INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2005-36

Optimal Collusion under Cost Asymmetry

J. THAL[1]

[1] University of Toulouse 1 and CREST-LEI, 28 rue des Saints-Pères, 75007 Paris.
I am grateful, above all, to my Ph.D. advisor Patrick Rey.
This paper has also benefited from comments by Mohamed Belhaj, Bruno Jullien, Ulrich Kamecke, Lucy White, Joe Harrigton, Andreas Blume, Dan Kovenock, and participants of the 2004 Enter Jamboree, the 2004 EARIE Meeting, the 2004 WZB/CEPR "Collusion and Cartels" conference, the Econometric Society World Congress 2005, and the European Winter Meeting of the Econometric Society 2005. Any remaining errors are my own.
The author can be contacted for comments at jeaninethal@gmail.com

Optimal Collusion under Cost Asymmetry

Jeanine Thal*

December 18, 2005

*University of Toulouse 1 and CREST-LEI. I am grateful, above all, to my Ph.D. advisor Patrick Rey. This paper has also benefitted from comments by Mohamed Belhaj, Bruno Jullien, Ulrich Kamecke, Lucy White, Joe Harrington, Andreas Blume, Dan Kovenock, and participants of the 2004 Enter Jamboree, the 2004 EARIE meeting, the 2004 WZB/CEPR "Collusion and Cartels" conference, the Econometric Society World Congress 2005, and the European Winter Meeting of the Econometric Society 2005. Any remaining errors are my own. The author can be contacted for comments at jeaninethal@gmail.com.

Abstract:

Cost asymmetry is generally thought to hinder collusion because a more efficient firm has both less to gain from collusion and less to fear from retaliation. Our paper reexamines this conventional wisdom and characterizes optimal collusion without any prior restriction on the class of strategies. We first stress that firms can "collude" on retaliation schemes that maximally punish even the most efficient firm. This implies that whenever collusion is sustainable under cost symmetry, some collusion is also sustainable under cost asymmetry; efficient collusion, however, remains more difficult to sustain when costs are asymmetric. Finally, we show that, in the presence of side payments, cost asymmetry generally facilitates collusion.

Résumé:

L'asymétrie de coût est généralement perçue comme un facteur qui empêche la collusion car une firme efficace a moins à gagner de la co-opération et moins à craindre d'une punition. Cet article réexamine le rôle de l'asymétrie des coûts en caractérisant la collusion optimale sans aucune restriction a priori sur les stratégies considérées. Nous soulignions que les firmes peuvent se mettre d'accord sur un schéma de punitions maximales pour toutes les firmes, y compris la plus efficace. Ceci implique que lorsque la collusion est soutenable entre firmes symétriques, il existe des accords de cartel soutenables entre firmes asymétriques ; néanmoins, l'asymétrie de coût rend la collusion efficace plus difficile. Finalement, nous montrons qu'en présence de transferts monétaires entre firmes l'asymétrie de coût facilite la collusion en général.

1 Introduction

Economists and policymakers generally agree that cost asymmetry hinders collusion. In his classical industrial organization textbook for example, Scherer (1980) states that "...the more cost functions differ from firm to firm, the more trouble firms will have maintaining a common price policy". The US Merger Guidelines refer to some of the underlying arguments for this conventional wisdom when stating that "...the extent of homogeneity may be relevant both for the ability to reach terms of coordination and to detect or punish deviations from those terms".

There are three main reasons why cost asymmetry is thought to hinder collusion.[1] Firstly, coordination problems are obviously more complex when firms have divergent preferences concerning collusive prices and there are no natural focal points. Secondly, it may be difficult to convince an efficient firm to join a cartel, since it earns relatively high profits even under competition. Thirdly, cost asymmetry may also hinder the *sustainability* of collusion, since (i) it may be more difficult to retaliate against an efficient firm in case it deviates from the cartel agreement, and (ii) a more efficient firm may gain relatively more from deviating in the short-term.

This paper focuses on the sustainability of collusion, thus largely ignoring issues of coordination and participation.[2] Our aim is to analyze the maximum scope for collusion. Threats of severe retalitation against cheating firms are clearly optimal for cartel stability, since they reduce deviation incentives and thereby facilitate co-operation. Of course, punishment threats must be credible to be effective. We illustrate that it is possible to design tough punishments even for the most efficient firm: firms can "collude" on punishments that leave the cheating firm with minmax continuation profits, no matter whether the deviator has high or low costs. Thus, cost asymmetry weakens retaliation only if there is some reason why firms should use standard trigger strategies or other

[1] See Ivaldi et al. (2003) for an overview of the different arguments.

[2] Schmalensee (1987) applies a variety of selection criteria to model the choice of price and output quotas by an asymmetric cartel, subject to the constraint that each firm is at least as well off as without collusion. His paper, however, does not explicitly examine whether a selected outcome is also sustainable.

We, on the other hand, apply standard repeated game theory, which is well-suited to study sustainability but not so much coordination or participation issues. In repeated games, there is generally a large multitude of equilibria and no uncontested method to select one of these. Firms might even be "locked" into a bad equilibrium, in which some firm's profits are lower than in the standard competitive equilibrium. Nevertheless, our analysis gives some guidance as to which equilibria firms may reasonably select, since we characterize the Pareto frontier of all sustainable outcomes.

restricted forms of punishments instead of these maximal punishments.

Whether a more efficient firm nonetheless has relatively stronger incentives to deviate depends on the level of the collusive price. Suppose that the cartel members have different (constant) marginal costs, but produce perfect substitutes and split the market equally. If the price is equal to the most efficient firm's monopoly price, then by undercutting the collusive price each firm can improve its short-term profits by the same proportion. When there are two firms for example, each firm can achieve a short-term gain of 50%. Short-term deviation incentives are hence symmetric. When punishments are indeed maximal for all firms, collusion is thus sustainable whenever it is sustainable under cost symmetry.

This conclusion differs from those in the previous literature that has focused on grim trigger strategies.[3] Bae (1987) as well as Harrington (1991), whose frameworks very closely ressemble ours, determine the set of prices and output quotas sustainable by grim trigger strategies. They find that cost asymmetry always hinders collusion, even when allowing for inefficient allocations from the viewpoint of the cartel.

When the focus is on efficient collusion, on the other hand, our qualitative results are in line with the previous literature. Unless only the most efficient firm produces, static efficiency requires a price above the most efficient firm's monopoly price. For such prices, however, the most efficient firm has a disproportionally high deviation gain: it not only gains market share, but also switches to its profit-maximizing price. Firms with higher monopoly prices have relatively less to gain from a deviation. Collusion on a (statically) efficient allocation will thus be more difficult to sustain under cost asymmetry than under cost symmetry.

Another contribution of our paper is to analyze the role of side transfers. As Bain (1948) has argued more than 50 years ago, if firms have different marginal costs, the maximization of industry profits by a cartel requires side payments: without transfers, some production must be allocated to high-cost firms to induce their compliance. While antitrust rules typically prohibit direct transfers, there is evidence that some (illegal) cartels nevertheless use illegal payments. In the Florida bid rigging scheme for providing school milk, for example, dairies used side payments to compensate cartel members for refraining from bidding.[4] Other cartels used compensation schemes to discourage cheating. In

[3]As far as we are aware, the only authors who have used optimal punishments in a similar framework are Bernheim & Whinston (1990); since their focus is on a different issue, however, they do not characterize these punishments in detail or analyze the impact of cost asymmetry on sustainability.

[4]See Pesendorfer (2000).

the worldwide lysine cartel, for example, firms with realized market shares above their allotments had to compensate the other firms through inter-firm sales.[5] In the New York trash haulers cartel, "[an] undercover police detective posing as a carting executive paid more than $790,000 in "dues" to the [trash haulers'] associations and in compensation to other carters".[6]

Our analysis confirms that side payments facilitate collusion between asymmetric firms; more suprisingly, it also shows that cost asymmetry generally facilitates collusion when side payments are feasible. The latter result runs completely counter to the conventional wisdom on the impact of cost asymmetries. Side transfers allow firms to increase the total pie by allocating more production to the most efficient firm without inducing a deviation by a less efficient firm. In a way, firms agree on a mutually beneficial scheme of compensation payments for being inactive.

On the theoretical side, our paper relates to the literature on collusion under various forms of cost asymmetry. In the existing literature, the authors often either choose to or are bound to impose some restrictions on the strategies considered. Rothschild (1999) and Vasconcelos (2005) both deal with collusion under cost asymmetry when firms compete à la Cournot. Rothschild uses standard grim trigger strategies. Vasconcelos looks for more general punishments in the class of equilibria with proportional market shares on all equilibrium paths; he shows that optimal punishments, with a stick-and-carrot structure as proposed by Abreu (1986, 1988), exist within this restricted class of equilibria. For a limited range of parameters, these punishments are also maximal and would thus be optimal even without any restrictions.

In the related literature on collusion with asymmetric capacity constraints where firms compete in prices, the characterization of optimal punishments is unfortunately quite difficult. While Lambson (1987) shows that optimal punishments exist in models with symmetric capacity constraints, Lambson (1994) provides only a partial characterization in the asymmetric case. The impact of asymmetry in capacities on collusive sustainability was studied by Davidson & Deneckere (1990) in the context of grim trigger strategies. Compte, Jenny & Rey (2000) extend this analysis and allow for harsher punishments, but again restrict attention to a particular class of equilibria where market shares along any punishment path are the same as under collusion.

[5]See Hammond (2005). Similar compensation schemes were also employed in the citric acid cartel (see European Commission (2001)), or the sodium gluconate cartel (see European Commission (2002)).

[6]See Porter (2004) for this citation from the *New York Times*, June 23, 1995.

Lambson (1995) allows for small asymmetries in marginal costs as well as in capacity constraints and discount rates. He shows that, in nearly symmetric Bertrand games, optimal punishments leave firms with minmax discounted profits. Our paper generalizes this result to all situations in which the only dimension of heterogeneity is in marginal costs.

