

Welfare-enhancing collusion in the presence of a competitive fringe

Juan-Pablo Montero and Juan Ignacio Guzmán*

May 11, 2006

Abstract

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 will never observe an output-expanding collusion in a price-setting game). We also find that the time at which maximal collusion is most difficult to sustain can be either at booms or recessions.

1 Introduction

In Table 1 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 output during the periods of low demand (denoted as Quota status and associated to lower spot prices in the London Metal Exchange) but failed to extend

*Montero (jmontero@faceapuc.cl) is Associate Professor of Economics at the Catholic University of Chile (PUC) and Research Associate at the MIT Center for Energy and Environmental Policy Research; Guzmán (jguzmanb@puc.cl) is PhD candidate at PUC. We thank Claudio Agostini, Jeremy Bulow, Joe Harrington, Bill Hogan, Jason Lepore, Salvador Valdés, Felipe Zurita and seminars participants for helpful discussions and comments. This work was completed while Montero was visiting the Harvard's Kennedy School of Government under a Repsol YPF-Harvard KSG Research Fellowship. Montero also thanks Fondecyt (Grant # 1051008) for financial support.

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 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 and discuss its welfare implications. 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 in which such collusive characterization may apply. There are basically two reasons for that. First, in these markets a firm's strategic variable is its level of production while prices are cleared, say, in a metal exchange. Second, collusive efforts, if any, are likely to be carried out by a fraction of the industry (typically, the largest firms) leaving, for incentive compatibility reasons, 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 often refer to the group of non-cartel firms as competitive fringe and to the group of potential cartel firms as strategic or large firms).⁴

It is important to make clear that the possibility of having an output-expanding collusion,

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

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

³From reading some news 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).

⁴Although the collusion literature usually assumes a structure of identical firms, this heterogenous structure in which a few large firms compete with many smaller firms has long been recognized (Arant, 1956; Pindyck, 1979).

and hence, lower prices, is totally unrelated to this idea that a cartel should prevent prices going too high as to induce the development of a substitute product (or the discovery of new mineral deposits) that can erode a fraction of the current demand.⁵ Besides that this can be easily added to our model by including a probability of discovery increasing in prices (more reasonably perhaps, in the average price of some period of time), our model is constructed simply upon the presence of a known fringe of small suppliers that run and shut down their production units so that at all moment the unit-cost of the marginal fringe firm is equal to the equilibrium price. In addition and consistent with practical observation, we assume that the entry or exit of large firms is a rare event.

Our results are entirely explained by market-interaction forces among existing players. As demand expands, fringe members increase output with no regard of the effect that such increase can have on the equilibrium price. Conversely, large firms do take into account the effect that their output decisions have on the equilibrium price, and consequently, they "limit" their Nash-Cournot output expansions as to accommodate for the fringe (price-taking) expansion. Because of the strategic substitution between the output of fringe firms and that of large firms (Bulow et al., 1985; Fudenberg and Tirole, 1984), it may indeed be optimal for the large firms to coordinate in a joint output expansion beyond their Nash-Cournot levels. The price drop caused by the strategic firms' over-expansion is more than offset by their market share increase.

Whether it is optimal for the large firms to implement such an output-expanding collusion as opposed to a traditional output-contracting collusion will ultimately depend on the fringe's (non-collusive) output, which in turn, will depend on cost differences between large and fringe firms and on the magnitude of the demand shocks. One can always find 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 that in which fringe costs generate both output-expanding collusion during booms and output-contracting collusion during recessions.

One way to appreciate these results more fully it is to contrast them with those obtained for a price-setting game with differentiated products.⁶ Think, for example, of a conventional Hotelling linear city in which there are two large supermarkets in each extreme of the city and

⁵It is not new in the literature that the introduction of a competitive fringe can alter existing results. Roridan (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.

⁶If products are perfectly homogeneous it is immediate that we can never have an output-expanding collusion.

a large number of small stores located downtown (i.e., in the middle of the city) competing in prices.⁷ Following the arguments above, one could conjecture that for a fringe sufficiently efficient that enjoys a large market share it may be optimal for the two supermarkets to coordinate on jointly pricing below their non-collusive (Nash-Bertrand) levels and expand their market share accordingly. It turns out this is never the case and the reason is the strategic complementarity between prices charged by the strategic firms and by the fringe firms. If large firms lower their prices, fringe firms' equilibrium response is to lower theirs preventing large firms from gaining market share.

The presence of an important fraction of non-cartel firms has also implications for cartel firms' ability to sustain the collusive outcome throughout the business cycle. Using the same i.i.d. demand shocks of Rotemberg and Saloner (1986), we show that it is no longer true that is more difficult for firms to sustain maximal collusion during booms than during recessions. Depending on fringe's costs relative to large firms' and on the possible realizations of demand, there will be cases in which it is more difficult for firms to sustain (maximal) collusion during periods of low demand.

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 have important policy 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.⁸ Based on the aggregate data of Table 1, 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 the next section (Section 2) we present the model and derive the (non-collusive) Nash-Cournot equilibrium. In Section 3 we present the maximal collusive equilibrium and demonstrate the possibility of an output-expanding collusion (we also demonstrate that the latter is never the case in a price-setting game). In Section 4, we study the cartel stability along the business cycle. The numerical exercise based on the copper

⁷Note that in equilibrium all fringe firms that are called to produce charge the same price which is equal to the unit-cost of the marginal fringe firm.

⁸It is interesting to contrast these observations with recent events in the copper industry. Back in 2001 when prices were at record low, the three largest firms in the industry made fairly simultaneous announcements of supply restrictions (either as lower production or inventory holdings) that were eventually carried out. Today, prices are at record high and some of the same firms are announcing very aggressive expansion plans. Obviously, in the absence of a more detailed empirical analysis we cannot tell whether these firms are engaged in some sort of tacit collusion or are rather optimally adjusting their supplies (including capacities and inventories) as they move from one non-collusive equilibrium to another. Nevertheless, this paper should prove useful in structuring such an empirical test.

cartel data is in Section 5. Concluding remarks follow.

