CESifo WORKING PAPERS

10170
2022

December 2022

Collusion Sustainability with a Capacity Constrained Firm

Leonardo Madio, Aldo Pignataro

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10170*

Collusion Sustainability with a Capacity Constrained Firm

Abstract

We study an infinitely repeated oligopoly game in which firms compete on quantity and one of them is capacity constrained. We show that collusion sustainability is non-monotonic in the size of the capacity constrained firm, which has little incentive to deviate from a cartel. We also present conditions for the emergence of a partial cartel, with the capacity constrained firm being excluded by the large firms or self-excluded. In the latter case, we show under which circumstances the small firm induces a partial conspiracy that is Pareto-dominant. Implications for cartel identification and enforcement are finally discussed.

JEL-Codes: D210, L130, L410.

Keywords: antitrust, capacity constraints, collusion, partial cartel.

Leonardo Madio
Department of Economics and Management
University of Padova
Via del Santo 33
Italy – 35123 Padova
leonardo.madio@unipd.it

Aldo Pignataro
Italian Regulatory Authority for Energy, Networks and Environment (ARERA)
Milan / Italy
aldo.pignataro@unicatt.it

This version: December, 2022
We are grateful to Alan Beggs and two anonymous refeeres for thorough and insightful comments and suggestions. We also thank Giacomo Calzolari, Stefano Colombo, Niccoló Comini, Yassine Lefouili, Ramakanta Patra, Carlo Scarpa, Shiva Shekhar, Sandro Shelegia, Geert van Moer, and participants at the IIOC (2021) and OLIGO (2021) conferences for helpful suggestions on early versions of the paper. The views and opinions expressed in this article do not necessarily reflect the official policy and position of the Authority for which Aldo Pignataro works. Usual disclaimers apply. Leonardo acknowledges financial support from the Unicredit Foundation (Foscolo Europe fellowship).

1 Introduction

It is generally understood that small firms have smaller incentives to join a cartel than large firms (e.g., Bos and Harrington 2010; Compte et al. 2002). The empirical evidence showed that while large firms take part in a conspiracy (if present), the behavior of small firms is mixed, with some cartels featuring their presence and others excluding them. On the one hand, small firms, which are often constrained in their capacity, might collude to benefit from a price hike or avoid punishment from cartel members. For example, Marshall et al. (2019) reported several cases in which a fringe joined a cartel spontaneously (e.g., cartels in *Rubber Chemicals* and *Zinc Phosphate*) and others in which they were pressure to joint the cartel by large firms (e.g., cartels in *Electrical and Mechanical*, *Carbon and Graphite Products*, *Citric Acid* and *Graphite Electrodes*). On the other hand, small firms might prefer not to collude (e.g., cartels in *Vitamins*) or be excluded by large firms because of their weak competitive threat outside the cartel (e.g., cartels in *Industrial Tubes*).

We provide a rationale for the emergence of all-inclusive and partial cartels and how cartel stability is shaped by the size of a small, capacity constrained, firm. Identifying conditions for the emergence of cartels and their structure is key to design effective anti-cartel enforcement (Bos et al., 2018), especially when collusion is harder to detect.[1] This is even more relevant when considering cartels operating in markets in which quantity matters and capacity withholding, or more generally output restrictions, might result in both overcharges and loss of welfare caused by artificial scarcity (Rausser and Stuermer, 2020).[2]

[1]For instance, the OECD (2018) has raised several concerns about the use of algorithms and machine learning that makes collusion easier to sustain and more difficult to be detected (see e.g., Calvano et al. 2020a,b; Harrington 2018; Klein 2021). Due to the rapid evolution of the digital markets, it is therefore critical to identify those factors that facilitate collusion and their application of artificial intelligence in more traditional markets, e.g., gasoline (Assad et al., 2020) and energy (Abada and Lambin, 2020). In the latter case, repeated interactions, quantity competition, and capacity constraints render collusion more akin to the one we focus on.

[2]Rausser and Stuermer (2020) studied the long-term consequences of cartels, focusing on the copper market cartels. In this market, firms compete in quantity and the cartels increase prices implicitly by restricting output and stockpiling. They estimated cartel damages, finding that output damage is more important than price damage. Moreover, there is a fairly large literature dealing with the quantification of cartels' damages. For instance, Connor and Bolotova (2006) showed that cartels entail a notable economic cost for final consumers which consists of a median overcharge price between 17% and 30%, while Smuda (2013) found a mean of 20%. Such a reduction is further harm of a cooperative

In this paper, we explore the link between capacity and collusion and identify conditions for the emergence of a partial cartel in which either the small firm is excluded or is self-excluded. We use a simple oligopoly model in which firms compete on quantity. One of the firms is capacity constrained, whereas others have sufficiently large (or unlimited) capacity. There are several reasons for why a firm might be constrained in capacity, which can depend on the past financial constraints, variation of demand over time, or the intention to merge with other firms in the future.[3] We study cartel stability according to the size of the small firm — i.e., when the small firm has stringent, intermediate, or soft capacity constraints. The cases differ depending on whether the capacity constraint is binding under collusion, competition or deviation.

Our first result shows that the sustainability of a full participation cartel is non-monotonic in the capacity of the small firm. If this firm is small enough (such that it has a meaningless influence on cartel output), there is no strictly positive incentive to deviate from the cartel. As the size of this firm grows, the cartel gets more unstable because of the externality on other cartel members, which have to withhold their capacity. If the capacity of the small firm is intermediate, then collusion becomes easier to sustain again. In this case, all cartel members can at least produce what is agreed within the conspiracy and, importantly, the deviation of the small firm is sub-optimal given its limited capacity. Indeed, the extra-capacity of the small firm compared to the cartel quantity is a threat to firms firms outside the cartel. If the capacity of the small firm is large enough, cartel stability becomes independent of the size of the capacity constrained firm as the large firms' incentive to deviate is independent as well, while the small firm is once again constrained in its deviation. These results are robust to different degrees of product differentiation across firms.

agreement for would-be purchasers, i.e., they would be willing to buy a good at a competitive price but not at the cartel price.