Our analysis proceeds as follows. Section 2 sets out the framework. Section 3 discusses optimal punishments in models of repeated price setting when firms have asymmetric costs. Section 4 deals with stationary collusion without side payments (the appendix contains a proof that justifies the focus of our analysis on stationary collusion). We first derive the set of all sustainable collusive outcomes as a function of the discount factor. Next, we restrict attention to efficient collusion. Here, we distinguish between stationary collusion on a statically Pareto-efficient outcome, and fully efficient collusion on a lottery. We also derive the Pareto frontier of sustainable allocations, i.e. the subset of Pareto undominated allocations within the set of all sustainable allocations. In section 5, we allow for side payments. We again derive the set of all sustainable stationary allocations, and discuss the difference with the previous results. Section 6 concludes.

2 Framework

We consider a simple model of infinitely repeated Bertrand competition between $n \geq 2$ firms indexed by $i = 1, 2, ..., n$. Entry by other firms is blockaded; it may, however, happen that not all n firms indeed sell in equilibrium. Firms produce perfect substitutes, but face different constant marginal costs of production $c_1 < c_2 < ... < c_n$. The demand function for the good, $D(p)$, is continuous and strictly downward sloping with $\lim_{p \longrightarrow 0} D(p) = \infty$.

The monopoly profit function of firm i is:

$$\pi_i(p) = (p - c_i)D(p).$$

These functions are assumed to be twice differentiable and strictly concave. We denote the unique monopoly price of firm i by p_i^m. A standard argument then ensures that $p_1^m < p_2^m < ... < p_n^m$.

Unless stated otherwise, we assume that the cost advantage of the most efficient firm 1 compared to firm 2 is non-drastic:

$$p_1^m > c_2.$$

In this set-up, we analyze the subgame perfect equilibria of a noncooperative supergame where firms simultaneously set prices in every period. The vector of prices in period t is denoted by $P^t = (p_1^t, p_2^t, ..., p_n^t)$. Since the firms produce perfect substitutes, the whole demand in any period goes to the lowest price firm(s). In case of a price tie at the lowest price, consumers are indifferent between a number of sellers and we may split demand between all firms that charge the lowest price in any way consistent with the equilibrium.[7] Firms have to serve the entire demand at any given price. The market sharing rule in period t is hence a vector $s^t = (s_1^t, s_2^t, ..., s_n^t)$ such that $s_i^t \in [0,1]$ for all i, $\sum_i s_i^t = 1$, and $s_i^t = 0$ if $p_i^t \notin \min_{j\in[1,n]}\{p_j^t\}$. Each firm aims to maximize its discounted stream of profits, where the (common and constant) discount rate is $\delta \in (0,1)$. Prices are publicly observable, and firms have perfect memory; they can thus condition their actions on past prices.

3 Minmax Punishments

For tacit collusion to be succesful, firms need to agree on some credible retaliation mechanism to deter profitable short-term deviations. The scope for collusion is greatest if a firm that deviates from the collusive agreement is punished as harshly as possible. By the same logic, it is easiest to punish a firm if deviations from the prescribed punishment are retaliated against as severely as possible.

The minmax of each firm's profit is zero in our model: while a firm can always avoid negative profits by charging a price above its marginal cost, any other firm can drive its profits down to zero by undercutting its price. The most severe punishments that can possibly be imposed on a deviator thus have a continuation value of zero. Obviously, if firms are able to credibly "collude" on punishment strategies such that any deviation by a particular firm triggers a minmax punishment for this firm, then those punishment strategies are optimal, since they maximize the scope for collusion.[8]

Under cost symmetry, it is easy to punish all firms maximally by standard grim trigger

[7] When firms are symmetric, it seems rather natural to assume that in case of a tie the market is split symmetrically. When firms are asymmetric, however, there is no similarly convincing assumption, and any arbitrary restriction may rule out interesting equilibria. See Simon & Zame (1990) for a discussion of the motivations to endogenize sharing rules in discontinuous games.

[8] We focus on punishment strategies such that *any* deviation by a particular firm, be it from collusion or from a punishment already in play, triggers the same punishment. Abreu (1988) shows that this does not imply any loss of generality. If several firms deviate simultaneously, no punishment is started.

strategies: firms simply agree to revert to the competitive equilibrium for ever upon any deviation. Under cost asymmetry, the conventional competitive equilibrium is such that consumers pay a price c_2, and the low-cost firm serves the entire demand earning $\pi_1(c_2) > 0$.[9] Standard grim trigger strategies hence provide minmax punishments for all firms except the most efficient one, which explains the conventional wisdom that retaliation against an efficient firm is difficult.

As we will show, however, it is rather straightforward to design self-enforcing agreements such that a deviation by any firm, even the most efficient one, is punished maximally. The punishments we construct to retaliate against deviations by the low-cost firm have a stick-and-carrot structure.[10] For some time, the efficient firm serves the entire demand at a price below its own marginal cost, thus making losses. Then, the carrot phase of the punishment starts, and firms switch to some (sustainable) cooperative outcome that generates positive profits. The low-cost firm is willing to stick to its own punishment although it makes temporary losses, since compliance yields a reward in the future. To punish deviations by any other firm, we simply rely on reversion to the competitive equilibrium.

Lemma 1 *Suppose that (stationary)*[11] *collusion on some price $p^* > c_2$ and the market sharing rule s^* can be supported by minmax punishments. Select any $p^P < c_1$ and T such that*

$$\sum_{t=1}^{T} \delta^{t-1}\pi_1(p^P) + \frac{\delta^T}{1-\delta} s_1^* \pi_1(p^*) = 0, \tag{1}$$

and let $\varepsilon \in [0, c_1 - p^P]$.

Then the following punishment strategies minmax deviators and are (jointly) credible:

- *Upon any deviation by firm 1, firms play the following sequence of prices and market sharing rules starting from the first period after the deviation $t = 1$:*

$$\left[\left\{ \left(p^P, p^P + \varepsilon, ..., p^P + \varepsilon\right), (1, 0, ..., 0) \right\}_{t=1}^{T}, \left\{ (p^*, ..., p^*), s^* \right\}_{t=T+1}^{\infty} \right].$$

[9] Note that for c_2 to be the equilibrium price, it is not necessary that all firms charge c_2. One profile of (undominated) strategies that supports the competitive equilibrium for small enough η is the following: the low-cost firm charges c_2, and any firm $i \neq 1$ randomizes uniformly over $[c_i, c_i + \eta]$. See Blume (2003) for a proof of this in the case of two firms.

[10] Abreu (1986) was the first to propose punishments with such a strucure.

[11] We show in the appendix that whenever (i.e. for any discount factor for which) non-stationary collusion can be supported by minmax punishments, there also exists some stationary outcome that can be supported by minmax punishments.

- *Upon any deviation by a firm $i \neq 1$, firms revert to the competitive equilibrium for ever from the first period after the deviation onwards.*

Proof. The proposed punishments minmax deviators by construction. First, by (1) firm 1's punishment leaves zero continuation profits to 1. Second, any firm $i \neq 1$ earns zero profits in the competitive equilibrium where it does not make any sales.

It remains to establish that the proposed strategy profile is indeed credible. For this we need to show that no firm has an incentive to deviate from any punishment at any stage and be punished in turn. As usual, it suffices to consider one-shot deviations.

It is straightforward that no firm has an incentive to deviate from a punishment that prescribes the infinite repetition of the competitive equilibrium: no firm can make a short-term gain by deviating from a static equilibrium, and a deviation starts a minmax punishment for the deviator.

To show that no firm has a strict incentive to deviate from 1's punishment, it is sufficient to consider deviations at $t = 1$. Clearly, no firm has an incentive to deviate for $t > T$ if the proposed punishment strategies are indeed credible, since by assumption the collusive path can be supported by minmax punishments. Also, a firm has no incentive to deviate in any period $t \in [2, T]$ if it has no incentive to deviate at $t = 1$: The short-term gains from a deviation are the same at any stage $t \in [1, T]$, whereas the cost of foregoing the future switch to collusion increases with t.

Firm 1's best possible deviation from its own punishment at $t = 1$ is to charge a price above $p^P + \varepsilon$ to earn zero instead of negative profits in the first period. This deviation would trigger the restart of firm 1's punishment with zero continuation profits. The firm is hence indifferent between complying and deviating optimally.

A firm $i \neq 1$ cannot benefit by deviating from 1's punishment at $t = 1$ either: a deviation could not generate any short-term benefit but would nonetheless trigger a minmax punishment for firm i. Hence, no firm $i \neq 1$ wants to deviate from 1's punishment:

$$0 + \delta \times 0 = 0 \leq 0 + \frac{\delta^T}{1-\delta} s_i \pi_i(p^*).^{12}$$

■

Note that it would also be possible to minmax all firms by means of "non-standard" grim trigger strategies. The low-cost firm's punishment then consists of reversion to

[12] The assumption that the collusive path is sustainable by minmax punishments trivially implies that all firms earn non-negative profits in every stage of the collusive path.

the non-conventional static equilibrium in which all firms charge c_1, and the low-cost firm serves the entire demand.[13] This punishment seem unreasonable, however, since it involves the play of weakly dominated strategies: for a high-cost firm charging a price below its own cost forever is never better than charging a price at or above its own cost forever. Any arbitrarily small risk of mistakes by the other firms would hence suffice to make such strategies unappealing.

The punishment strategies we construct do not involve any weakly dominated strategies; since the efficient firm's punishment has a stick-and-carrot structure, below-cost pricing is only temporary for all firms. Compliance with the punishments is therefore the unique best reply to a strategy by the other firms that consists of undercutting the stick price before switching to collusion conditional on compliance.[14] From (1), it is clear that reversion to the non-conventional equilibrium with price c_1 is the limit of the efficient firm's stick-and-carrot punishment as $T \to \infty$. Non-standard grim trigger strategies, with reversion either to the non-conventional static equilibrium or the static competitive equilibrium depending on the identity of the deviator, are hence the limit of a profile of undominated strategies.

Finally, let us make some remarks concerning the "plausibility" of stick-and-carrot punishments. The structure of such punishments is very natural: having detected a deviation, firms first fight each other vigorously for some time before returning collusion. Compared to grim trigger punishments, stick-and-carrot punishments also have the advantage of not being prone to renegotiation anymore once the stick phase is over. They are also truly targeted at the deviator, while minmax grim trigger strategies punish all firms, not only the deviator, down to zero.

There may be limits on the severity of punishments in the real world, for example due to budget constraints or a lack of commitment not to renegotiate, especially when executives change over time. On the other hand, cartel members credibly threatened

[13]To sustain an equilibrium with price c_1, at least one of the less efficient firms has to charge this price, otherwise the low-cost firm would want to increase its price.