2 Oligopoly-fringe model

2.1 Notation

A group of n identical (strategic) firms ($i = 1, \dots, n$) and a competitive fringe consisting of a continuum of firms (indexed by j) produce some commodity in an infinite-horizon setting. At the beginning of each period, firms simultaneously choose their production levels and the price is cleared (in a metal exchange) according to the inverse demand curve $P(\theta, Q) = \theta P(Q)$ with $P'(Q) < 0$, where Q is total production and $\theta \in [\underline{\theta}, \bar{\theta}]$ is, as in Green and Porter (1984), a multiplicative demand shock which is observed by all firms before they engage in production.⁹

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.¹⁰ 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 $C''_f(Q_f) > 0$ and $C'''_f(Q_f) \leq 0$,¹¹ where $Q_f = \int q_{fj} dj$ is fringe firms' 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 $P(Q) = a - bQ$, that strategic firms have no production costs and that the fringe' aggregate marginal cost curve is $C'_f(Q_f) = cQ_f$, where a , b and c are strictly positive parameters.¹²

⁹Below we explain how the results change under an additive shock (i.e., $P(Q, \theta) = P(Q) + \theta$).

¹⁰Including 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.

¹¹If anything, $C'_f(Q_f)$ should be strictly concave (i.e., $C'''_f(Q_f) < 0$) capturing the practical observation that fringe's output is more elastic at higher prices (e.g., Crowson, 1999 and 2003).

¹²That 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.

2.2 The (non-collusive) 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 Stackelberg timing within in 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 (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.¹³ Without such cooperation, the static game necessarily collapses into a simultaneous move game.¹⁴

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

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

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

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 or below the clearing price.

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

$$\theta P'(Q^{nc}) Q_s^{nc} / n + \theta P(Q^{nc}) - C'_s(Q_s^{nc} / n) = 0 \quad (3)$$

$$\theta P(Q^{nc}) = C'_f(Q_f^{nc}) \quad (4)$$

¹³Suppose 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.

¹⁴One of the first oligopoly-fringe models that explicitly adopts this simultaneous-move assumption is Salant's (1976) model for the oil market.

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 $Q_f^{nc}(\theta)$.¹⁵

3 Collusive equilibria

It is well known that in a 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 easy or difficult 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,¹⁶ in this section we are interested in finding the best collusive agreement for the strategic firms.

A natural point of departure in this simultaneous-move game is to compute what we call the static monopoly outcome that results from taking the group of large firms as a single player playing a one-period game against the fringe. The equilibrium is obtained from intersecting the static reaction (best-response) function of the group of fringe firms with the "static best-response function" of the group of large firms. Based on the works of Fudenberg and Levine (1989) and Fudenberg, Kreps and Maskin (1990) on repeated games with long-run and short-run players,¹⁷ one might argue that the static monopoly outcome is the best collusive agreement attainable for the group of large firms if fringe firms are thought to be short-run players that play only once and in addition have no means to learn about previous play. However, this is a poor characterization of smaller firms in most commodity markets. 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.

More importantly, when fringe firms observe previous play it is natural to think that the

¹⁵The 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 $\partial Q_s^{nc}(\theta)/\partial\theta < 0$ and $\partial Q_f^{nc}(\theta)/\partial\theta > 0$; but we do not require these properties in any of our Propositions.

¹⁶The discount factor of fringe firms is irrelevant since they always operate along their static best-response function.

¹⁷Fudenberg and Levine (1989) is not strictly a supergame as Fudenberg et al.(1990) and ours. They consider a single long-run player and introduce a bit of uncertainty about its type. It does not seem straightforward to us how to extend this incomplete information approach to the case of many long-run players each of them attempting to build reputation.

group of large firms can strictly improve upon the static monopoly outcome by following (equilibrium) strategies that are dependent on the possible histories of the game. In other words, the expectations of fringe firms as to what the large firms will produce are now sensitive to previous play of large firms which in turn allows the group of large firms to credibly communicate its commitment to (profitably) depart from the static monopoly outcome.¹⁸ We will refer to the most preferred of these dynamic outcomes as the maximal collusive agreement. Before describing its properties, we will present the static monopoly outcome because it will help us to more easily convey the intuition behind our main result.

3.1 Static monopoly outcome

The static best-response function of fringe firms,¹⁹ which we denote by the aggregate function $Q_f^*(Q_s)$, is implicitly given by $\theta P(Q_f + Q_s) = C'_f(Q_f)$. On the other hand, the static best-response function of the group/cartel of large firms to an aggregate play of Q_f by the fringe firms is given by

$$Q_s^*(Q_f) = \arg \max_{Q_s} \{ \theta P(Q_s + Q_f) Q_s - n C_s(Q_s/n) \}$$

Given a discount factor close enough to one, the stability of the cartel of large firms is subgame perfect and it is assumed here that this is correctly anticipated by fringe firms. Hence, the static monopoly outcome is obtained from the intersection of $Q_f^*(Q_s)$ and $Q_s^*(Q_f)$. Denoting by Q_s^0 and Q_f^0 the corresponding equilibrium quantities, we have

$$\theta P'(Q^0) Q_s^0 + \theta P(Q^0) - C'_s(Q_s^0/n) = 0 \tag{5}$$

$$\theta P(Q^0) = C'_f(Q_f^0) \tag{6}$$

¹⁸Unlike in Fudenberg and Levine (1989) where there is a single long-run player, here we have two or more large firms which allow them by the threat of falling into a price war, as in Gul (1987) for the durable-good duopolists, to build commitment despite fringe firms do not observe previous play. This would require, however, to impose fixed beliefs as to what fringe firms expect large firms to produce in each period. We could just assume that, but it is more natural to us to think that fringe firms' beliefs are not (exogenously) fixed but rather sensitive to past play.