[3]Other reasons for which a firm can have limited capacity result from the dynamics of firms in the industry, due for example to the exit of a firm from the market, which would be the inability of small firms to adjust to the new market conditions. Heterogeneity in capacity can also be dependent on the presence of government incentives to invest in capacity that occur in different periods, resulting in some firms benefiting from them and other firms being exempted. All in all, there are different reasons that might induce firms to invest in capacity and a complete analysis of these incentives is beyond the scope of the paper.

Our second result concerns the underlying conditions for a partial cartel to emerge. A not all-inclusive conspiracy can emerge for two reasons: (i) the small firm finds it optimal not to participate in the cartel and announces its intention; (ii) the small firm is excluded *ex-ante* by the large firms. We find that there exists a parameter range in which the small firm can profitably increase its quantity and the large firms prefer to tolerate such a deviation, as the capacity constraint limits the harm for the colluding firms. Moreover, the small firm may induce a Pareto efficient solution for all market players by announcing not to participate in a cartel when a full participation cartel is not sustainable. Intuitively, as a cartel is more sustainable for fewer and more symmetric participants (Ivaldi et al., 2003), a partial cartel is more likely to emerge as an alternative to a full participation cartel for an intermediate capacity of the small firm.

Alternatively, the large firms might prefer not to run an all-inclusive cartel in order to mitigate the externality that an asymmetric and larger cartel would cause. Large firms can orchestrate the conspiracy on their own without coordination with the small firm. However, the gain from excluding the small firm should be balanced against the cost of facing that firm outside the cartel, exerting competitive pressure. When the capacity constrained firm is sufficiently small, the competitive threat outside the cartel exerted by the small is limited, and thus the large firms prefer an exclusionary strategy.[4]

Among the different industries our model speaks to, the electricity market is certainly the leading example for the ingredients we introduce. This industry is characterized by repeated interactions between a few firms with asymmetric capacity and by quantity-setting.[5] Indeed, firms are in the condition to withhold capacity in order to raise prices during peak hours. Moreover, there is fairly large evidence documenting the ability of few large generators to exercise market power and sustain collusion, which is often enforced

[4]This result is in line with the findings of Harrington et al. (2018). The authors showed empirically that German cement manufacturers were executing a bribery scheme between 1991 and 2002 only if the threat of non-cartel competitors was considered sufficiently aggressive. The paper also provides a theoretical appraisal of the incentives of the cartel members.

[5]For example, the British electricity market in 1989 was composed of National Power, controlling 30 GW of generating capacity, PowerGen, controlling 20 GW, and Nuclear Electric with a small quota of 9 GW. Similarly, the Spanish electricity market in 1998 was composed of Endesa and Iberdrola, which controlled 84% of generating capacity, and Unión Fenosa and Hidrocántabrico controlling the remaining 16% (see Dechenaux and Kovenock 2007).

through strategic capacity withholding (see e.g., Bergler et al. 2017, Fabra and Toro 2005, Fogelberg and Lazarczyk 2014, Kwoka and Sabodash 2011, Wolfram 1998, 1999).[6] In this industry, the emergence of cases is therefore not surprising (e.g., *Tolling Edipower*).[7]

The rest of the article is organized as follows. In the next section, we discuss our contribution to the literature. In Section 3, we introduce the baseline model. In Section 4, we characterize the equilibrium of the stage game. Section 5 analyzes an all-inclusive cartel. In Section 6, we consider the incentives of unconstrained and capacity constrained firms to run a partial cartel. Section 7 discusses some extensions and provides some robustness checks. Finally, we present concluding remarks and discuss policy implications.

2 Related Literature

This paper contributes and relates to several strands of the literature.

Asymmetries and collusion. Economists and policy-makers tend to agree that symmetry between firms facilitates the sustainability of a tacit cartel (Bernheim and Whinston, 1990; Ivaldi et al., 2003). Despite this common wisdom, several papers analyze the relationship between cartel stability and asymmetries in costs (e.g., Miklos-Thal 2011, Rothschild 1999 and Vasconcelos 2005), product differentiation (e.g., Bos and Marini 2019; Bos et al. 2020; Chang 1991; Lambertini and Sasaki 1999; Ross 1992; Tyagi 1999) and the number of products (e.g., Khün 2004). While all these models identify the different incentives to cheat among firms and the difficulty in punishing deviating behaviors, results often differ depending on the cost structure, type of competition, and punishment strategies. This literature shares with us the idea that even small degrees of firms' asymmetries might affect collusion sustainability, thus explaining the emergence of heterogeneous and asymmetric cartels (see e.g., see also empirical evidence as in Davies and Olczak 2008;

[6]These markets have been extensively examined from a theoretical point of view by Benjamin (2016), Dechenaux and Kovenock (2007) and Fabra (2003), who provided evidence about the relationship between capacity withholding and collusion sustainability.

[7]The Italian Antitrust Authority (AGCM) contested a possible agreement between power generators (which are capacity constrained by definition) in Sicily. They aimed at coordinating their strategies, to keep high prices during peak hours, through capacity withholding, and possibly gaining more from secondary transactions. See the AGCM case *I721 - Tolling Edipower*.

Levenstein and Suslow 2006). We focus on the interplay between capacity constraints and collusion sustainability.

Capacity constraints and collusion. An active body of literature has sought to study collusion and capacity constraints (Bos and Harrington, 2010, 2015; Compte et al., 2002; De Roos, 2004; Fabra, 2006; Garrod and Olczak, 2018). A common result in this literature is that small firms have fewer incentives to join a conspiracy than large firms. In particular, Compte et al. (2002), in a Bertrand-Edgeworth setting, showed that while collusion becomes more difficult to sustain when the aggregate capacity is large, the presence of asymmetric capacities helps collusion when the aggregate capacity is limited. Similarly, Bos and Harrington (2010) studied endogenous cartel formation when capacity constraints are heterogeneous across firms. They showed that small firms would be more likely not to join the cartel while taking advantage of the cartel price set by the larger firms.