[14]If one or several firms are so inefficient that their marginal costs lie *above* the collusive price p^*, then these firms' prices along the low-cost firm's punishments must be adapted to rule out weakly dominated strategies. More precisely, such firms must charge prices above their own marginal cost instead of p^* during the carrot phase of 1's punishment (alternatively, they could also charge a higher price in all periods). This does not have any impact on the sustainability of collusion at price p^*, since in any case firms with higher marginal costs will not receive any positive market share in schemes that minmize deviation incentives.

other firms even with arson or physical violence in some industries. Porter (2004) reports evidence of such threats in a cartel between Italian bakers in Greenwich village in the 1980's, as well as in the New York trash haulers cartel. Trash haulers in Los Angeles refrained from illegal threats but agreed to punish competitors by offering below-cost rates, thus imposing temporary negative profits. It is important to bear in mind, however, that the discussion about the severity of punishments is a priori independent of cost asymmetry. The crucial ingredient needed for our result that collusion is sustainable under cost symmetry whenever sustainable under cost asymmetry is that each firm's continuation profits along its own punishment is equal to the same proportion of that firm's discounted cartel profits, not that punishments are maximal.

4 Collusion without Side Payments

We now turn to the question of which agreements firms can sustain by optimal punishments. We first derive the set of all sustainable stationary collusive outcomes. Next, we restrict attention to allocations that are efficient in the sense of being Pareto-optimal for the firms. In this context, it is important to distinguish between collusive agreements such that both firms share the market in each period, and lotteries that grant temporary monopoly positions, since firms are able to achieve higher expected profits with lotteries. We then check when a given efficient agreement, be it deterministic or a lottery, is indeed sustainable. Finally, we analyze what we call the constrained Pareto frontier, i.e. the subset of undominated allocations (for the firms) within the set of sustainable (stationary) outcomes.

4.1 Sustainability

We define a collusive outcome by a vector (p, s), where $p > c_2$ is the firms' collusive price and s the market sharing rule. Our focus is hence on collusive paths in which all active firms sell at a common price in every period, and this price as well as market shares are constant over time. In the appendix, we show that these restrictions do not limit the scope for sustainability in the sense that whenever the discount factor is high enough to sustain some non-stationary collusive path, then there also exists a stationary collusive path.

A collusive path is sustainable if and only if it can be supported by the most severe subgame perfect punishment strategies. In the following, we will characterize the set of stationary paths sustainable when the punishment triggered by any deviation is maximal for the deviator. Lemma 1 then implies that if this set is non-empty, firms can indeed "collude" on maximal punishments for all firms.

Note first that no collusive scheme ever assigns a positive market share to a firm whose cost is above the collusive price; otherwise, such a firm would make negative profits by sticking to collusion, whereas it could earn zero continuation profits by deviating to a higher price and be punished in turn. Firms with too high marginal costs will therefore not participate in the collusive agreement, but rather play the role of potential entrants in the industry.[15] We define the set of "active firms" by

$$A(p) = \{i \mid c_i < p\}.$$

Sustainability of collusion then boils down to the requirement that none of the active firms has an incentive to deviate from the collusive outcome.[16]

The optimal short-term deviation for an active firm $i \in A(p)$ is to charge p_i^m if the collusive price lies above p_i^m, and to slightly undercut its rivals' price otherwise. The non-deviation constraint of any active firm i is hence

$$\frac{1}{1-\delta} s_i \pi_i(p) \geq \pi_i\left(\min[p, p_i^m]\right) \text{ for } i \in A(p) \tag{C_i}$$

A collusive outcome (p, s) is sustainable by maximal punishments if and only if it satisfies conditions C_i for all $i \in A(p)$.

Adding up the non-deviation conditions C_i of all active firms using the fact that their market shares must add up to one yields the following necessary condition for collusion at price p:

$$\delta \geq \widetilde{\delta}(p), \tag{2}$$

where

$$\widetilde{\delta}(p) \equiv \frac{\sum_{i \in A(p)} \frac{\pi_i\left(\min[p, p_i^m]\right)}{\pi_i(p)} - 1}{\sum_{i \in A(p)} \frac{\pi_i\left(\min[p, p_i^m]\right)}{\pi_i(p)}}.$$

[15] We implicitly assume that each of the firms $1, ..., n$ can freely enter the market at any time. Entry of outsiders, on the other hand, is blockaded. The source of the advantage of inactive incumbents over outsiders could for example be a patent or a licence that cannot be traded freely.

[16] If $c_i = p$, firm i's non-deviation constraint is satisfied trivially, since both collusive and deviation profits are zero in this case. Granting a positive market share to firm i would then hinder collusion, since some other firm's market share would need to be reduced. We therefore take it as given that $\sum_{i \in A(p)} s_i = 1$. This assumption will simplify the exposition, but does not influence the critical discount factor.

It is easy to see that this condition is not only necessary but also sufficient: whenever (2) is satisfied, there exists a vector of market shares s such that the non-deviation conditions C_i are satisfied for all active firms.

We denote the number of active firms, that is of elements in $A(p)$, by $m(p) \in [2, n]$. Clearly, $m(p)$ is (weakly) increasing: as the collusive price rises, more firms could profitably undercut p and must therefore join the collusive agreement for it to remain sustained.

For collusive prices $p \in (c_2, p_1^m]$, the critical discount factor is $\widetilde{\delta}(p) = \frac{m(p)-1}{m(p)}$, which is also the threshold for collusion between $m(p)$ symmetric firms. This result arises because in this case each active firm's deviation would consist in slightly undercutting its rival, and all firms' punishments impose zero continuation profits. Each cartel member's deviation incentives then only depend on its market share relative to the discount factor, and the non-deviation constraints are symmetric: for all $i \in A(p)$,

$$s_i \geq 1 - \delta. \tag{C_i'}$$

Note that even for prices below the most efficient firm's monopoly price, the critical discount factor may exhibit upward jumps if the number of active firms $m(p)$ increases, so that the market must be shared by a larger number of firms to preserve collusion. Suppose for example that $n = 3$, and $c_2 < c_3 < p_1^m$. Then the critical discount factor is $\frac{1}{2}$ for $p \in (c_2, c_3]$, but $\frac{2}{3}$ for $p \in (c_3, p_1^m]$.

For $p > p_1^m$, on the other hand, the discount factor threshold $\widetilde{\delta}(p)$ strictly increases even if the number of active firms remains constant. This result is driven by the wedge between a firm's stand-alone collusive profits $\pi_i(p)$ and its deviation profits $\pi_i(p_i^m)$ whenever $p > p_i^m$. Given any market sharing rule, firm i's incentive to deviate is then clearly higher the larger the difference between the collusive price and its own monopoly price. For $p > p_1^m$, this difference is positive for at least the most efficient firm 1, which drives up the critical discount factor. If the collusive price exceeds the monopoly prices of several firms, this effect is further reinforced.

The critical discount factor is thus increasing in p for two reasons: (i) by creating or increasing the wedge between stand-alone collusive profits and short-term deviation profits, higher prices may increase the deviation incentives of already active firms, and (ii) a price increase may attract "entry", which in turn forces firms to share the market with more firms in order to preserve collusion.

The minimum market share that must be granted to an active firm $i \in A(p)$ such that collusion at price p is indeed sustainable for some discount factor $\delta \geq \widetilde{\delta}(p)$ is

$$\widetilde{s}_i(p,\delta) \equiv (1-\delta)\frac{\pi_i\left(\min[p, p_i^m]\right)}{\pi_i(p)}.$$

This lower bound $\widetilde{s}_i(p,\delta)$ is such that C_i is binding. Moreover, each firm's market share is restricted upwards by the other firms' non-deviation constraints. In particular, the maximum market share that can possibly be granted to firm i without triggering a deviation by any other firm is $1 - \sum_{j \in A(p) \neq i} \widetilde{s}_j(p,\delta)$.

For $p \in (c_2, p_1^m]$, when the non-deviation constraints are independent of the collusive price, market shares are restricted by $s_{i \in A(p)} \in [1-\delta, 1-(m(p)-1)(1-\delta)]$ and $\sum_{i \in A(p)} s_i = 1$, as under cost symmetry between $m(p)$ firms. For prices above p_i^m, the lower bound on firm i's market share, $\widetilde{s}_i(p,\delta)$, strictly increases with the price to accommodate i's increasing deviation incentives.

Figure 1 provides a graphical representation of the sustainable outcomes for different discount factors when there are only two firms. Note that the couple (p, s_1) completely defines an allocation in this case. As under cost symmetry, only equal market sharing rules are sustainable for $\delta = \frac{1}{2}$: the set of sustainable outcomes is $\Delta(\frac{1}{2}) = \left\{(p, s_1) \mid p \in (c_2, p_1^m], s_1 = \frac{1}{2}\right\}$. For $\delta' \in (\frac{1}{2}, \widetilde{\delta}(p_2^m)]$, the set of sustainable allocations $\Delta(\delta')$ includes all outcomes left of or on the line labelled $C_2(\delta')$, along which the high-cost firm is indifferent between complying and deviating, as well as right of or on the line labelled $C_1(\delta')$, along which the low-cost firm is indifferent between deviating and complying. Both non-deviation constraints are binding at the maximum collusive price: $p' = \widetilde{\delta}^{-1}(\delta')$. Clearly, as the discount factor increases, the set of sustainable allocations becomes larger and larger. First, the maximum sustainable price, $\widetilde{\delta}^{-1}(\delta)$, increases with the discount factor. Second, for any given sustainable price, the range of sustainable market sharing rules expands in both directions as the discount factor rises.

Drastic Cost Difference So far, we have assumed that the cost difference was sufficiently small so that $c_2 < p_1^m$. If instead the low-cost firm enjoys a drastic cost advantage, it charges its monopoly price p_1^m and earns monopoly profits in the static competitive equilibrium. No collusion is then sustainable by grim trigger strategies where firms revert to the competitive equilibrium.