¹⁹Note that since fringe firms are infinitesimally small they always play along their static best-response curve regardless of their life horizon.

where $Q^0 = Q_s^0 + Q_f^0$.²⁰ Solving we obtain the static-monopoly equilibrium quantities $Q_s^0(\theta)$ and $Q_f^0(\theta)$.²¹

By comparing equilibrium conditions (3)–(4) with (5)–(6), it holds that

Lemma 1 $Q_s^0(\theta) < Q_s^{nc}(\theta)$ and $P(Q^0(\theta)) > P(Q^{nc}(\theta))$ for all θ .

Proof. Straightforward since $P'(Q) < 0$ and $n \geq 2$. ■

This is the conventional view regarding the operation of a cartel in that it always reduces output to lift prices. In the presence of a competitive fringe there is a caveat, however. When the fringe's output is large enough, it may be optimal for the large firms to stick to the non-collusive equilibrium (and fringe firms anticipate that). From conventional monopoly theory, we would say that in such cases large firms do not want to restrict output any further because they face a too elastic (residual) demand.

This possibility can be easily illustrated for our simplified (i.e., linear) model. Let π_s^0 and π_s^{nc} be the strategic firm's profits under static monopoly and Nash-Cournot, respectively.²² It can be shown that

$$\pi_s^0(\theta) > \pi_s^{nc}(\theta) \iff \theta < (\sqrt{n} - 1)c/b \equiv \tilde{\theta} \quad (7)$$

Expression (7) indicates that for demand shocks sufficiently large it is optimal for large firms to follow Nash-Cournot strategies instead of static monopoly pricing. Note that the smaller the value of c and/or n the fewer the times at which large firms want to (statically) collude. The reason is that in this example large firms' (non-collusive) market share is increasing in n and c .²³

²⁰We have not been explicit about what happens when a large firm deviates from the (collusive) equilibrium path. A reasonable (subgame perfect) punishment path is a return to Nash-Cournot but with fringe firms playing, on aggregate, Q_f^0 (not Q_f^{nc}). This is because in this "one-shot game" fringe firms' beliefs are fixed in that they expect the large group to always play Q_s^0 . There are no deviations in equilibrium, however.

²¹For the simplified model these quantities are

$$Q_s^0 = \frac{ac}{b(2c + \theta b)} \quad \text{and} \quad Q_f^0 = \frac{\theta a}{2c + \theta b}$$

²²The profits for the simplified model are

$$\pi_s^{nc} = \frac{\theta}{b} \left[\frac{ac}{c(n+1) + \theta b} \right]^2 \quad \text{and} \quad \pi_s^0 = \frac{\theta}{nb} \left[\frac{ac}{2c + \theta b} \right]^2$$

²³That large firms may find it optimal to remain at their Nash-Cournot levels is not unique to the quantity competition assumption. In fact, in the Hotelling-city example of the introduction the Nash-Bertrand prices and the static monopoly prices are the same.

3.2 Maximal collusion

Assuming that all players acting at date t have observed the history of play up to date t , it is possible to construct history-dependent strategy profiles with payoffs for the large firms that are strictly higher than those under the static monopoly outcome (and easier to sustain). While the latter is one of the multiple possible equilibria in this repeated game, here we are interested in finding the maximal collusive agreement, that is, the one that gives large firms the highest attainable payoff for a discount factor close enough to one. Let Q_s^m and Q_f^m denote the (aggregate) quantities corresponding to the maximal collusive equilibrium.

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

$$Q_s^m = \arg \max_{Q_s} \{\theta P(Q_s + Q_f(Q_s))Q_s - nC_s(Q_s/n)\} \quad (8)$$

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

$$\theta P(Q_f + Q_s) = C'_f(Q_f) \quad (9)$$

Replacing $Q_f(Q_s)$ from (9) into (8), the strategic firms' maximal collusive outcome solves

$$\frac{C''_f(Q_f^m)\theta P'(Q^m)}{C''_f(Q_f^m) - \theta P'(Q^m)} Q_s^m + \theta P(Q^m) - C'_s(Q_s^m/n) = 0 \quad (10)$$

$$\theta P(Q^m) = C'_f(Q_f^m) \quad (11)$$

where $Q^m = Q_s^m + Q_f^m$. Solving we obtain the collusive equilibrium strategies $Q_s^m(\theta)$ and $Q_f^m(\theta)$.^{24,25}

Note that having Q_f as a function of Q_s in (8) resembles a Stackelberg (static) game but it is only because in this repeated game large firms anticipate and use fringe firms' equilibrium

²⁴For the simplified model the equilibrium quantities are

$$Q_s^m = \frac{a}{2b} \quad \text{and} \quad Q_f^m = \frac{\theta a}{2(c + \theta b)}$$

Note that $\partial Q_f^m(\theta)/\partial \theta > 0$.

²⁵We have been silent about the implementation of the (maximal) collusive agreement. We can think of the following set of (symmetric) trigger strategies (which depend on the realization of θ): In period 0, strategic firm i 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} .

response in constructing its optimal action and not because some first-mover advantage. By comparing equilibrium conditions (3)–(4) with (10)–(11) we can establish

Proposition 1 *There is a level of demand $\hat{\theta}$ for which $Q^m(\hat{\theta}) = Q^{nc}(\hat{\theta}) \equiv \hat{Q}(\hat{\theta})$, $Q_f^m(\hat{\theta}) = Q_f^{nc}(\hat{\theta}) \equiv \hat{Q}_f(\hat{\theta})$ and $Q_s^m(\hat{\theta}) = Q_s^{nc}(\hat{\theta}) \equiv \hat{Q}_s(\hat{\theta})$. This (unique) level of demand is found by replacing the definitions of $\hat{Q}(\hat{\theta})$ and $\hat{Q}_f(\hat{\theta})$ into $\hat{\theta} = -(n-1)C_f''(\hat{Q}_f(\hat{\theta}))/P'(\hat{Q}(\hat{\theta}))$ and solving. In addition, if θ is greater (lower) than $\hat{\theta}$, then Q^m is greater (lower) than Q^{nc} .*