Our approach and mechanisms differ in several dimensions. While these papers consider price competition, we analyze a quantity-setting game with (potentially) differentiated products. We focus on quantity-setting as quantity represents the strategic variable for which firms are asymmetric and several markets in which capacity matters are characterized by quantity-setting (e.g., energy sector, copper). In such cases, a capacity constrained firm is not allowed to freely choose the quantity that would maximize profits. An important difference between our framework and theirs is the focus on a small capacity constrained firm rather than on multiple heterogeneous firms. Focusing on a capacity constrained firm allows us to isolate the effect of one source of asymmetry from the aggregate variation in the available capacity. For example, a capacity constrained firm may not be in the condition to match cartel requests or be aggressive in the market if defecting, revealing therefore the real threat of a small firm and affecting the consequent response of the cartel members (see De Roos 2004).[8] Finally, another difference between our setting and theirs is the focus on differentiated products. Such a difference matters for the incentives of the large firms to sustain a cartel. Contrary to common wisdom for

[8]Similar results are also likely to apply in the presence of two large unconstrained firms and a fringe of (symmetric) capacity constrained firms.

which a higher degree of substitutability across products hinders collusion (Deneckere, 1983; Wernerfelt, 1989), we show that collusion might be facilitated, other things equal, in the presence of a capacity constrained firm if its capacity is intermediate.

Our paper adds to this literature in that we find that small firm has a larger incentive to join a cartel when sufficiently small as the limited capacity would discipline their participation in the cartel. By contrast, large firms would prefer to exclude the small firm and mitigate the externality that an additional cartel member would create as long as the cost of competing with the small firm is bearable. The two effects together highlight a *novel* non-monotonicity in the sustainability of a full cartel. This setting is reminiscent of Brock and Scheinkman (1985), who documented how limited capacity alters the incentive to collude and how the entry of one capacity constrained firm can foster the sustainability of a tacit cartel. In their paper, the presence of a large number of firms leads to an increase of the total capacity and, therefore, the price falls in case of retaliation. In our model, this result also emerges if the capacity of the small firm is slightly above that threshold.

Partial cartel. This paper also shares connections with the literature on partial cartels. The closest studies are Brandão et al. (2014), De Roos and Smirnov (2021), Hasnas and Wey (2015), and Shaffer (1995). Shaffer (1995) endogenized cartel formation in the presence of a Cournot fringe and showed that the latter prefer to compete against the cartel with simultaneous moves. However, the cartel can induce a Stackelberg equilibrium in which it moves first if it can threaten effectively the fringe firms in light of its significant market power. In our model, we introduce capacity constraints and show that the small firm might refrain from joining a cartel and this can represent a Pareto-dominant solution.

Brandão et al. (2014) examined partial cartel sustainability between two incumbents, and the role played by asymmetries between firms and the order of the moves (e.g., simultaneous or sequential). They found that, depending on differences in efficiency, full collusion between incumbents and an entrant can be accommodated more or less often than partial collusion between incumbents. In this paper, instead, it is the size of the capacity constrained firm that might facilitate or hinder collusion in its different forms.

Finally, Hasnas and Wey (2015) showed the existence of partial cartels in a price-setting game in which products are differentiated both horizontally and vertically. They analyzed the incentive to collude between three heterogeneous firms and prove that the critical discount factor of the high-quality sellers is key to sustain a cooperative agreement. Our findings echo this result when considering the self-enforceable constraint to sustain a cartel for the large firms. However, focusing on asymmetries in capacity constraints allows us to study the role of the residual demand and its impact on the emergence of partial cartels.

3 The Model

Consider an industry with $n = 3$ firms playing an infinitely repeated quantity-setting game and selling differentiated products at a constant marginal cost, which is normalized to zero without loss of generality.[9] For the sake of simplicity, let us assume that one firm (Firm 1, *henceforth*) has a limited capacity $k > 0$, while the other two firms (Firm 2 and Firm 3, *henceforth*) are symmetric and operate with unlimited capacity, i.e., they can costlessly adjust their supply rapidly to satisfy the entire demand. Assume a representative consumer with preferences à la Singh and Vives (1984).[10] When buying from firm $i = 1, 2, 3$, with $j \neq i$, the problem of a representative consumer is

$$\max_{q_1,q_2,q_3} \sum_{i=1,2,3} \Big(U(q_i) - p_i q_i\Big) \tag{1}$$

where p_i is the price of firm i and the smooth and strictly concave utility function is

$$U(q_i) = q_i\Big(\theta - \frac{1}{2}\beta q_i - \gamma \sum_{j\neq i} q_j\Big). \tag{2}$$

[9]This assumption is common in many theoretical papers. A notable exception is Bos et al. (2020) who showed that in the presence of costs varying with quantity or quality results might change. Although asymmetric capacity constraints might be considered a limit case of asymmetric marginal costs, the two cases starkly differ in the repeated game, as the incentives to participate in a cartel are stronger for a capacity constrained rather than for a less efficient firm. Details are available upon request.

[10]This model can be traced back to Levitan and Shubik (1978), as discussed in a recent contribution by Choné and Linnemer (2020).

Let $\gamma \in [0, \beta)$, with $\beta > 0$, be a parameter measuring substitutability between products such that $\gamma = 0$ implies independence, while products become perfect substitutes as γ approaches β. Thus, the case of homogeneous products is a special case of the current setup. Denote $\theta > 0$ the state of the *aggregate demand* in the economy, which is assumed to be sufficiently large so that consumers always buy at equilibrium. Maximizing with respect to q_i, with $i = 1, 2, 3$ yields the following inverse demand function:

$$p_i(q_i, q_{-i}) = \theta - \beta q_i - \gamma \sum_{j \neq i} q_j. \tag{3}$$

Firms are long-lived and discount future profits at a common discount factor, $\delta \in (0, 1)$, whereas consumers are short-sighted. In every period ($\tau = 1, 2..., \infty$), firms choose their quantities simultaneously. There is perfect monitoring, which implies that all past decisions are common knowledge. We follow Friedman (1971) and assume that firms adopt a grim trigger strategy: collusion is sustained by means of infinite Nash reversion, such that any deviation from the cooperative agreement is punished by playing the competitive equilibrium of the stage game in all the subsequent periods. We also assume colluding firms share the market equally, unless the small firm has tight constraints. The equilibrium concept is Subgame Perfect Nash Equilibrium (SPNE). Therefore, collusion is sustainable as SPNE if, and only if,

$$\frac{\pi_i^c}{1 - \delta} \geq \pi_i^d + \frac{\delta}{1 - \delta} \pi_i^p.$$

where π_i^c, π_i^d and π_i^p denote the one-shot collusion, deviation and punishment profits respectively. This implies that a cartel is stable as long as the long-term profits from colluding are larger than the short-term gains from cheating. Therefore, there exists a critical value of the discount factor

$$\delta_i = \frac{\pi_i^d - \pi^c}{\pi^d - \pi^*} \tag{4}$$

such that the cartel is stable if $\delta \geq \max_{i \in \{1,2,3\}} \{\delta_i\}$.