Once we consider optimal punishments, however, collusion among several firms is still sustainable. First note that firms can credibly collude on maximal punishments irrespective of the size of the cost difference. The characterization of the set of sustainable

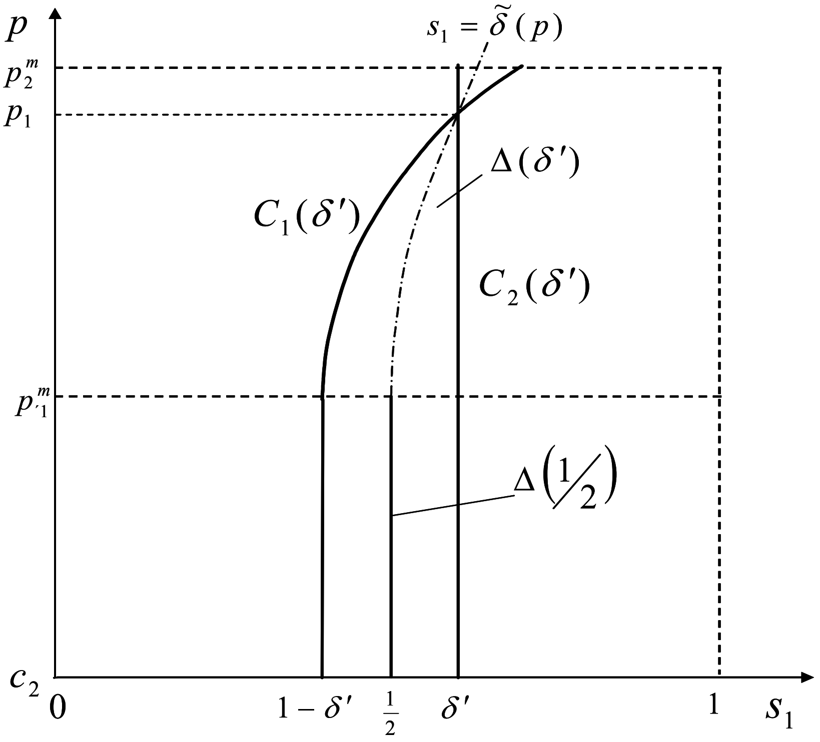

Figure 1: Sustainable Collusive Allocations without Side Payments

collusive allocations then remains unchanged, except that any collusive price above c_2 (so that at least two firms are active) lies stricly above p_1^m now. The low-cost firm's optimal deviation is therefore always to charge its monopoly price, and collusion (even between only two firms) is only possible for discount factors strictly above $\frac{\pi_1(p_1^m)}{\pi_1(c_2)+\pi_1(p_1^m)} > \frac{1}{2}$. Collusion between several firms is thus sustainable even when the cost difference is drastic; the critical discount factor, however, is higher. Note also that when the cost difference is drastic, in any collusive scheme the efficient firm is locked into a "bad" equilibrium where its profits are lower than in the competitive equilibrium.

4.2 Efficient Collusion

In this section, we consider the sustainability of allocations that are Pareto-optimal for the firms. We first restrict attention to statically efficient allocations, and show that it is more difficult to sustain such allocations under cost asymmetry than under cost symmetry. Then we drop the restriction of stationarity, which is a serious one here because firms can improve Pareto efficiency by taking turns being the monopolist. We therefore also analyze the sustainability of collusion on a lottery for the monopoly position, and show that efficient collusion remains more difficult than under cost symmetry. For simplicity, we restrict attention to an industry with only two firms, i.e. $n = 2$, in this section.

4.2.1 Pareto-Efficient Production

Let us first analyze the efficient allocation of production between two cost asymmetric firms in the absence of side payments, neglecting the issue of collusive sustainability. We consider the following problem:

$$\max_{\{p,s_1\}} [s_1\pi_1(p)]^\alpha \left[(1-s_1)\pi_2(p)\right]^{1-\alpha}, \ \alpha \in [0,1]. \tag{P1}$$

Solving $(P1)$ for every $\alpha \in [0,1]$ yields a simple characterization of all Pareto-efficient outcomes for the firms.[17] The solution for each α is characterized by:

$$s_1 = \alpha, \tag{3}$$

and

$$-\alpha\frac{\pi_1'(p)}{\pi_1(p)} = (1-\alpha)\frac{\pi_2'(p)}{\pi_2(p)}. \tag{4}$$

[17] See exercise 6.1 in Tirole (1988) for a detailed treatment of an equivalent problem.

As α varies between 0 and 1, the optimal market sharing rule varies between 0 and 1, and the optimal price between the two firms' monopoly prices. The two firms' isoprofit lines are tangent at any Pareto optimum.

Combining (3) and (4) yields the following one-to-one relationship between the low-cost firm's market share and the price:

$$s^O(p) = \frac{(c_2 - c_1)D(p) + (p - c_2)\pi_1'(p)}{(c_2 - c_1)D(p)}, p \in [p_1^m, p_2^m]. \tag{5}$$

As can be checked easily, $s^O(\cdot)$ is strictly downward-sloping. The inverse of $s^O(\cdot)$ will be denoted by $p^O(\cdot)$.

4.2.2 Stationary Collusion on Pareto-Efficient Outcomes

Let us now analyze the sustainability of Pareto-efficient outcomes as characterized by condition (5).

It is easy to check that the minimum discount factor for which *some* Pareto-efficient outcome is sustainable must be such that both firms are indifferent between colluding and deviating. To see this, suppose first that some allocation $(p^O(s_1), s_1)$ is sustainable at discount factor δ, but (at least) one firm strictly prefers compliance. Then, by continuity, there exists another Pareto-efficient allocation sustainable at lower discount factors. First, firms can move along the Pareto frontier, into the direction preferred by the firm with the binding non-deviation constraint, to an allocation at which both firms strictly prefer compliance. Second, if both firms strictly prefer compliance at δ, they will also be willing to collude at a slightly lower discount factor. Alternatively, suppose that both firms are indeed indifferent between deviating and complying from $(p^O(s_1), s_1)$ for discount factor δ. By definition, any Pareto-efficient allocation different from $(p^O(s_1), s_1)$ is strictly better for one of the firms but strictly worse for the other firm. Thus, for δ, one of the firms would deviate from any Pareto-optimal allocation other than $(p^O(s_1), s_1)$. We can conclude that no other Pareto-optimal allocation is sustainable for the same (or a lower) discount factor.

Figure 2 pictures the minimum discount factor $\widehat{\delta}$ for which some efficient collusion is sustainable. The corresponding allocation is denoted by $(\widehat{p}, \widehat{s})$. As just explained, both firms must be indifferent between colluding and deviating. Since both non-deviation constraints are binding at price $\widehat{p}$ only if $\delta = \widetilde{\delta}(\widehat{p})$, and for this discount factor only the market sharing rule $\widetilde{\delta}(\widehat{p})$ is sustainable, it must be that $\widehat{s} = \widetilde{\delta}(\widehat{p})$. Moreover, for $(\widehat{p}, \widehat{s})$ to be Pareto-efficient, it must lie on $s^O(p)$. Since $\widetilde{\delta}(p)$ strictly increases in p, while $s^O(p)$ strictly decreases in p, there indeed exists a unique allocation $(\widehat{p}, \widehat{s})$ which is both Pareto-efficient

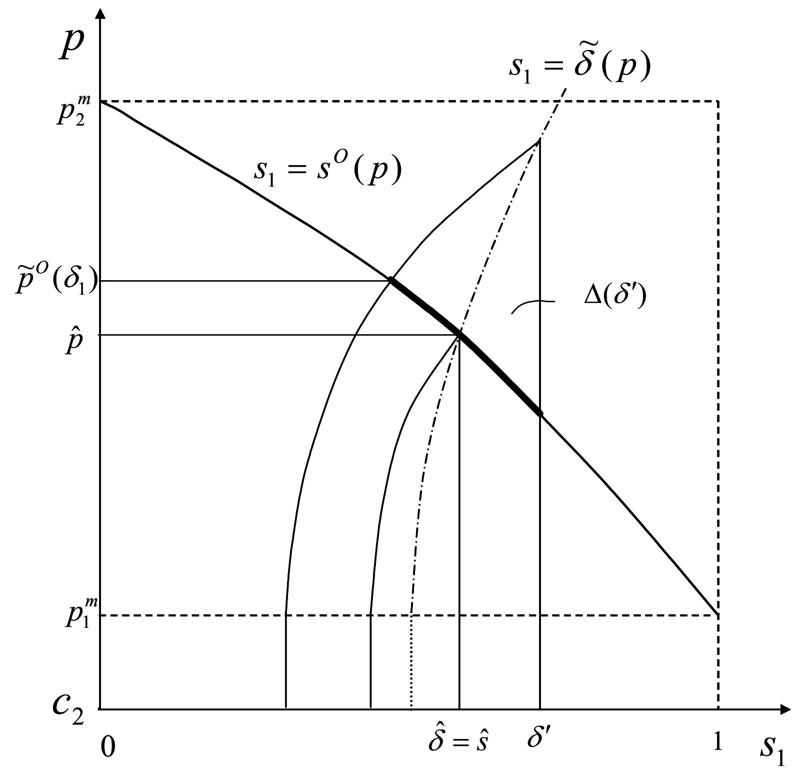

Figure 2: Efficient Stationary Collusion

and such that both firms are indifferent between colluding and deviating. Graphically, the allocation $(\widehat{p}, \widehat{s})$ can thus be found where $s^O(p)$ cuts $\widetilde{\delta}(p)$, and the corresponding discount factor threshold is $\widehat{\delta} = \widehat{s}$. For discount factors above $\widehat{\delta}$, a whole range of efficient allocations is sustainable. This is illustrated in the figure for $\delta' > \widehat{\delta}$.

The following proposition summarizes these results (a formal proof is relegated to the appendix):

Proposition 2 *Let $\widehat{p} \in (p_1^m, p_2^m)$ be uniquely defined by $s^O(\widehat{p}) = \widetilde{\delta}(\widehat{p})$, and let $\widehat{\delta} \equiv \widetilde{\delta}(\widehat{p})$. Moreover, let $\widetilde{p}^O(\delta)$ be the function implicitly defined by $s^O(\widetilde{p}^O(\delta)) = \widetilde{s}_1(\widetilde{p}^O(\delta), \delta)$, and note that $\widehat{p} = \widetilde{p}^O(\widehat{\delta})$. Then,*

(i) *for $\delta < \widehat{\delta}$, no Pareto-efficient allocation is sustainable.*

(ii) *for $\delta \geq \widehat{\delta}$, all Pareto-efficient allocations $\left(p^O(s_1), s_1\right)$ with market sharing rules s_1 in the range $\left[s^O(\widetilde{p}^O(\delta)), \delta\right]$ are sustainable. In particular, for $\delta = \widehat{\delta}$, the unique sustainable Pareto-efficient allocation is $\left(\widehat{p}, \widehat{\delta}\right)$.*

Under cost symmetry, the discount factor for some efficient collusion is $\frac{1}{2}$: collusion on the common monopoly price is possible for any discount factor above this threshold if firms split the market equally. Since $\widehat{\delta} > \frac{1}{2}$, cost asymmetry thus hinders the sustainability of efficient collusion.