Proof. See the Appendix ■

This proposition opens up the possibility for an output-expanding collusion provided that $\hat{\theta}$ exists, i.e., $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$, which will always be the case for $C_f''(\cdot)$ sufficiently small. More interesting, the collusive agreement may include both 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., $\theta < \hat{\theta}$). In providing more intuition for our results, it is useful to present the following result first

Proposition 2 *Were the strategic and fringe firms competing in prices (i.e., upward sloping reaction functions), it would have been never optimal for the strategic firms to jointly price below their (non-collusive) Nash-Bertrand price levels.*

Proof. See the Appendix. ■

Propositions 1 and 2 indicate that an output-expanding, or equivalently, a price-reducing collusion, is only a possibility under quantity competition and never under price competition. This observation can be understood as the balance of two effects that in price competition work in the same direction while in quantity competition work in opposite directions. We can think of the first effect as a static effect. In a one-shot simultaneous-move game the best collusive outcome for the group of large firms is the static monopoly outcome of Section 3.1 which always have large firms reducing output (increasing prices) below (above) their Nash-Cournot (-Bertrand) levels.

The second effect is a dynamic effect that comes from the fact that in a repeated game the group of large firms can use history-dependent strategies to credibly build commitment towards the implementation of a better outcome than the one-shot simultaneous-move outcome.²⁶ Unlike the static effect, the dynamic effect works in opposite directions depending on the type of

²⁶This is equivalent to the reputation effect of Fudenberg and Levine (1989) for the case of a single (long-run) player playing a simultaneous-move stage game against a sequence of short-run opponents (i.e., that play only once).

competition and is because of the strategic substitutability of quantities versus the strategic complementarity of prices (Bulow et al., 1985; Fudenberg and Tirole, 1984).

When firms compete in prices and large firms jointly lower (increase) their prices, fringe firms' equilibrium response is to lower (increase) theirs. Note from eq. (18) in the Appendix that $\partial p_f(p_{s1}, \dots, p_{sn})/\partial p_{si} > 0$ for all i , where p_f and p_{si} are, respectively, the prices charged by fringe firms and strategic firm i . Knowing this, the dynamic effect makes large firms to price even higher relative to the static-monopoly outcome. On the other hand, when firms compete in quantities and large firms jointly increase their productions, fringe firms' equilibrium response is not to increase their quantities but to reduce them (note from (9) that $Q'_f(Q_s) = \theta P'(Q^m)/[C''_f(Q^m) - \theta P'(Q^m)] < 0$). The dynamic effect now makes strategic firms to expand their production taking advantage of fringe's contracting response.²⁷

Whether it is optimal for strategic firms (competing in quantities) to follow an output-expanding collusion or not, at least for some levels of demand, will depend on whether the dynamic effect dominates the static effect. Provided that $\hat{\theta}$ exists, Proposition 1 establishes that for $\theta < \hat{\theta}$ the static effect dominates while for $\theta > \hat{\theta}$ the dynamic effect dominates. The reason why the dynamic effect dominates the static effect as demand raises is because fringe's (non-collusive) output is increasing in θ .²⁸ More interestingly, one can imagine situations in which a collusive agreement among the large firms can increase overall welfare as long as periods of output-contracting collusion are followed by periods of output-expanding collusion.²⁹

Let illustrate some of these results for our simplified model (i.e., $P(Q) = a - bQ$, $C_s(q_s) = 0$ and $C'_f(Q_f) = cQ_f$). From Proposition 1, the level of demand for which the optimal collusive outcome coincides with the Nash-Cournot outcome is $\hat{\theta} = (n-1)c/b$.³⁰ Consequently, a collusive agreement among the strategic firms will lead to higher output (and, hence, to lower prices)

²⁷To understand this further, think of a holding company that owns all of the large firms. The (Nash-Cournot) outcome of the one-period simultaneous-move game between this holding company and the fringe is the static monopoly outcome derived in Section 3.1. In a simultaneous-move repeated game and where fringe firms observe all previous play the holding company can do strictly better by committing to a larger (Stackelberg) quantity in all possible demand realizations (provided that the discount factor is close enough to one).

²⁸Note that for the existence of $\hat{\theta}$ we do not require that the fringe's (Nash-Cournot) market share be also increasing in θ . In fact, one can find situations in which the fringe's market share is sinking with θ , yet, the large firms want to implement an output-expanding collusion (we did so in the simplified model for c relatively large and $C_s(q_s) = c_s q_s^2$ with $c_s > 0$).

²⁹Bulow et al. (1986) 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.

³⁰Note from expression (7) that $\tilde{\theta} < \hat{\theta}$, which may call some readers' attention. The reason is that the static effect is still present beyond $\tilde{\theta}$ and is only at $\hat{\theta}$ when is exactly offset by the dynamic effect.

as long as $\theta > (n-1)c/b$; a condition that can be conveniently expressed in terms of Cournot market shares as follows (see footnote 15 for the values of Q_f^{nc} and Q_s^{nc})

$$\theta > \frac{(n-1)c}{b} \iff \frac{Q_f^{nc}}{Q_s^{nc}} > \frac{n-1}{2n-1} \quad (12)$$

When the fringe's market share in the Nash-Cournot equilibrium is larger than $(n-1)/(2n-1)$, the group of strategic firms faces such a elastic "residual" demand that it is optimal for them to coordinate in an output-expanding collusive agreement.