4 Stage Game

In this section,we focus on the competitive equilibrium that characterize the punishment phase in case of any deviation from the collusion path.

Consider first the case in which Firm 1's capacity is not binding. Then, quantity is symmetric across firms, which maximize their own profit functions as follows

$$\max_{q_i} q_i(\theta - \beta q_i - \gamma \sum_{j \neq i} q_j) \qquad i = 1, 2, 3.$$

From the first order conditions, we obtain the following equilibrium quantities and profits

$$q_i = \frac{\theta}{2(\beta + \gamma)} \equiv q^*, \qquad \pi_i = \frac{\beta\theta^2}{4(\beta + \gamma)^2} \equiv \pi^*. \tag{5}$$

Consider now the case in which the capacity constraint is binding under competition — i.e., $k < q^*$. This implies that $q_1^* = k$, while the quantity set by the unconstrained firms is the result of the following maximization problem

$$\max_{q_i} q_i(\theta - \beta q_i - \gamma(k + q_j)) \qquad i = 2, 3,$$

where q_j is the quantity set by the unconstrained rival. Differentiating with respect to q_i yields the following best-reply function, which is decreasing both in the capacity constraint of Firm 1 and the quantity set by the remaining rival,

$$q_i(q_j, k) = \frac{1}{2\beta}(\theta - \gamma(q_j + k)).$$

Thus, equilibrium quantities and profits of Firm 2 and 3 are respectively

$$q_2^*(k) = q_3^*(k) = \frac{\theta - k\gamma}{2\beta + \gamma}, \qquad \pi_2^*(k) = \pi_3^*(k) = \frac{\beta(\theta - k\gamma)^2}{(2\beta + \gamma)^2}. \tag{6}$$

Firm 1 operates at full capacity k and obtains

$$\pi_1^*(k) = \frac{k\left(\beta(2\theta - k\gamma) - \theta\gamma + 2k(\gamma^2 - \beta^2)\right)}{2\beta + \gamma}. \tag{7}$$

Therefore, there is a threshold of k such that: (i) for $k \geq q^*$, the stage game features a symmetric equilibrium in which firms set a quantity $q^* = \frac{\theta}{2(\beta+\gamma)}$; (ii) for $k < q^*$, there is an asymmetric equilibrium in which the capacity constrained firm sets a quantity equal to k, whereas the other firms set $q_i^*(k) = \frac{\theta - k\gamma}{2\beta+\gamma}$ with $i = 2, 3$.

5 Full Cartel

In this section, we study the role of capacity constraints in shaping the sustainability of a full participation cartel. To isolate effects at stake, we first present a benchmark model in which there is no capacity constraint and all firms are symmetric. We then study firms' incentives to collude for different capacity constraints of the small firm.

5.1 Benchmark: no capacity constraints

Suppose all firms are symmetric and there is unconstrained capacity.[11] Suppose that a cartel exists, it is stable and all-inclusive, and maximizes the joint profits, i.e.,

$$\max_{q_i} \sum_{i=1,2,3} q_i(\theta - \beta q_i - \gamma \sum_{j \neq i} q_j),$$

which leads to the following symmetric equilibrium quantities and profits

$$q_i = \frac{\theta}{2(\beta + 2\gamma)} \equiv q^c, \qquad \pi_i = \frac{\theta^2}{4(\beta + 2\gamma)} \equiv \pi^c, \tag{8}$$

with the superscript 'c' for *collusion* when Firm 1 does not have binding capacity. Note that $q^c < q^*$. A cartel is sustainable in the long-run if no firm has incentive to deviate.

[11] This standard textbook example allows us to gain insight on how capacity constraints may affect firms' profits and the sustainability of collusion.

The deviation quantity of firm i, denoted by q_i^d, solves

$$\max_{q_i^d} q_i^d \left(\theta - \beta q_i^d - 2\gamma \frac{\theta}{2(\beta + 2\gamma)} \right).$$

Therefore, we have the following deviation quantity and profit,

$$q_i^d = \frac{\theta(\beta+\gamma)}{2\beta(2\gamma+\beta)} \equiv q^d, \qquad \pi_i^d = \frac{\theta^2(\beta+\gamma)^2}{4\beta(\beta+2\gamma)^2} \equiv \pi^d. \tag{9}$$

Note that $q^d > q^*$. By comparing deviation, collusion and punishment profits in the benchmark case, one can verify under which condition a cartel is sustainable, that is by using (4), if

$$\delta \geq \frac{(\beta+\gamma)^2}{2\beta^2 + 4\beta\gamma + \gamma^2} \equiv \delta^*, \tag{10}$$

The critical discount factor is increasing in γ — i.e., the larger the degree of substitutability among products, the less sustainable the cartel.

5.2 The role of capacity constraints

The above benchmark allows us to identify the classical incentives to leave a cartel in the absence of capacity constraints. When capacity constraints are present, four different regions of k emerge: (i) $k \in [0, q^c)$ when Firm 1's capacity constraint is always binding; (ii) $k \in [q^c, q^*)$ which implies that Firm 1's capacity constraint is not binding only if firms collude and collusion is sustainable; (iii) $k \in [q^*, q^d)$ if Firm 1's capacity constraint is not binding in equilibrium but only if it deviates from the cartel; (iv) $k \geq q^d$, which is the benchmark case described in Section 5.1.

Having already discussed how market forces behave in the region (iv), in what follows, we study the sustainability of the cartel in the remaining regions, that is when the critical discount factor might depend on k. Differentiating (4) with respect to k we obtain

$$\frac{\partial \tilde{\delta}}{\partial k} = \frac{1}{(\pi^d - \pi^*)^2} \left((\pi^d - \pi^c) \frac{\partial \pi^*}{\partial k} + (\pi^c - \pi^*) \frac{\partial \pi^d}{\partial k} + (\pi^* - \pi^d) \frac{\partial \pi^c}{\partial k} \right) \tag{11}$$

where subscripts are omitted. In what follows, we identify the sign of the partial derivatives (whenever differentiable in k) and discuss the main effects at stake for stringent (i), intermediate (ii) and soft (iii) capacity constraints.