4.2.3 Fully Efficient Collusion on a Lottery

So far, we have analyzed stationary collusion on statically efficient allocations. It is however straightforward to see that the firms could achieve a Pareto improvement by taking turns being the monopolist. Consider any statically efficient stationary allocation $\left(p^O(s_1), s_1\right)$: the firms' per-period profits are $s_1\pi_1\left(p^O(s_1)\right)$ and $(1-s_1)\pi_2\left(p^O(s_1)\right)$. Alternatively, let the low-cost firm be a monopolist with probability s, and the high-cost firm with probability $(1-s_1)$. Now, the low-cost firm's expected per-period profits are $s_1\pi_1\left(p_1^m\right)$, and the high-cost firm's expected per-period profits are $(1-s_1)\pi_2\left(p_2^m\right)$. Thus, both firms are better-off in terms of expected profits. In other words, the Pareto profit frontier of problem $(P1)$ is convex, and fully efficient collusion must allow for alternating or random monopolies.

To achieve fully efficient collusion, firms must be able to use correlated strategies. We will suppose that, by throwing a dice or observing some common signal at the beginning of each period, the firms can assign the monopoly to the low-cost firm with probability ρ, and to the high-cost firm with probability $(1-\rho)$. The probability ρ is part of the collusive arrangement, much as the market share earlier on. If the low-cost firm is the monopolist, it charges its monopoly price p_1^m and serves the whole market, while the high-cost firm charges (slightly more than) p_1^m and makes no sales at all. If the high-cost firm is the monopolist, it charges p_2^m and serves the whole market, while the high-cost firm charges (slightly more than) p_2^m and makes no sales at all. This way, the firms can achieve any point on the unconstrained linear Pareto profit frontier by letting ρ vary between zero and one. There is no loss of generality in restricting attention to lotteries with a stationary ρ.[18]

[18] To see this, suppose firms collude on a path of probabilities $\{\rho_t\}_{t=0}^{\infty}$ from today $(t=0)$ to infinity. The probabilities ρ_t can also be interpreted as expected probabilities when firms collude only on the probability distribution(s) from which the ρ_ts are drawn. Define

$$\rho \equiv \frac{(1-\delta)}{\delta}\sum_{T=\widehat{t}+1}^{\infty}\delta_T^{T-\widehat{t}}\rho_T$$

as the constant probability which would yield the same continuation payoff from time $\widehat{t}$ onwards as the sequence $\{\rho_t\}_{t=\widehat{t}}^{\infty}$. Then, the non-deviation constraints of the non-stationary lottery at time $\widehat{t}$ coincide with the non-deviation constraints of a stationary lottery with probability ρ. Since the non-deviation

It is easy to generalize the optimal stick-and-carrot punishments to allow for lotteries when firms cooperate, i.e. along the carrot phase of punishments. The stick price as well as the length of the stick period simply need to be such that the low-cost firm's *expected* continuation profits are zero. Consequently, the optimal punishments are maximal again.

Obviously, each firm's incentives to deviate are strongest when its rival holds the monopoly position. The low-cost firm's relevant non-deviation constraint is then

$$\pi_1(p_1^m) \leq 0 + \frac{\delta}{1-\delta}\rho\pi_1(p_1^m).$$

Similarly, the high-cost firm's relevant non-deviation constraint is

$$\pi_2(p_1^m) \leq 0 + \frac{\delta}{1-\delta}(1-\rho)\pi_2(p_2^m).$$

These conditions imply that fully efficient collusion can be sustained if and only if

$$\delta \geq \frac{\pi_2(p_2^m) + \pi_2(p_1^m)}{2\pi_2(p_2^m) + \pi_2(p_1^m)} > \frac{1}{2}. \tag{6}$$

Hence, fully efficient collusion requires discount factors above the threshold for efficient collusion under cost symmetry when no randomisation is needed to achieve full efficiency.

4.3 The Pareto Frontier of Sustainable Allocations

We now analyze the subset of Pareto undominated allocations within the set of sustainable outcomes for each discount factor for $n = 2$. Unlike in the previous section, we do not ask when a given efficient allocation is sustainable, but rather which of the sustainable allocations for a given discount factor are undominated. This approach takes account of the methodological point, underlined by Harrington (1991), that an allocation only provides a sensible collusive outcome if it is indeed implementable by a self-enforcing agreement. By restricting attention to the set of sustainable collusive equilibria a priori, the firms automatically solve this implementation problem.

Proposition 3 *The set of Pareto-undominated sustainable allocations is*

$$\Omega(\delta) = \Delta(\delta) \cap \left[\left\{(p, s_1) \mid \left(p \in \left[p_1^m, p^O(\delta)\right), s_1 = \delta\right)\right\} \cup \left\{(p, s_1) \mid p = p^O(s_1), s_1 \in [0,1]\right\}\right].$$

constraints must be satisfied in every time period, this implies that there is no loss of generality from considering only stationary lotteries.

Proof. First of all, unconstrained Pareto optimal allocations are obviously undominated if sustainable. For $\delta \geq \widehat{\delta}$, the set of Pareto undominated sustainable allocations therefore always includes part of the Pareto frontier $p^O(s_1)$.

Note also that any allocation (p, s_1) in $\Delta(\delta)$ with $p < p_1^m$ is Pareto dominated by the allocation (p_1^m, s_1), which is also included in $\Delta(\delta)$.

Therefore, we can restrict attention to sustainable allocations with prices at least equal to the low-cost firm's monopoly price. For such prices, the low-cost firm's isoprofit lines are strictly increasing and concave in the (s_1, p) space. In fact, the isoprofit line for profit level $\Pi = (1-\delta)\pi_1(p_1^m)$ coincides with $C_1(\delta)$. Profit levels are increasing in the southeast direction, as the low-cost firm prefers a higher market share s_1 and prices closer to its own monopoly price.

The high-cost firm's isoprofit lines are increasing and convex in the (s_1, p) space. For allocations below $p^O(s_1)$ they are flatter than, for allocations on $p^O(s_1)$ tangent to, and for allocation above $p^O(s_1)$ steeper than the isoprofit lines of the low-cost firm. Moreover, profit levels are increasing in the northwest direction, as the high-cost firm prefers a higher market share $(1 - s_1)$ and higher prices.

Given this, it is straightforward that any sustainable allocation strictly above $p^O(s_1)$ is Pareto-dominated: moving along the low-cost firm's isoprofit curve towards $p^O(s_1)$ always increases the high-cost firm's profits without hindering collusive sustainability. Now consider any allocation strictly below $p^O(s_1)$. If firms are able to move northeast along the low-cost firm's isoprofit line without violating sustainability, a Pareto improvement within $\Delta(\delta)$ can be achieved: the high-cost firm is strictly better off thanks to the price increase although its market share $(1 - s_1)$ is lower. The only sustainable allocations strictly below $p^O(s_1)$ that are not dominated are then those for which the high-cost firm's non-deviation constraint is binding, i.e. $s_1 = \delta$, so that no further northeast moves are feasible. ■

Undominated sustainable allocations thus either lie on the high-cost firm's non-deviation constraint or/and are unconstrained Pareto optima. In the former case, prices lie between the low-cost firm's monopoly price and $p^O(\delta)$. Figure 3 illustrates the sets of Pareto-undominated sustainable allocations for two different discount factors, δ_1 and δ_2, one below and one above $\widehat{\delta}$.

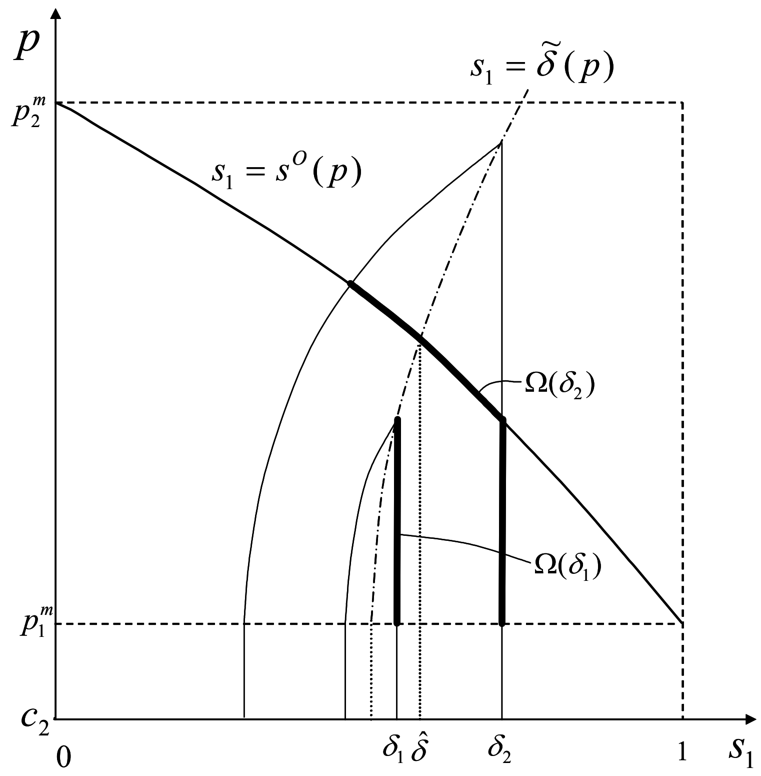

Figure 3: The Pareto Frontier of Sustainable Allocations

5 Collusion with Side Payments

Side payments are often ruled out in the literature on collusion,[19] since antitrust law forbids overt monetary transfers between firms in most jurisdictions. Nonetheless, as shown by the examples in the introduction, side payments are sometimes part of cartel agreements.