Since the fringe's market share is increasing in θ and decreasing in c , condition (12) is more likely to hold when the demand is large and the fringe is more efficient (i.e., has low c). For example, if $n = 2$ the collusive agreement will take the conventional output-contracting shape only when the fringe's market share is less than $1/3$. More generally, any time the fringe's Cournot market share is above $1/2$ the collusive agreement is to expand output above Cournot levels regardless of the number of firms. These results have also implications for understanding the shape that a collusive agreement can take among a subset of (heterogeneous) strategic firms. The fewer the strategic firms taking part in the collusive agreement the more likely it will contain at least some periods of (collusive) output expansion.

Before turning into the sustainability of the collusive agreement, there are two assumptions that deserve some discussion. First, if we relax the concavity assumption of the fringe's marginal cost (i.e., $C_f'''(Q_f) \leq 0$) and let it become mildly convex none of our results would change (see Proof of Proposition 1). If, however, we depart from practical observation and let the fringe's supply function $C_f'(Q_f)$ be highly inelastic above certain price level, then $\hat{\theta}$ may not exist.

Second, we have seen that with multiplicative demand shocks collusion can bring both output-contraction during recessions and output-expansion during booms. With additive demand shocks, i.e., $P(Q, \theta) = P(Q) + \theta$, the condition that separates output-contracting collusion from output-expanding collusion reduces to whether $-(n-1)C_f''/P'$ is greater or lower than the unity. 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 throughout the entire business cycle, which seems unpleasant. If, on the other hand, the elasticity of the fringe's supply function is increasing in prices (i.e., $C_f'''(Q_f) < 0$), we could again have a mixed collusive regime with output contraction for $\theta < \check{\theta}$ and output expansion for $\theta > \check{\theta}$, where $\check{\theta}$ solves $-(n-1)C_f''(Q_f(\check{\theta}))/P'(Q(\check{\theta})) = 1$. It is possible that in reality shocks are best modeled

as combination of both specifications.³¹

4 Collusion over the business cycle

We have characterized the maximal collusive agreement but said nothing on how difficult is for the strategic firms to sustain such an agreement under varying demand conditions. The question of whether 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 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's (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.

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

$$\pi^m(\theta) + \delta V^m \geq \pi^d(\theta) + \delta V^p \quad (13)$$

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 appear to us as most reasonable, particularly because of the fringe presence. Expression (13) adopts this view,³² 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 (i.e., move in the Nash-Cournot direction). 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. This invites us to speculate, at least in theory, about the possibility that a fringe firm

³¹See also Turnovsky (1976) for a technical discussion on the use of multiplicative shocks as opposed to additive shocks.

³²We also consider the optimal penal codes of Abreu (1986, 1988) and find no qualitative changes in our results.

could sabotage one of the strategic firms' production sufficiently enough as to force the latter to implement an (optimal) deviation and, hence, trigger a return to Nash-Cournot prices.³³

We can now use (13) 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.³⁴ Then, the critical demand level θ^c at which 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 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 will allow us to obtain tractable expressions for $\pi^d(\theta)$ and $\pi^m(\theta)$. Solving we obtain³⁵

$$\underline{\delta}(\theta) = \frac{\theta[(n-1)c - \theta b]^2}{(c + \theta b)^2} K \quad (14)$$

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 $[\underline{\theta}, \bar{\theta}]$ of \mathfrak{R}^+ . Despite both V^m and V^p depend on the actual support (and distribution) of θ , they enter as constant terms in (14), 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 $\bar{\theta}$). More importantly, depending on the values of $\underline{\theta}$ and $\bar{\theta}$ one can construct cases in which the critical time is either at booms (e.g., $[\underline{\theta} = 0, \bar{\theta} = \theta_1]$, $[\underline{\theta} = \hat{\theta}, \bar{\theta} > \hat{\theta}]$) or at recessions (e.g., $[\underline{\theta} = \theta_1, \bar{\theta} = \hat{\theta}]$). The latter example is most interesting because even if we restrict attention to output-contracting collusion, that is, $\bar{\theta} \leq \hat{\theta}$, we do not need invoke Green and Porter's (1984) imperfect monitoring to generate price wars at recessions.³⁶

³³Obviously, the incentive for a strategic firm's to sabotage the production of another strategic firm are always smaller than the incentives to deviate from the collusive agreement itself because the sabotage (regardless of whether it is detected or not by the affected firm) involves the simultaneous deviation of two firms which is always less profitable than the deviation of a single firm.

³⁴It is not difficult to show for the simple model that sustaining maximal collusion is easier than sustaining the static monopoly outcome. Since $\pi^m(\theta) > \pi^0(\theta)$, it suffices to show that $\pi^d(\theta)$ is smaller than the one-period profit from deviating from the monopoly outcome. After some algebraic manipulation we obtain that this holds for $\theta \leq (\sqrt{n} - 1)c/b$, but from Lemma 2 we know this is the relevant range.

³⁵ $\pi^d(\theta) = \theta b [q^d(\theta)]^2$, 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 .

³⁶The Bagwell and Staiger's (1997) model of serially correlated demand shocks is also able to generate price wars at recessions for some parameter values.

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

Proposition 3 *The time at which is more difficult for large firms to sustain maximal collusion can be either at booms or recessions.*

Proof. See the Appendix ■

The above result entirely hinges 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 fringe's costs, $\hat{\theta}$ moves to the right (see Proposition 1) and eventually falls outside the support of θ . In the limit, when fringe's costs are so high that its market share goes to zero, we return to Rotemberg and Saloner' (1986) prediction that in the absence of fringe firms collusion is more difficult to sustain during booms (i.e., at the largest θ).³⁷

5 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 collusion during booms. To illustrate this possibility, consider the following numerical exercise. Assume that the cartel was able to sustain maximal collusion in each of the six periods described in Table 1 and use the (aggregate) price and quantity data of each of those periods to recover cost and demand parameters. Then, use these parameters to predict what would have been the non-collusive equilibrium in each period.