Case (i): Stringent capacity constraint. Suppose Firm 1 has binding capacity, i.e., $k \in [0, q^c)$. Within and outside a cartel, this firm cannot do anything different from making available its entire capacity k. In this case, the cartel solves the following problem,

$$\max_{q_i} \sum_{i=2,3} q_i(\theta - \beta q_i - \gamma q_j - \gamma k) + k(\theta - \beta k - \gamma(q_i + q_j)).$$

In turn, equilibrium quantity and profits for Firm 2 and 3 are

$$q_2^c(k) = q_3^c(k) = \frac{\theta - 2k\gamma}{2(\beta + \gamma)}, \qquad \pi_2^c(k) = \pi_3^c(k) = \frac{\theta(\theta - 2k\gamma)}{4(\beta + \gamma)} \tag{12}$$

which decrease in k. Indeed, the more stringent the capacity constraint, the larger the market shares that Firm 2 and 3 can serve. Instead, Firm 1 operates at full capacity and its collusive quantity coincides with k, which leads to

$$\pi_1^c(k) = \frac{k\left(\beta(\theta - k\gamma) + k(2\gamma^2 - \beta^2)\right)}{\beta + \gamma}. \tag{13}$$

In this region of parameters, the profit of all firms depend on k. However, Firm 1 might deviate from a collusive agreement only by offering a quantity lower than k, but this would have a negative impact on its profits. In turn, there is no strictly positive incentive to deviate and, therefore, the critical discount factor above which Firm 1 has incentives to participate in the cartel is equal to zero. Accordingly, the sustainability of the cartel for a given $k \in [0, q^c)$ depends entirely on the incentives of the large firm. We focus on (11) for the two unconstrained firms[12] and observe that Firm 2 (or alternatively 3), is impacted as follows

$$\frac{\partial \pi_2^d(k)}{\partial k}\Big|_{k\in[0,q^c)} < 0, \quad \frac{\partial \pi_2^*(k)}{\partial k}\Big|_{k\in[0,q^c)} < 0, \quad \frac{\partial \pi_2^c(k)}{\partial k}\Big|_{k\in[0,q^c)} < 0, \tag{14}$$

[12] All the technical details are relegated to the Appendix.

Interestingly, as k increases, all profits decrease and the sign of (11) is a-priori ambiguous, as shown in the following

$$\left.\frac{\partial \tilde{\delta}_2}{\partial k}\right|_{k \in [0,q^c)} = \frac{1}{(\pi_2^d(k) - \pi_2^*(k))^2}\Bigg(\underbrace{(\pi_2^d(k) - \pi_2^c(k))\frac{\partial \pi_2^*(k)}{\partial k}}_{(-)} + \underbrace{(\pi_2^c(k) - \pi_2^*(k))\frac{\partial \pi_2^d(k)}{\partial k}}_{(-)} + \underbrace{(\pi_2^*(k) - \pi_2^d(k))\frac{\partial \pi_2^c}{\partial k}}_{(+)}\Bigg).$$

Note that the effect of k on the critical discount factor is positive only if the last term — i.e., the market sharing effect under collusion — is sufficiently positive. In the Appendix, we observe that the (negative) impact of k on collusive profits outweighs the (negative) impact of k on both deviation and punishment profits — i.e., $|{}^{\partial \pi_2^c(k)}/_{\partial k}| > |{}^{\partial \pi_2^d(k)}/_{\partial k}|$ and $|{}^{\partial \pi_2^c(k)}/_{\partial k}| > |{}^{\partial \pi_2^*(k)}/_{\partial k}|$ with $i = 2, 3$. Therefore, the market sharing effect prevails of the others and the critical discount factor is increasing in k.

The intuition behind the above result is as follows. For $k = 0$, there is a standard duopoly with only two unconstrained firms. Thus, the cartel is sustainable if these two firms have no incentives to deviate. A strictly positive k implies the entry of a new rival in the market that destabilizes the cartel: even if the third firm is capacity constrained, its entry in the market makes collusion more difficult to sustain. At first sight, it might seem that more symmetry among firms — i.e., the lower the difference in the available capacity — makes collusion harder to sustain, which is in sharp contrast with most of the literature on cartel sustainability (see e.g., Ivaldi et al. 2003 and Motta 2004). Nevertheless, for stringent capacity, any increase in k does not eliminate the asymmetry between constrained and unconstrained firms as the constrained firm continues to be limited in size.[13]

Case (ii): Intermediate capacity constraint. Suppose Firm 1 is not so small, that is, it has binding capacity under competition but not under cartel, i.e., $k \in [q^c, q^*)$. This

[13]Note that, at $k = 0$, an increase in k introduces an asymmetry in the market, making collusion harder. However, as previously described, we focus on an industry with $n = 3$ firms and we do not consider the entrance of a new small firm, which clearly makes the cartel more unstable. Note that the critical discount factor (shown in the Appendix) is increasing in γ, whereas the critical discount factor for the small firm is equal to 0. As a result, the difference in the two discount factors is also increasing in the degree of substitutability, rendering the incentives to join a cartel more asymmetric.

implies that quantity and collusive profits coincide with (8). In this region of parameters, the optimal deviation for this firm consists in offering its entire capacity k, because the optimal unconstrained deviation would exceed its capacity.

In this scenario, the small firm has now a higher incentive to leave the cartel as a larger capacity implies a higher deviation profit. Punishment profits are increasing in k, thereby making collusion more difficult to sustain because any deviation is punished less harshly, cartel payoffs are constant and independent of Firm 1's capacity, i.e., ${}^{\partial \pi^c}/_{\partial k} = 0$. Therefore, we can rewrite (11) as

$$\left.\frac{\partial \tilde{\delta}_1}{\partial k}\right|_{k\in[q^c,q^*)} = \frac{1}{(\pi_1^d(k) - \pi_1^*(k))^2}\Bigg(\underbrace{(\pi_1^d(k) - \pi^c)\frac{\partial \pi_1^*(k)}{\partial k}}_{(+)} + \underbrace{(\pi^c - \pi_1^*(k))\frac{\partial \pi_1^d(k)}{\partial k}}_{(+)}\Bigg) > 0$$

showing a positive impact on the critical discount factor.