If antitrust institutions indeed monitor direct payments well, more indirect transfers are conceivable. Competitors that are also business partners, for example, can quite easily manipulate transfer prices. Joint ventures may also serve as a vehicle for such transfers. Moreover, many industries exhibit a high degree of (both active and passive) cross-ownership between firms. Firms are thus often direct claimants to a share of the profits of one of their rivals.[20] Finally, payments in kind may sometimes be another way to avoid antitrust suspicions.

[19] One exception is Jehiel (1992).

[20] The effect of cross-ownership on collusion has recently been analyzed by Gilo, Spiegel & Moshe (2004) who find that cross-ownership, even if passive, usually facilitates collusion. Active ownership reinforces this effect.

In the following analysis, there are no restrictions at all on side payments. Thus, firms can collude on outcomes in which only the efficient firm produces, but transfers part of its profits to inefficient firms. This is clearly an extreme case that does not reflect reality, yet it allows us to identify the mechanisms by which cost asymmetry affects collusive sustainability when side payments are feasible. The main qualitative insight that side payments decrease the relative deviation incentives of inefficient firms and therefore facilitate some collusion carries over if the extent of side payments is limited.

5.1 Sustainable Outcomes and Efficiency

We now consider the following stage game. At the beginning of each period, the firms simultaneously quote prices. Then, all prices are observed and the lowest price firm(s) serve(s) the entire demand. Finally, firms can make side payments to share the profits earned in that period.

We can restrict attention to collusive allocations in which the most efficient firm carries out all the production in every period: letting any other firm produce a positive quantity in some or all periods would lower collusive profits, but not deviation profits. A collusive outcome is then defined by a vector (p, S), where $p > c_2$ is the collusive price, and $S = (S_1, S_2, ..., S_n) \in [0,1]^n$ such that $\sum_i S_i = 1$ the profit sharing rule. When complying with the collusive agreement, all firms quote price p, the low-cost firm then serves the entire demand, and finally makes a side payment equal to $S_i \pi_1(p)$ to each firm $i \neq 1$.

Note that the low-cost firm has no reason to make positive side payments to firms with marginal costs above the adopted price p, since those firms cannot credibly threaten to undercut the collusive price. Hence, only sufficiently efficient firms need to receive positive transfers to prevent deviations; the set of such firms coincides with the set of "active firms" $A(p) = \{i \mid c_i < p\}$ in the analysis without side payments. The collusive profit sharing rule is then such that $S_i > 0$ for all $i \in A(p)$ and $\sum_{i \in A(p)} S_i = 1$.

We use the same minmax punishment strategies as in the previous analysis. Since firms cannot be punished more severely, there is no point in introducing side payments during punishment phases.

The low-cost firm could optimally deviate from the collusive outcome by charging $\min[p, p_1^m]$ and then refusing to make any side payments. The low-cost firm's no-deviation constraint is thus

$$\pi_1 \left(\min[p, p_1^m]\right) \leq \frac{1}{1-\delta} S_1 \pi_1(p). \qquad (D_1)$$

The best deviation of any other active firm $i \in A(p) \setminus 1$ would be to either slightly undercut p if the collusive price lies below its own monopoly price, or to charge p_i^m otherwise. Such a deviation would not only trigger the start of i's punishment in the next period, but also make i lose the side payment from 1 in the deviation period: knowing a punishment will start in the next period, firm 1 no longer has any incentives to make side payments. The non-deviation constraint of any firm $i \in A(p) \setminus 1$ is hence

$$\pi_i \left(\min[p, p_i^m]\right) \leq \frac{1}{1-\delta} S_i \pi_1(p). \tag{D_i}$$

A collusive outcome (p, S) is then sustainable if and only if conditions D_i are satisfied for all firms $i \in A(p)$. The implied necessary and sufficient condition for collusion at price p is:

$$\delta \geq \widetilde{\delta}^T(p),$$

where

$$\widetilde{\delta}^T(p) \equiv \frac{\sum_{i \in A(p)} \pi_i(\min[p, p_i^m]) - \pi_1(p)}{\sum_{i \in A(p)} \pi_i(\min[p, p_i^m])}. \tag{7}$$

Any collusive outcome such that $p = p_1^m$ is efficient, since firms cannot jointly gain by either changing the price or reallocating production. The critical discount factor for efficient collusion is thus

$$\widetilde{\delta}^T(p_1^m) = 1 - \frac{(p_1^m - c_1)}{\sum_{i \in A(p_1^m)} (p_1^m - c_i)}.$$

Since $c_i > c_1$ for all $i \neq 1$, and $A(p_1^m)$ includes at least two firms,

$$\widetilde{\delta}^T(p_1^m) < 1 - \frac{1}{m(p_1^m)}.$$

The critical discount factor is thus lower both than $\widetilde{\delta}(p_1^m)$, the threshold for collusion without side payments, and than the critical discount factor for collusion between $m(p_1^m)$ symmetric firms.[21] The reason behind this is that the short-term deviation profits, $\pi_i(p_1^m)$, of the firms that receive side payments firm are strictly below firm 1's stand-alone profits, $\pi_1(p_1^m)$. The deviation incentives of all cartel members except the most efficient firm are hence lower than without side payments when firms agree on a sharing rule $s = S$. Moreover, a high-cost firm gains relatively less from deviating than any member of a symmetric cartel: if side payments are feasible, collusion with a more efficient competitor is more "valuable" than collusion with an identical firm.

[21]The feasibility of side payments is irrelevant under cost symmetry. In the absence of fixed costs, no advantage can be derived from allocating production to only one of the firms.

The results of the following proposition build on our main point that side payments decrease the relative deviation incentives of inefficient firms. In the proof (which is in the appendix), we compare the critical discount factor required to sustain collusion in different situations to derive the two main conclusions of this section: (i) side payments facilitate collusion between cost asymmetric firms, and (ii) when side payments are feasible, cost asymmetry facilitates collusion.

Proposition 4 *If side payments are feasible, then*

1. *collusion on any price $p > c_2$ is easier to sustain than when side payments are infeasible.*

2. *collusion on any price $p > c_2$ is easier to sustain between an efficient firm with cost c_1 and $(m-1)$ less efficient firms with marginal costs between c_2 and p than between m efficient firms each with marginal cost c_1.*

3. *it is easier to sustain collusion on some price $p \in (c_2, p_1^m]$ between an efficient firm with cost c_1 and $(m-1)$ less efficient firms with marginal costs between c_2 and p than it is to sustain profit-maximizing collusion between m symmetric firms (with any level of marginal costs).*

Finally it is worth noting that the threshold $\widetilde{\delta}^T(p)$ is strictly increasing everywhere. For $p < p_1^m$, any price reduction alleviates the inefficient firms' non-deviation constraints without affecting the low-cost firm's deviation incentives. In fact, $\widetilde{\delta}^T(p) \to 0$ as $p \to c_2$, so that some collusion is sustainable for any $\delta > 0$.[22]

For $p > p_1^m$, a price rise clearly increases the low-cost firm's deviation incentives by driving a wedge between deviation profits, $\pi_1(p_1^m)$, and stand-alone collusive profits, $\pi_1(p)$. Moreover, a price rise also increases the deviation incentives of all other firms, since $\frac{\pi_i(\min[p, p_i^m])}{\pi_1(p)}$ is increasing in p for all $i \neq 1$. Figure 4 illustrates the sets of sustainable outcomes with side payments when there are only two firms. The non-deviation constraints are illustrated for two different discount factors $\widetilde{\delta}^T(p_1^m)$ and $\delta' > \widetilde{\delta}^T(p_1^m)$.

[22]If less efficient firms that have previously stopped production can only reenter with some time lag, the discount factor threshold does not tend towards zero. However, collusion is still sustainable for discount factors below $\frac{1}{2}$ if the time lag is not too long. In particular, if the time lag is only one period, the discount factor threshold for efficient collusion, $\widetilde{\delta}^T(p_1^m)$, remains unchanged.

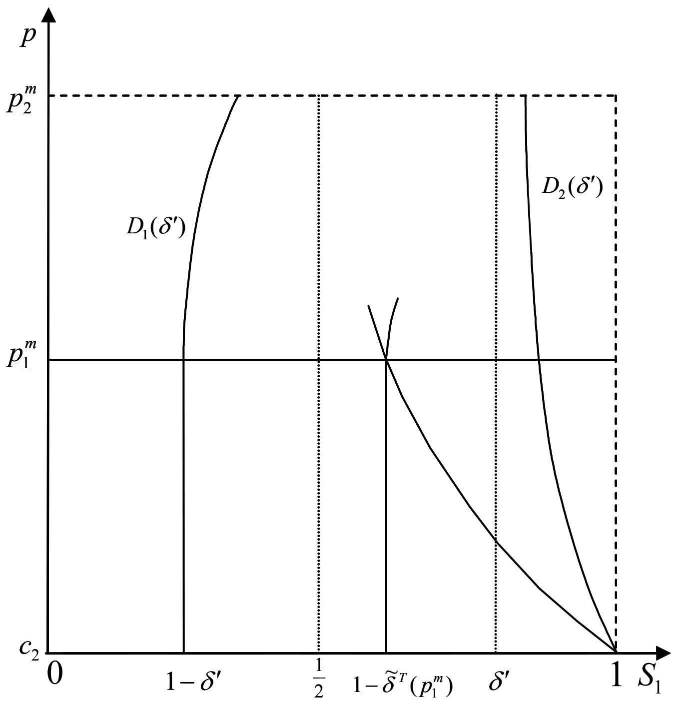

Figure 4: Sustainable Collusive Allocations with Side Payments

As already noted by Bernheim & Whinston (1990), an interesting comparison can be made between these results and those obtained in the context of multi-market contact. When each active firm has a marginal cost advantage in one market, multi-market contact allows the pooling of non-deviation constraints: by shifting sales towards the most efficient firm in each market, collusive profits go up, and the gains from deviating fall. The same mechanism is at work here: side payments allow a shift of sales to the most efficient firm, which raises collusive profits and decreases the deviation gains of less efficient firms.

5.2 The Pareto Frontier of Sustainable Allocations

We now comment on the Pareto frontier of sustainable allocations when there are only two firms and side payments are feasible. As already mentioned, any allocation in which only the low-cost firm is an active producer and maximizes profits by setting its monopoly price is Pareto optimal for the firms.