In carrying out the exercise, we assume that (i) the five cartel members are identical, (ii) the demand in period $t = 1, \dots, 6$ is $P_t(Q_t, \theta_t) = \theta_t(a - bQ_t)$, (iii) the marginal cost function of each of the cartel members is $c_{st}q_{st}^\gamma$, and (iv) the fringe's marginal cost function is $c_{ft}Q_{ft}^\eta$. The parameters to be estimated are θ_t , a , b , c_{st} , γ , c_{ft} , η . We have more parameters than equilibrium equations so we are forced to make some (reasonable) arbitrary selections. We set $b = 0.7$ to work with demand elasticity numbers around -0.35 ; similar to those in Agostini (2006) and the studies cited therein. In addition, we set $\gamma = \eta = 0.4$. We do not have a good

³⁷Strictly 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. (14) reduces to $\lim_{c \rightarrow \infty} \underline{\delta}(\theta) = \theta(n-1)^2 K$, where $K = a^2/16n^2b(V^m - V^p)$ and V^m and V^p correspond to the no-fringe values.

reason to differentiate between γ and δ and these numbers produce less variation among the c_{kt} 's ($k = s, f$), which we think should not vary much in a four-year period. Besides, lower numbers (e.g., $\gamma = \eta = 0.1$) produce the unreasonably scenario of output-expanding collusion at all periods 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, $\theta_3 = 1$.

Results, which are merely for illustrative purposes and not aimed at testing hypothesis, are reported in Table 2. The next three columns following the period column show demand and costs parameters for the six periods.³⁸ In the fifth, sixth and seventh columns we reproduce the (assumed) maximal collusion levels of Table 1 (quantities are again at their annual rates) to facilitate the comparison with the hypothetical Nash-Cournot levels of the following three columns. As predicted by our theory, the non-collusive prices are lower during recessions ($t = 1, 2, 4$ and 6) but higher during booms ($t = 3$ 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 prices are less volatile and that a one cent off during booms add more to consumer surplus than a one cent off during recessions, it may well be that the copper cartel of 1935-39 did not have a negative impact on welfare but the opposite. Obviously, this is just an hypothesis that has yet to be tested econometrically.

6 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 welfare-enhancing 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 price wars in recessions (i.e., procyclical pricing). More generally, it may be equally difficult for large firms

³⁸Notice the variation of the cartel firms' cost parameters (i.e., c_s 's), particularly the low numbers in $t = 3$ and 5 . Besides indicating that even lower numbers (i.e., higher variation) would have resulted had we assumed return to Nash-Cournot during these two booms, it may be that these low numbers reflect an asymmetric expansion with greater participation of lower cost firms.

to sustain maximal collusion during booms than 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,³⁹ 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.

Our theory may even suggest to revisit the operation of JEC railroad cartel itself. By looking at Porter's (1983) estimations, it appears that price wars were more frequent during periods in which the Great Lakes were open to navigation, that is, periods in which the cartel faced a competitive fringe (i.e., lake steamers and sailships). If there was some strategic substitutability between the cartel's pricing decisions and the fringe's capacity decisions, one could argue that some of Porter's price wars (during navigation weeks) may end up reclassified as periods of low collusive pricing.

Appendix

Proof of Proposition 1: The first part is straightforward. For $Q^m = Q^{nc}$ we need $C_f''(\cdot)\theta P'(\cdot)/(C_f''(\cdot) - \theta P'(\cdot)) = \theta P'(\cdot)/n$ which rearranged leads to $\theta = -(n-1)C_f''(\cdot)/P'(\cdot)$.

³⁹It would also be less of a problem if large firms manage, as in mineral markets, an in-house inventory to be built up during recessions and withdrawn during booms.

For the second part, we need to show that if we are at $\theta = \hat{\theta}$ and let θ go up by a marginal amount, say, to θ' , the term $C_f''/(C_f'' - \theta P')$ in (10) suffers a greater fall (recall that $P' < 0$) 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 (10) to continue holding at $\theta' = \hat{\theta} + d\theta$. Therefore, we need to show

$$\frac{d}{d\theta} \left(\frac{C_f''(Q_f^m(\theta))}{C_f''(Q_f^m(\theta)) - \theta P'(Q^m(\theta))} \right) < 0 \quad (15)$$

Totally differentiating and rearranging, condition (15) reduces to

$$-C_f''' \cdot \theta P' \cdot (dQ_f^m(\theta)/d\theta) + C_f'' \cdot (P' + \theta P'' \cdot (dQ^m(\theta)/d\theta)) < 0$$

But since $C_f'''(Q_f) \leq 0$ by assumption, $P' + \theta P'' \cdot (dQ^m(\theta)/d\theta) \equiv d[\theta P'(Q^m(\theta))]/d\theta < 0$ (recall that $\lim_{\theta \rightarrow \infty} \theta P'(Q) = -\infty$), and both $dQ(\theta)/d\theta$ and $dQ_f(\theta)/d\theta$ are, by construction, positive for all θ , expression (15) holds, which in turn implies that $dQ^m(\theta)/d\theta|_{\theta=\hat{\theta}} > dQ^{nc}(\theta)/d\theta|_{\theta=\hat{\theta}}$.⁴⁰ It remains to show that $\hat{\theta}$ is unique. Suppose the contrary that there exists some $\check{\theta} \neq \hat{\theta}$ for which $Q^m(\check{\theta}) = Q^{nc}(\check{\theta})$. But if (15) holds for $\hat{\theta}$ it cannot simultaneously hold for $\check{\theta}$, which would be a contradiction.