Nevertheless, what guides the sustainability of the cartel is the critical discount factor of the unconstrained firms. For them, the optimal deviation quantity and profits coincide with (9), that is, the strategy adopted when all firms are fully symmetric. Indeed, both deviation profits and cartel profits do not depend on k, which then implies that the sign of (11) is the same as the sign of the effect of k on the competition profit. This is negative (as shown in the Appendix). The intuition is that the extra-capacity of the small firm compared to the cartel quantity represents a competitive threat outside the cartel. Thus, holding other things equal, a cartel becomes more stable.[14]

Case (iii): Soft capacity constraint. Suppose now Firm 1's capacity is binding under deviation but not under competition, i.e., $k \in [q^*, q^d)$. In this case, unconstrained firms' profits are not affected by the size of k in any of the considered scenario. This means that the critical discount factor for unconstrained firms is flat and equal to δ^* for any $k \geq q^*$. The collusive quantity and profit of Firm 1 still coincide with (8), therefore its optimal choice is between restricting capacity to its cartel level q^c or engaging in a

[14] Note that this result brings out an alert toward those policies fostering the investments in more capacity. See, for example, in the energy market, the capacity remuneration mechanisms introduced in many European countries in order to achieve long-run efficiency goals.

strategic deviation, with the maximum quantity to be set equal to $k < q^d$. Therefore, k has an effect only on Firm 1's deviation profits (as shown in the Appendix), being its capacity sufficiently large such that the equilibrium quantities, both under collusion and competition, are the same across firms. Clearly, the larger the Firm 1's capacity the larger the deviation profits as far as $k < q^d$. Hence, it is intuitive that the critical discount factor for Firm 1 is increasing in k and approaches δ^* as k tends to q^d.

Finally, as a special case, for any $k \geq q^d$, symmetry and unconstrained capacity is restored and the critical discount factor is equal to δ^* for all firms.

As the sign of the partial derivatives in (11) might or might not depend on k, with some effects being negative while others positive in the relevant parameter ranges, this entails an interesting non-monotonicity. We conclude the following.

Proposition 1. *A full participation cartel is sustainable if and only if unconstrained firms have no incentive to deviate. The critical discount factor*

- *for the unconstrained firms is strictly increasing in k for $k \in (0, q^c)$ and weakly decreasing otherwise;*
- *for the capacity constrained firm is zero for $k \in (0, q^c)$ and weakly increasing in k otherwise.*

The two critical discount factors coincide and become equal to δ^*, being independent of k, for any $k \geq q^d$.

Proof. See Appendix. □

Proposition 1 shows that the critical discount factor of the capacity constrained firm weakly increases in k. When its capacity is low enough — i.e., $k \in [0, q^c)$ — the capacity constrained firm does not have a strictly positive incentive to leave the cartel since its critical discount factor is flat and equal to zero. When its capacity constraint is intermediate or soft — i.e., $k \in [q^c, q^d)$ — the critical discount factor becomes positive and (strictly) increases in k. In turn, a deviation becomes relatively more likely. Finally, the critical discount factor continues to increase in k, up to the point in which it becomes flat and all firms are symmetric — i.e., for any $k \geq q^d$.

A second result concerns the incentive to deviate for the unconstrained firms is non-monotonic. When the capacity constrained firm is small, the critical discount factor strictly increases in k which renders collusion harder to sustain. In this case, the entry of a small firm in the market reduces cartel profits more than deviation and punishment profits. Interestingly, there exists a turning point, which coincides with $k = q^c$, that creates a non-monotonicity, i.e., the critical discount factor to be weakly decreasing in k (because cartel profits are no longer affected by any increase in k) up to the point it becomes flat and independent of k — i.e., $k \geq q^*$. Albeit decreasing in k, the critical discount of the large firms remains always higher than that of the capacity constraint firm, which then implies that if they have no incentive to deviate from a collusive equilibrium, then the small firm has no incentive as well.

Figure 1 illustrates the evolution of the two critical discount factors as a function of k. The dashed line identifies the critical discount factor of Firm 1, whereas the solid line identifies the critical discount factor of the unconstrained firms.[15]

Moreover, we also observe that, contrary to conventional wisdom, small firms have a larger incentive to be part of a conspiracy the smaller they are. Noteworthy, the size of the capacity constrained firm also matters for the incentives of the other firms to sustain a cartel and their incentives to leave the conspiracy are (weakly) non-monotonic in Firm 1's capacity. This result is reported in the following corollary.

Corollary 1. *A capacity constrained firm has a higher incentive to be part of a conspiracy the smaller its capacity.*

So far, we have shown how the size of the small firm impacts the collusive outcome. However, also product substitutability, proxied by γ, is likely to play a role. Previous

[15] Our motivating examples mostly referred to energy markets in which products are homogeneous. In the Appendix, we show that the figure can be replicated identically when allowing for homogeneous products and perfect substitutability, i.e., $\beta = \gamma$.

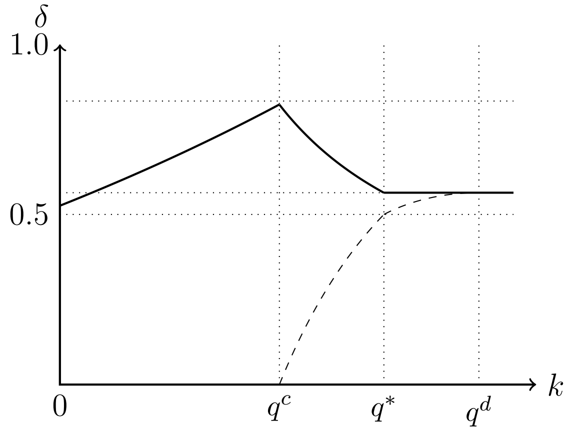

Figure 1: Critical discount factors for the capacity constrained firm (dashed line) and the unconstrained firms (solid line) as a function of k ($\theta = 3, \beta = 1.1, \gamma = 1$).

literature has shown that a higher degree of substitutability among products in quantity-setting games hinders collusion (Deneckere, 1983; Wernerfelt, 1989), unless, for example, there are network externalities (Song and Wang, 2017). In our framework, while for soft and stringent capacity constraints of the small firm, the larger firms have fewer incentives to join a cartel (as the critical discount factors increase in γ), in the presence of intermediate capacity the reverse holds. Specifically, in the latter case, as γ increases, the small firm exerts a competitive pressure during the punishment to the large firms which, in addition to the reduction in the deviation profits, more than compensates for the externality exerted under collusion. Using the terminology of Tyagi (1999), increased product substitutability leads to higher gains from collusion that more than compensate for the gain from deviation. As a result, the large firms have more incentives to form a cartel rather than competing aggressively with the small firm outside the cartel. The following proposition summarizes the above discussion.

