Patient Player,' *Econometrica*, 57, pp. 759–778.

Fudenberg, D.; Kreps, D. and Maskin, E., 1990, 'Repeated Games with Long-Run and Short-Run Players,' *Review of Economic Studies*, 57, pp. 555–573.

Green, E. and Porter, R., 1984, 'Noncooperative Collusion under Imperfect Price Information,' *Econometrica*, 52, pp. 87–100.

Haltiwanger, J. and Harrington, J. E., 1991, 'The Impact of Cyclical Demand Movements on Collusive Behavior,' *RAND Journal of Economics*, 22, pp. 89–106.

Herfindahl, O. C., 1959, *Copper Costs and Prices: 1870–1957* (The John Hopkins University Press, Baltimore, Maryland, U.S.A.).

Liski, M. and Montero, J.-P., 2006, 'Forward Trading and Collusion in Oligopoly,' *Journal of Economic Theory*, 131, pp. 212–230.

Mailath, G. and Samuelson, L., 2006, *Repeated Games and Reputations: Long-Run Relationships* (Oxford University Press, New York, New York, U.S.A.).

Pindyck, R., 1979, 'The Cartelization of World Commodity Markets,' *American Economic Review*, 69, pp. 154–158.

Porter, R., 1983, 'A Study of Cartel Stability: The Joint Executive Committee, 1880–1886,' *Bell Journal of Economics*, 14, pp. 301–314.

Riordan, M., 1998, 'Anticompetitive Vertical Integration by a Dominant Firm,' *American Economic Review*, 88, pp. 1232–1248.

Rotemberg, J. and Saloner, G., 1986, 'A Supergame-Theoretic Model of Price Wars during Booms,' *American Economic Review*, 76, pp. 390–407.

Salant, S. W., 1976, 'Exhaustible Resources and Industrial Structure: A Nash-Cournot Approach to the World Oil Market,' *Journal of Political Economy*, 84, pp. 1079–1093.

Staiger, R. and Wolak, F., 1992, 'Collusive Pricing with Capacity Constraints in the Presence of Demand Uncertainty,' *RAND Journal of Economics*, 23, pp. 203–220.

Walters, A., 1944, 'The International Copper Cartel,' *Southern Economic Journal*, 11, pp. 133–156.