Proof of Proposition 2: Consider a group of n strategic firms ($i = 1, \dots, n$) 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 than 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 $\mathbf{p}_s = (p_{s1}, \dots, p_{sn})$ 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 si-

⁴⁰It is easy to see that (15) holds when $C_f'(Q_f)$ and $P(Q)$ are linear.

multaneously solving each firm's problem

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

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

Then, the Nash-Bertrand equilibrium outcome is given by

$$D_{si}(p_{si}^{nb}, p_{-si}^{nb}, p_f^{nb}) + [p_{si}^{nb} - C'_s(D_{si})] \frac{\partial D_{si}}{\partial p_{si}} = 0 \quad (16)$$

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

On the other hand, and following Section 3.2, the maximal collusive outcome for the strategic firms is obtained by solving

$$\max_{p_{s1}, \dots, p_{sn}} \sum_{i=1}^n \{D_{si}(p_{si}, \mathbf{p}_{-si}, p_f(\mathbf{p}_s)) p_i - C_s(D_{si}(p_{si}, \mathbf{p}_{-si}, p_f(\mathbf{p}_s)))\} \quad (17)$$

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

$$C'_f(D_f(p_f, \mathbf{p}_s)) = p_f \quad (18)$$

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

$$D_{si}(p_i^m, p_{-i}^m, p_f^m) + \sum_{k=1}^n [p_{si}^m - C'_s(D_{sk})] \left(\frac{\partial D_{sk}}{\partial p_{si}} + \frac{\partial D_{sk}}{\partial p_f} \frac{\partial p_f}{\partial p_{si}} \right) = 0 \quad \text{for all } i = 1, \dots, n \quad (19)$$

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 (19) and (16) is a stream of various positive terms (those price effects internalized in the collusive agreement); therefore, it is immediate that $D_{si}(p_i^m, p_{-i}^m, p_f^m) < D_{si}(p_i^{nb}, p_{-i}^{nb}, p_f^{nb})$ for all $i = 1, \dots, n$, and with that, $D_f(p_f^m, \mathbf{p}_s^m) < D_f(p_f^{nb}, \mathbf{p}_s^{nb})$. Note that in the Hotelling example of the introduction, the strategic firms do not directly face each other, so $\partial D_{s1} / \partial p_{s2} = \partial D_{s2} / \partial p_{s1} = 0$.

Proof of Proposition 3: We simply need to reproduce the relevant characteristics of Figure 1 for the general model (whether, for example, the global maximum for the function $\underline{\delta}(\theta)$ is to

the right or left of $\hat{\theta}$ is not relevant for the validity of the proposition). Using the no-deviation condition (13) and the definitions for $V^m \equiv E_\theta[\pi^m(\theta)]/(1 - \delta)$ and $V^p \equiv E_\theta[\pi^{nc}(\theta)]/(1 - \delta)$, 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., $[\underline{\theta}, \bar{\theta}]$, will only scale $\Delta(\theta)$ up or down. For $\theta = 0$, we have that $\pi^d(\theta) = \pi^m(\theta) = 0$, so $\Delta(\theta) \rightarrow \infty$ and $\underline{\delta}(\theta) = 0$. For $\theta = \hat{\theta}$, we have that $\pi^d(\theta) = \pi^m(\theta) > 0$, so $\Delta(\theta) \rightarrow \infty$ 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 $0 < \underline{\delta}(\theta) < 1$. However, $\underline{\delta}(\theta)$ does not converge to the unity because $\pi^m(\theta)$, and hence $\pi^d(\theta)$, are bounded by the fringe presence.⁴¹

References

- [1] Abreu, D. (1986), Extremal equilibria of oligopolistic supergames, *Journal of Economic Theory* 39, 191-225.
- [2] Abreu, D. (1988), On the theory of infinitely repeated games with discounting, *Econometrica* 56, 383-396.
- [3] Agostini, C. (2006), Estimating market power in the US Copper Industry, *Review of Industrial Organization* 28, 17-39.
- [4] Arant, W. (1956), Competition of the few among the many, *Quarterly Journal of Economics* 70, 327-345.
- [5] Bagwell, K. and Staiger, R. (1997), Collusion over the Business Cycle, *RAND Journal of Economics* 28, 82-106.
- [6] Bulow, J.I., Geanakoplos, J.D. and Klemperer, P.D. (1985), Multimarket oligopoly: strategic substitutes and complements, *Journal of Political Economy* 93, 488-511.
- [7] Crowson, P. (2003), Mine size and the structure of costs, *Resources Policy* 29, 15-36.

⁴¹Let \bar{Q} , where $P(\bar{Q}) = 0$, be the (totally inelastic) demand as θ approaches infinity. The equilibrium price cannot go above \bar{P} , where $\bar{P} = C'_f(\bar{Q})$.

- [8] Crowson, P. (1999), *Inside Mining*, Mining Journal Books Limited, London, UK.
- [9] Ellison, G. (1994), Theories of Cartel Stability and the Joint Executive Committee, *RAND Journal of Economics* 25, 37-57.
- [10] Green, E. and Porter, R. (1984), Noncooperative collusion under imperfect price information, *Econometrica* 52, 87-100.
- [11] Fudenberg, D. and Tirole, J. (1984), The fat cat effect, the puppy dog ploy and the lean and hungry look, *American Economic Review Papers and Proceedings* 74, 361-368.
- [12] Fudenberg, D. and Levine, D. (1989), Reputation and equilibrium selection in games with a patient player, *Econometrica* 57, 759-778.
- [13] Fudenberg, D., D. Kreps and E. Maskin (1990), Repeated games with long-run and short-run players, *Review of Economic Studies* 57, 555-573.
- [14] Gul, F. (1987), Noncooperative collusion in durable goods oligopoly, *RAND Journal of Economics* 18, 248-254.
- [15] Haltiwanger, J., and J.E. Harrington (1991), The impact of cyclical demand movements on collusive behavior, *RAND Journal of Economics* 22, 89-106.
- [16] Herfindahl, O. (1959), *Copper Costs and Prices: 1870-1957*, The John Hopkins University Press, Baltimore, Maryland.
- [17] Liski, M. and Montero, J.-P. (2006), Forward trading and collusion in oligopoly, *Journal of Economic Theory*, forthcoming.
- [18] Pindyck, R. (1979), The cartelization of world commodity markets, *American Economic Review Papers and Proceedings* 69, 154-158.
- [19] Porter, R. (1983), A study of cartel stability: The Joint Executive Committee, 1880-1886, *Bell Journal of Economics* 14, 301-314.
- [20] Riordan, M. (1998), Anticompetitive vertical integration by a dominant firm, *American Economic Review* 88, 1232-1248.
- [21] Rotemberg, J. and Saloner, G. (1986), A supergame-theoretic model of price wars during booms, *American Economic Review* 76, 390-407.