Interestingly, even if some unconstrained efficient outcomes are sustainable, the firms will not always want to select one of them. To see this, consider any given discount factor $\delta > \widetilde{\delta}^T(p_1^m)$. Then, the preferred outcome of the high-cost firm is $(p_1^m, 1-\delta)$. The low-cost firm's preferred allocation however is *not* such that the price is p_1^m and S as large as possible given the price. In fact, the low-cost firm prefers to move to a price strictly below p_1^m. Such a move has a negative second-order effect on $\pi_1(p_1^m)$, but this effect is more than offset by a positive first-order effect on the low-cost firm's share S, since it alleviates the high-cost firm's no-deviation constraint.

It can be shown that the Pareto frontier of sustainable allocations has a similar shape to that without side payments. For very small discount factors, only allocations on the high-cost firm's non-deviation constraint $D_2(\delta)$ are undominated. For large enough discount factors, the Pareto frontier has two sections: the sustainable part of the unconstrained Pareto frontier ($p = p_1^m$), and part of $D_1(\delta)$ (namely, between p_1^m and some price strictly below p_1^m).

6 Concluding Remarks

By using optimal punishments and allowing for side payments, this paper addresses two largely unexplored aspects in the existing literature on collusion between cost asymmetric firms. We have derived three main results: (i) Without side payments, *some* collusion is sustainable under cost asymmetry whenever collusion is sustainable under cost symmetry. (ii) Without side payments, *efficient* collusion is more difficult when costs are asymmetric. (iii) With side payments, cost asymmetries facilitate collusion. Hence, the general conclusion that cost asymmetry hinders the sustainability of collusion needs to be nuanced. The key policy implication is that the feasibility of side payments between cartel members plays a particularly important role when firms have asymmetric cost structures.

We characterize the maximum scope for collusion in the textbook model of Bertrand competition under cost asymmetry. Our paper thus provides a theoretical benchmark for models that incorporate further restrictions, e.g. budget constraints. Other interesting avenues for future research may be to explicitly model the costs associated with disguising side transfers, or to use more general cost structure to check the robustness of our results.

7 Appendix

Non-Stationary Collusion A (possibly) non-stationary collusive path τ_0 consists of a price vector and a market sharing rule for each period from today to infinity: $\tau_0 = \{P^t, s^t\}_{t=0}^{\infty}$ such that $s_i^t \in [0,1]$, $\sum_i s_i^t = 1$, $s_i^t = 0$ if $p_i^t \notin \min_j\{p_j^t\}$.

Any sustainable collusive path can be supported by the most severe punishment strategies if these are (jointly) credible. Lemma 1 in the main text shows that there exists a subgame perfect strategy profile such that every firm is punished down to minmax continuation profits. Our focus on minmax punishments in the following is therefore without any loss of generality.

We first show that restricting attention to collusive paths along which firms charge a common price in every period does not limit the scope of our analysis of sustainability:

Lemma 5 *Suppose collusion on a path where different firms charge different prices in some period(s) is sustainable. Then there also exists a sustainable collusive path such that all firms charge a common price in each period.*

Proof. Consider any collusive path $\tau_0 = \{P^t, s^t\}_{t=0}^{\infty}$ such that $p_i^t \neq p_j^t$ for some $i \neq j$ at some $t \in [0, \infty)$. Then consider the path $\widehat{\tau}_0 = \left\{\widehat{P}^t, \widehat{s}^t\right\}_{t=0}^{\infty}$ such that $\widehat{p}_i^t = \min_j\{p_j^t\}$ for all i and all t, and $\widehat{s}^t = s^t$ for all t. We now show that if τ_0 is sustainable, then $\widehat{\tau}_0$ is also sustainable.

Each firm's per period profits along τ_0 and $\widehat{\tau}_0$ are the same; the prices at which the good is sold to consumers are the same on both paths, namely $\min_j\{p_j^t\}$ in period t, and market shares are identical. Short-term deviation profits in any period, however, are (weakly) higher along τ_0. The short-term profitability of a deviation only depends on the lowest price charged by any of the other firms in the deviation period. Since $\min_{j\neq i}\{\widehat{p}_j^t\} = \min_j\{p_j^t\} \leq \min_{j\neq i}\{p_j^t\}$ for all i and t, the scope for deviations is therefore always at least as wide along τ_0 as along $\widehat{\tau}_0$. ■

We now turn to non-stationary collusion, assuming that firms indeed charge a common price in every period.

Lemma 6 *If non-stationary collusion is sustainable, then some stationary collusion is also sustainable. More precisely, if non-stationary collusion on a path with $\overline{p} \equiv sup\{p_t\}_{t\geq 0}$ is sustainable, then stationary collusion on any price $p \in (c_2, \min[\overline{p}, p_1^m]]$ is also sustainable.*

Proof. When a deviation triggers a minmax punishment, firm i's non-deviation constraints along collusive path τ_0 are

$$\pi_i(\min[p_i^t, p_i^m]) \leq \sum_{T=t}^{\infty} \delta^{T-t} s_i^T \pi_i(p_i^T) \text{ for } t \in [0, \infty). \tag{8}$$

By backward induction, if one of the non-deviation constraints is violated in any future period, collusion will already break down today.

First suppose that $\overline{p} \equiv sup\{p_t\}_{t \geq 0}$ is finite. Then for any $\varepsilon > 0$, there exists a period t_ε such that $\frac{\pi_i(p_{t_\varepsilon})}{\pi_i(\overline{p})} > (1-\varepsilon)$ for all i. Note also that for small enough ε (in fact, for any $\varepsilon \leq 1$), $m(p_{t_\varepsilon}) = m(\overline{p})$. In the following, we will focus on this case.

We know that $p_1^m < p_2^m < ... < p_n^m$. There is hence some $I \in [0, m(\overline{p})]$ such that $\overline{p} > p_i^m$ for all $i \leq I$, but $\overline{p} \leq p_i^m$ for all $i > I$. For sufficiently small ε, this implies that $p_{t_\varepsilon} > p_i^m$ for all $i \leq I$. The non-deviation constraints in period t_ε for the I most efficient firms ($i \leq I$) are then

$$\pi_i(p_i^m) \leq \sum_{T=t_\varepsilon}^{\infty} \delta^{T-t_\varepsilon} s_T^i \pi_i(p_T). \tag{9}$$

By the definition of the monopoly price, $\pi_i(p_T) \leq \pi_i(p_i^m)$ for all T. Therefore, (9) implies

$$\pi_i(p_i^m) \leq \sum_{T=t_\varepsilon}^{\infty} \delta^{T-t_\varepsilon} s_T^i \pi_i(p_i^m) \text{ for } i \leq I,$$

or equivalently

$$1 \leq \sum_{T=t_\varepsilon}^{\infty} \delta^{T-t_\varepsilon} s_T^i \text{ for } i \leq I. \tag{10}$$

For the less efficient firms $i \in [I+1, m(\overline{p})]$, the non-deviation constraints in period t_ε are

$$\pi_i(p_{t_\varepsilon}) \leq \sum_{T=t_\varepsilon}^{\infty} \delta^{T-t_\varepsilon} s_T^i \pi_i(p_T).$$

Since $p_t \leq \overline{p}$ for all t by definition, $\overline{p} \leq p_i^m$ for $i > I$, and $\pi_i(\cdot)$ is increasing below p_i^m, this condition implies that

$$\pi_i(p_{t_\varepsilon}) \leq \sum_{T=t_\varepsilon}^{\infty} \delta^{T-t_\varepsilon} s_T^i \pi_i(\overline{p}) \text{ for } i > I,$$

which in turn implies

$$(1-\varepsilon) < \sum_{T=t_\varepsilon}^{\infty} \delta^{T-t_\varepsilon} s_T^i \text{ for } i > I. \tag{11}$$

Adding up the necessary conditions in (10) and (11) for all n firms, using the fact that the active firms' market shares add up to 1 in each period, yields the following necessary condition for collusion:

$$I + (m(\overline{p}) - I)(1-\varepsilon) \leq \frac{1}{1-\delta},$$

which rewrites as

$$1 - \frac{1}{m(\overline{p}) - (m(\overline{p}) - I)\varepsilon} \leq \delta.^{23} \tag{12}$$

Now suppose, by contradiction, that the discount factor is $\delta' < 1 - \frac{1}{m(\overline{p})}$. Then there always exists some sufficiently small ε' such that $\delta' < 1 - \frac{1}{m(\overline{p}) - (m(\overline{p}) - I)\varepsilon'}$, so that the necessary condition (12) is violated. This implies that $\delta \geq 1 - \frac{1}{m(\overline{p})}$ is a necessary condition for sustainability when $\overline{p}$ is finite.

We now compare the derived threshold $1 - \frac{1}{m(\overline{p})}$ to the critical discount factor for stationary collusion, $\widetilde{\delta}(p)$. First, if $\overline{p} \leq p_1^m$, then $\widetilde{\delta}(\overline{p}) = 1 - \frac{1}{m(\overline{p})}$. Since $\widetilde{\delta}(p)$ is (weakly) increasing in p, this implies that stationary collusion on any price $p \in (c_2, \overline{p}]$ is sustainable. Second, if $\overline{p} > p_1^m$, then $\widetilde{\delta}(p_1^m) = 1 - \frac{1}{m(p_1^m)} \leq 1 - \frac{1}{m(\overline{p})}$, since $m(p)$ is (weakly) increasing. Since $\widetilde{\delta}(p)$ is increasing, we then know that stationary collusion on any price $p \in (c_2, p_1^m]$ is sustainable.

We still need to consider the case when prices "explode" at some point approaching infinity. However, this analysis is trivial given the previous explanations. In fact, for large enough prices all n firms are "active" and each firm's optimal deviation is to charge its monopoly price. Thus, the necessary condition (12) is simply $1 - \frac{1}{n} \leq \delta$. Since $m(p_1^m) \leq n$, the critical discount factor for stationary collusion on the low-cost firm's monopoly price $\widetilde{\delta}(p_1^m) = 1 - \frac{1}{m(p_1^m)} \leq 1 - \frac{1}{n}$. This implies that stationary collusion on any price $p \in (c_2, p_1^m]$ is sustainable. ■

Proof of Proposition 1. As shown in the text preceding the proposition, the minimum discount factor for which some efficient collusion is sustainable must be such that both firms are indifferent between complying and deviating. The function $\widetilde{\delta}(p)$ was previously defined as the discount factor at which both non-deviation constraints are binding given the price, and for this discount factor only the market sharing rule $s_1 = \widetilde{\delta}(p)$ can be sustained. Define $\widehat{p}$ then as the price at which the two non-deviation constraints are simultaneously binding at a Pareto-efficient market sharing rule $s_1 = s^O(\widehat{p})$:

$$s^O(\widehat{p}) = \widetilde{\delta}(\widehat{p}).$$

[23] This inequality is in fact strict unless $I = n$. In the latter case, we are done at this point, as we have derived the necessary condition $\frac{n-1}{n} \leq \delta$.