Proposition 2. *The incentive to collude is U-shaped in the degree of product differentiation.*

As the cartel sustainability depends entirely by the incentives of the large firms, when γ increases collusion is more likely to occur, other things being equal, for intermediate capacity. This result, which echoes those of Ross (1992) and Tyagi (1999), bears some important implications for Antitrust Authorities as the incentives to collude become more similar across (asymmetric) firms. This is because the difference in the critical value factors between the small and the large firm decreases.[16]

6 Partial Cartel

So far, we have studied the sustainability of a cartel between all the market players and how the size of the small firm affects the incentives to be part of the conspiracy. However, there might be cases in which a partial cartel can be sustained only by large firms. To study under which conditions partial cartels can emerge, we study a sequential game in each period τ of the game. In the first stage, firms decide whether to join (or let the other firms join) the cartel. In the second period, firms decide about quantity. In Section 6.1, we examine the case in which the small firm can induce a partial cartel by credibly announcing not to join in, whereas in Section 6.2, we examine the case in which the large firms exclude the small firm from the conspiracy. Finally, in Section 6.3, we analyze all the equilibrium outcomes and identify cartel formation for different values of k.

6.1 Induced Partial Cartel

In this subsection, we study whether there is any incentive for a small firm to manipulate the other firms' strategies by announcing in advance that it will not participate to the cartel. Such a situation may arise when firms have informal meetings and leak their strategies, wherein firms, in each period τ, credibly commit (or announcements) on their future strategies before setting quantities. For example, such a credible announcement could be the decision not to take part in these meetings.

To study firms' incentives, we restrict attention to the case in which the small firm may have an incentive not to take part in the cartel (i.e., $k > q^c$), as for stringent capacity Firm 1 cannot increase its profit by withholding its capacity. Moreover, we also restrict

[16] Note that, in this parameter range, the critical discount factor of the small firm increases with γ.

attention to the case in which asymmetries among participants are still present (see e.g., Bos and Harrington 2010) as these make partial cartels more likely, thus $k < q^d$.

The underlying idea is that, being Firm 1 capacity constrained, if its commitment not to join the cartel at the initial stage of each period is sufficiently credible, then some deviations from the cartel quantity can be bearable by the larger firms. They may prefer to sustain a partial cartel instead of competing harshly with Firm 1.[17] This allows the small firm to act first by claiming not to take part in the cartel. This strategy can be profitable for Firm 1 when a partial cartel can be sustained by the larger firms, which make their decisions for given quantity of Firm 1.

Suppose the small firm credibly commits not to participate in the conspiracy and to produce at full capacity.[18] Such a choice is inferred by the quantity announced by Firm 1 in the first period. Therefore, the game becomes sequential and, in the second stage of each period, Firm 2 and 3 maximize their joint profits to obtain

$$q_i^{pc}(k) = \frac{\theta - k\gamma}{2(\beta+\gamma)}, \qquad \pi_i^{pc}(k) = \frac{(\theta - k\gamma)^2}{4(\beta+\gamma)}, \qquad i = 2, 3. \tag{15}$$

The quantities set in this case are larger than those in (8). In a partial cartel, the unconstrained firms no longer need to discount heavily the quantity k offered by the capacity constraint firm. Indeed, under partial cartel, unconstrained firms do not take into account the negative externality of their offers on the Firm 1's profit. This leads them to offer a higher quantity and even obtain higher profits compared to those obtained in an all-inclusive cartel, when k is not too large relative.[19]

By anticipating the behavior of the cartel members, Firm 1 always makes available its entire capacity, k, for any $k \in [q^c, q^d)$. The reason is that Firm 1 would find it optimal

[17] This is akin to the stability assumptions of a partial cartel used by Hasnas and Wey (2015): if one firm does not take part in the full cartel, the other firms will continue to collude if profitable.

[18] Note that, in this parameter range, Firm 1 may withhold its capacity and gain more than in a full cartel. However, making available its entire capacity maximizes its profit, then it represents the optimal deviation strategy for Firm 1. This result is proved by backward induction in the Appendix. Instead, if we simply assume that Firm 1 does not participate in the conspiracy and sets quantities as in a competitive market, it will produce at full capacity for any $k \leq q^*$, otherwise $q_1 = q^*$. However, this strategy would not represent an optimal deviation strategy as shown in the Appendix.

[19] In the following section, we characterize the threshold of k under which large firms' profits are higher under a partial cartel than under a full participation cartel.

to anticipate cartel members by committing to produce a higher quantity. However, as the optimal quantity is larger than q^d — the maximum quantity Firm 1 would be able to produce —, the best Firm 1 can do is to operate at full capacity and earn

$$\pi_1^{pc}(k) = \frac{k(\beta(\theta - k\gamma) + k(\gamma^2 - \beta^2))}{(\beta + \gamma)}. \tag{16}$$

The small firm faces a trade-off between participating in the collusion (and earning a stream of collusive profits as in (8)) and enjoying deviation profits in the subsequent periods (when the partial cartel is sustainable). Comparing static profits in the two cases, one can observe that the capacity constrained firm has an incentive to induce a partial cartel only if its capacity is not small enough. The rationale is the following. If Firm 1's capacity is sufficiently small in this parameter range and, hence, close to the equilibrium quantity under a full participation cartel, then the marginal gains from inducing a partial cartel are very limited as each cartel member expands individual equilibrium quantity, making Firm 1's deviation less appealing. Indeed, being part of a conspiracy is optimal for the small firm. On the contrary, when Firm 1's capacity is not sufficiently far from q^c, the marginal gains from operating at full capacity become large enough to make such a deviation strategy more profitable than participating in the conspiracy. This comparative statics allows us to conclude the following.