- [22] Salant, S.W. (1976), Exhaustible resources and industrial structure: A Nash-Cournot approach to the world oil market, *Journal of Political Economy* 84, 1079-1093.
- [23] Staiger, R. and Wolak, F. (1992), Collusive pricing with capacity constraints in the presence of demand uncertainty, *RAND Journal of Economics* 23, 203-220.
- [24] Turnovsky, S. (1976), The distribution of welfare gains from price stabilization: The case of multiplicative disturbances, *International Economic Review* 17, 133-148.
- [25] Walters, A. (1944), The International Copper Cartel, *Southern Economic Journal* 11, 133-156.

Table 1. Evolution of the copper cartel of 1935-1939

Period	Quota status	Cartel production		Noncartel production		London spot price	
		Production (annual rate)	Per cent change from preceding period	Production (annual rate)	Per cent change from preceding period	Cents per pound	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)

Table 2. Collusive and non-collusive equilibria for the copper cartel of 1935-39

t	θ	c_s	c_f	Q_s^m	Q_f^m	P^m	Q_s^{nc}	Q_f^{nc}	P^{nc}
1	0.23	2.19	2.00	559	408	8.2	679	337	7.6
2	0.27	2.31	2.41	598	368	9.5	739	290	8.6
3	1.00	1.44	3.16	917	439	13.4	847	498	14.0
4	0.49	1.75	2.22	762	481	9.7	798	457	9.5
5	1.36	1.12	2.38	948	504	10.6	624	800	12.8
6	0.51	2.02	2.22	736	511	10.0	743	508	9.9

Note: $a = 92.5$.

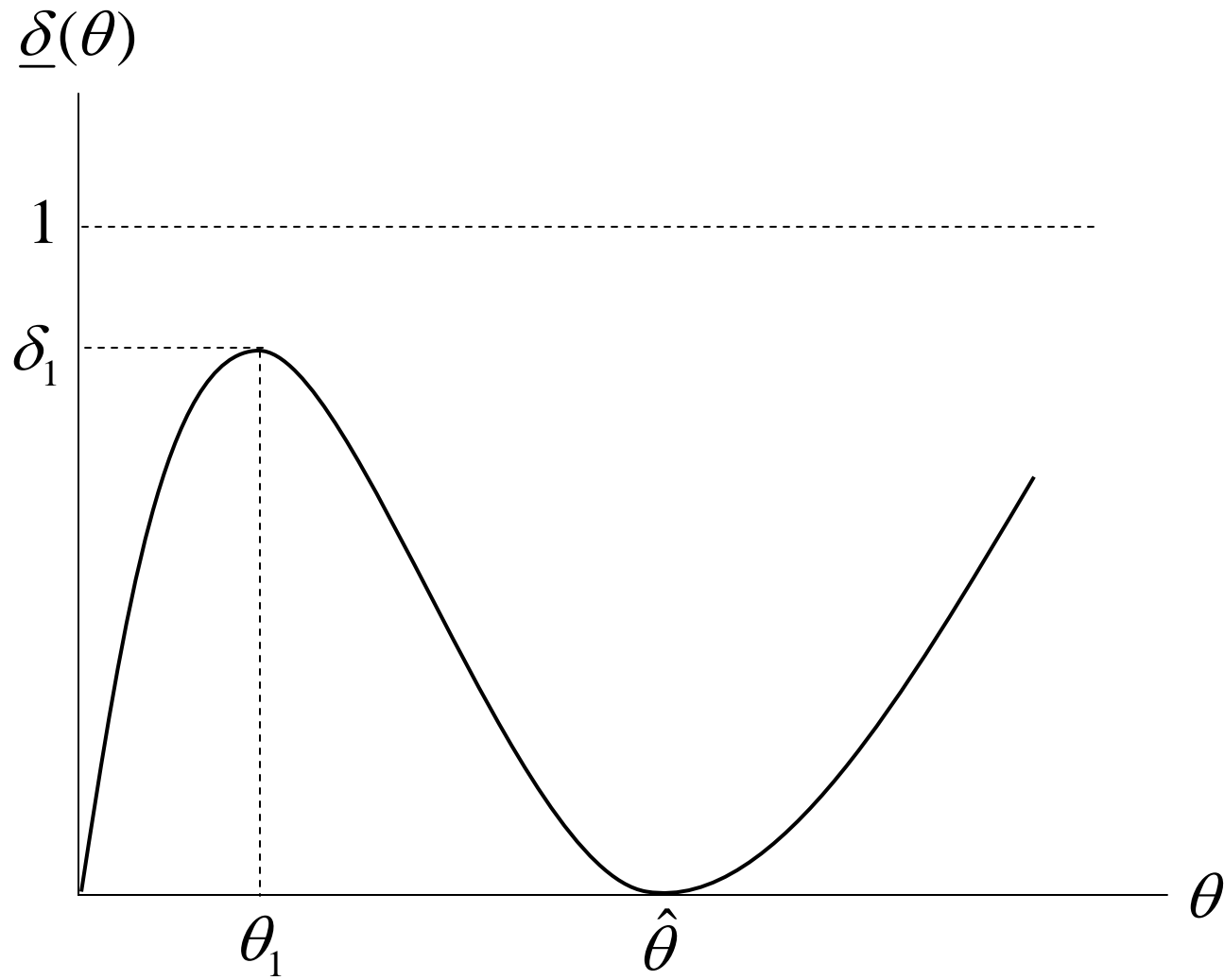


Figure 1. Critical time for maximal collusion