Such a $\widehat{p}$ exists and is unique because $s^O(p)$ is strictly decreasing with $s^O(p_1^m) = 1$ and $s^O(p_2^m) = 0$, while $\widetilde{\delta}(p)$ is strictly increasing over the relevant range and $\widetilde{\delta}(p_1^m) = \frac{1}{2}$. It also follows that $\widehat{p} \in (p_1^m, p_2^m)$, which implies that

$$\widehat{\delta} \equiv \widetilde{\delta}(\widehat{p}) > \frac{1}{2}.$$

The Pareto-efficient outcome $(\widehat{p}, \widehat{s})$, where $\widehat{s} \equiv s^O(\widehat{p}) = \widehat{\delta}$, is thus sustainable as long as

$$\delta \geq \widehat{\delta},$$

and $\widehat{\delta}$ is the lowest discount factor for which some efficient collusion is sustainable.

Now suppose $\delta > \widehat{\delta}$. Within the set of Pareto-efficient allocations, the high-cost firm's deviation incentives are increasing and the low-cost firm's deviation incentives decreasing with the collusive market sharing rule. The maximum s_1 must hence be such that the high-cost firm is just indifferent between complying and deviating from the efficient allocation, whereas the minimum s_1 is such that the low-cost firm is indifferent between complying and deviating.

Given any price and the discount factor, the low-cost firm's non-deviation constraint is binding at $\widetilde{s}_1(\delta, p)$. Let $\widetilde{p}^O(\delta)$ be the price as a function of the discount factor such that the low-cost firm's non-deviation constraint is binding at a Pareto-efficient market sharing rule:

$$s^O(\widetilde{p}^O(\delta)) = \widetilde{s}_1(\widetilde{p}^O(\delta), \delta).$$

The function $\widetilde{p}^O(\delta)$ is uniquely defined by this condition, and is strictly increasing in δ. The high-cost firm's non-deviation constraint is binding at a Pareto-efficient allocation market sharing rule for:

$$s^O(p) = \delta.$$

Thus, the range of market sharing rules for which Pareto-efficient collusion is sustainable becomes

$$\left[s^O(\widetilde{p}^O(\delta)), \delta\right].$$

This range is non-empty if and only if $\delta \geq \widehat{\delta}$. ■

Proof of Proposition 3. Part 1: The threshold factor for collusion when side payments are infeasible can be expressed as

$$\widetilde{\delta}(p) = 1 - \frac{1}{\sum_{i \in A(p)} \frac{\pi_i(\min[p, p_i^m])}{\pi_i(p)}},$$

while the threshold for collusion with side payments can be rewritten as

$$\widetilde{\delta}^T(p) \equiv 1 - \frac{1}{\sum_{i\in A(p)} \frac{\pi_i(\min[p,p_i^m])}{\pi_1(p)}}. \tag{13}$$

Since $\pi_1(p) > \pi_i(p)$ for all p and $i \neq 1$, and $A(p)$ includes at least two firms for $p > c_2$,

$$\sum_{i\in A(p)} \frac{\pi_i(\min[p,p_i^m])}{\pi_1(p)} < \sum_{i\in A(p)} \frac{\pi_i(\min[p,p_i^m])}{\pi_i(p)} \text{ for all } p > c_2.$$

From this inequality it directly follows that $\widetilde{\delta}(p) > \widetilde{\delta}^T(p)$ for all $p > c_2$, which establishes the statement of the proposition.

Part 2: Consider a situation with m firms that have symmetric marginal costs equal to c_1. The critical discount factor for collusion (with or without side payments) on some price $p > c_1$ is

$$\frac{m\frac{\pi_1(\min[p,p_1^m])}{\pi_1(p)} - 1}{m\frac{\pi_1(\min[p,p_1^m])}{\pi_1(p)}} = 1 - \frac{1}{m\frac{\pi_1(\min[p,p_1^m])}{\pi_1(p)}}. \tag{14}$$

We now compare (14) to the critical discount factor $\widetilde{\delta}^T(p)$ when firms have asymmetric costs, as given in (13). Since $\pi_1(\min[p,p_1^m]) > \pi_i(\min[p,p_i^m])$ for all p and $i \neq 1$, and $A(p)$ includes at least two firms for $p > c_2$,

$$\sum_{i\in A(p)} \frac{\pi_i(\min[p,p_i^m])}{\pi_1(p)} < m(p)\frac{\pi_1(\min[p,p_1^m])}{\pi_1(p)} \text{ for all } p > c_2.$$

This implies that $\widetilde{\delta}^T(p)$ is smaller than (14) for $m = m(p)$. In other words, the critical discount factor for collusion between one efficient firm and $(m-1)$ firms with marginal costs between c_2 and p is higher below the one for collusion between m efficient firms.

Part 3: The critical discount factor for collusion (with or without side payments) between m symmetric firms on any price between the firms' marginal cost and their common monopoly price is $\frac{m-1}{m}$. Moreover,

$$\widetilde{\delta}^T(p) = 1 - \frac{1}{\sum_{i\in A(p)} \frac{\pi_i(p)}{\pi_1(p)}} \text{ for } p \in (c_2, p_1^m].$$

Since $\pi_1(p) > \pi_i(p)$ for all $i \neq 1$, and $A(p)$ includes at least two firms for $p > c_2$,

$$\sum_{i\in A(p)} \frac{\pi_i(p)}{\pi_1(p)} < m(p) \text{ for } p > c_2.$$

It follows that $\widetilde{\delta}^T(p) < \frac{m(p)-1}{m(p)}$ for $p \in (c_2, p_1^m]$. In this price range, collusion between one efficient firm with cost c_1 and $(m-1)$ firms with marginal costs in $[c_2, p)$ is thus easier to sustain than collusion between m symmetric firms on their monopoly price. This establishes the statement of the proposition. ■

References

[1] Abreu, D. (1986), "Extremal Equilibria of Oligopolistic Supergames", *Journal of Economic Theory*, Vol.39: 191-225.

[2] Abreu, D. (1988), "Towards a Theory of Discounted Repeated Games", *Econometrica*, Vol.56: 383-396.

[3] Bae, H. (1987), "A Price-Setting Supergame between Two Heterogeneous Firms", *European Economic Review*, Vol.31: 1159-1171.

[4] Bain, J. S. (1948), "Output quotas in imperfect cartels", *Quarterly Journal of Economics*, Vol.62: 617-622.

[5] Bernheim, B. Douglas and Michael D. Whinston (1990), "Multimarket Contact and Collusive Behaviour", *Rand Journal of Economics*, 21(1): 1-26.

[6] Blume, A. (2003), "Bertrand without fudge", *Economics Letters*, 78: 167-168.

[7] Compte, O., Jenny, F., and Rey, P. (2000), "Capacity Constraints, Mergers and Collusion", *European Economic Review*, 46: 1-29.

[8] Davidson, C. and Denerecke, R. (1990), "Excess Capacity and Collusion", *International Economic Review*, 31(3): 521-541.

[9] European Commission (2001), "Commission Fines Five Companies in Citric Acid Cartel", Press Release, December 5, 2001.

[10] European Commission (2002), "Commission Fines Five Companies in Sodium Gluconate Cartel", Press Release, March 19, 2002.

[11] Gilo, D, Spiegel, Y. and Moshe, Y. (2004), "Partial Cross Ownership and Tacit Collusion", mimeo, Tel Aviv University.

[12] Hammond, S. D. (2005), "Caught in the Act: Inside an International Cartel", Department of Justice.

[13] Harrington, J. E. Jr. (1991), "The Determination of Price and Output Quotas in a Heterogeneous Cartel", *International Economic Review*, 32(4): 767-792.

[14] Jehiel, P. (1992), "Product differentiation and price collusion", *International Journal of Industrial Organization*, 10: 633-641.

[15] Ivaldi, M., Rey, P., Seabright, P. and J. Tirole (2003), "The Economics of Tacit Collusion", *IDEI Working Paper*, n. 186, Report for DG Competition, European Commission.

[16] Lambson, V. E. (1987), "Optimal Penal Codes in Price-setting Supergames with Capacity Constraints", *Review of Economic Studies*, 54: 385-297.

[17] Lambson, V. E. (1994), "Some Results on Optimal Penal Codes in Asymmetric Bertrand Supergames", *Journal of Economic Theory*, 62: 444-468.

[18] Lambson, V. E. (1995), "Optimal Penal Codes in Nearly Symmetric Bertrand Supergames with Capacity Constraints", *Journal of Mathematical Economics*, 24(1): 1-22.

[19] Pesendorfer, M. (2000)," A Study of Collusion in First-Price Auctions", *Review of Economic Studies*, 67(3): 281-411.

[20] Porter, R. H. (2004), "Detecting Collusion", The Center for the Study of Industrial Organization at Northwestern University, working paper #0051.

[21] Rothschild, R. (1999), "Cartel Stability when Costs are Heterogeneous", *International Journal of Industrial Organization*, 17: 717-734.

[22] Scherer, F. M. (1980), *Industrial Market Structure and Economic Performance*, Houghton Mifflin Company, Boston.

[23] Schmalensee, R. (1987), "Competitive Advantage and Collusive Equilibria", *International Journal of Industrial Organization*, 5: 351-368.

[24] Simon, L. K. and Zame, W. R. (1990), "Discontinuous Games and Endogenous Sharing Rules", *Econometrica*, 58(4): 861-872.

[25] Tirole, J. (1988), *The Theory of Industrial Organization*, MIT Press, Cambridge.

[26] Vasconcelos, H. (2005), "Tacit Collusion, Cost Asymmetries, and Mergers", *Rand Journal of Economics*, 36(1): 39-62.