Lemma 1. *There exists a threshold $\hat{k}_1 \in (q^c, q^*)$ such that the capacity constrained firm has incentive to induce a partial cartel only if $k \geq \hat{k}_1$, otherwise it prefers to join the cartel.*

Proof. See Appendix. □

While Firm 1 can induce the formation of a partial cartel, to exist such a cartel needs to be sustainable for the unconstrained firms. There exists a critical discount factor, denoted by $\delta^{pc,intermediate}$, above which Firm 2 and 3 have no incentives to deviate from a partial cartel. The rationale is that any deviation of the constrained firm is limited by its capacity, which is lower than the competitive level in this parameter range. Moreover, the negative effect of k on unconstrained firms' profits is the same regardless of whether their strategy entails sustaining a partial cartel, deviating from it or simply competing on quantities, rendering the critical discount factor independent of k. A similar argument is present for soft capacity constraints, wherein there exists a critical discount factor, denoted by $\delta^{pc,soft}$ which is always increasing in k, and above which a partial cartel is sustainable. The intuition is that when Firm 1 has soft capacity constraints, any increase in k represents a deviation quantity that goes beyond the optimal quantity arising under competition. In turn, while colluding and deviation profits are still negatively affected by an increase in k, competition profits are independent of the capacity of the small firm. Hence, the larger k, the larger the negative externality caused to the cartel participants and the less appealing the partial cartel.

Denote $\delta^{pc} = \{\delta^{pc,intermediate}, \delta^{pc,soft}\}$ the critical discount factor above which collusion is sustainable in the respective parameter ranges. The following lemma can be presented.

Lemma 2. *The two unconstrained firms have no incentive to deviate from a partial cartel only if $\delta \geq \delta^{pc}$.*

Proof. See Appendix. □

The above lemma identifies a critical value for which a partial cartel can take place. This area is defined by both the incentive of the capacity constrained firm to induce a partial cartel and the incentive of the unconstrained firms not to deviate from the partial cartel. The latter boundary implicitly identifies a threshold of k, defined hereafter as $\hat{k}_2$, above which Firm 1's capacity is so large that Firm 2 and 3's profits become lower than under symmetric competition. This makes Firm'1 deviation not acceptable for the large firms and a partial cartel is not sustainable. We conclude the following.

Proposition 3. *Suppose that Firm 1 can credibly commit to make available its entire capacity and a full cartel is sustainable. There exists a threshold $\hat{k}_2 > \hat{k}_1$ such that, for $k \in [\hat{k}_1, \hat{k}_2]$, the small firm induces the existence of a sustainable partial cartel if and only* *if* $\delta \geq \delta_2^{intermediate}$ *if Firm 1's capacity constraint is intermediate and* $\delta \geq \max\{\delta^{pc}, \delta^*\}$ *if Firm 1's capacity constraint is soft.*

Proof. See Appendix. □

The above proposition suggests that, on the one hand, the small firm may induce a partial cartel between unconstrained firms only if its capacity is sufficiently large to get a benefit from its cheating behavior. On the other hand, unconstrained firms have never incentive to participate in a partial cartel if the colluding profits are lower than those under competition. This may happen when the capacity of the small firm is so large that the negative externality on partial colluding profits more than offsets any competitive advantage given by the sustainability of a partial conspiracy. By contrast, if the capacity constraint is sufficiently small, sustaining a partial cartel allows the unconstrained firms to increase their profits with respect to those obtained under competition. In other words, the capacity constraint is so stringent that it allows the small firm to commit to increasing the quantity by an amount that is neither too large nor too detrimental to unconstrained firms' profits. Hence, if Firm 1 can credibly commit to make available its entire capacity, there is an intermediate region of k in which it can profitably induce the unconstrained firms to accommodate such a deviation and create a partial cartel.

The region of parameters in which the small firm can profitably induce the unconstrained firms to revert from a full cartel to a partial cartel is non-monotonic in k. For k sufficiently small, being a partial cartel easier to sustain rather than a full cartel, the weakly decreasing effect of k on the critical discount factor dominates. Instead, for k sufficiently large, any increase in k lowers the difference between partial cartel and competition profits. This makes a partial cartel more difficult to sustain. Note that, in this intermediate region of k, a full cartel becomes easier to sustain rather than a partial cartel for high enough values of k. This implies that, for high enough values of k, unconstrained firms sustain a partial cartel only if the discount factor is sufficiently high, while they revert to a full cartel if the discount factor is intermediate. Intuitively, for low values of the discount factor, cooperation is not feasible.

So far, we have studied the incentive of the capacity constrained firms to induce a partial cartel when a full cartel is sustainable. Suppose now that a full cartel is not sustainable, but the small firm can still credibly commit. It follows that unconstrained firms might coordinate themselves to form a partial cartel. On the one hand, the small firm has an incentive to produce at full capacity whenever its capacity is not too small. The reason is that making available its entire capacity is always beneficial if accommodated by the unconstrained firms. On the other hand, unconstrained firms are willing to accommodate such a strategy only if k is not too large, otherwise a partial cartel is not sustainable. As there exists an area in which partial cartel is more sustainable than a full cartel, then for an intermediate region of k, a partial cartel exists if the discount factor is neither too small nor too large — i.e., $\delta \geq \delta^{pc}$, but lower than the critical value above which a full cartel is sustainable. This result is stated in the following proposition.

Proposition 4. *Suppose that Firm 1 can credibly commit to make available its entire capacity but a full cartel is not sustainable. There exists a threshold $\hat{k}_3 > \hat{k}_1$ such that, for $k \in [\hat{k}_1, \hat{k}_3]$, a partial cartel is sustainable if and only if $\delta \in [\delta^{pc}, \delta_2^{intermediate}]$ if Firm 1's capacity constraint is intermediate and $\delta \in [\delta^{pc}, \delta^*)$ if Firm 1's capacity constraint is soft.*

Proof. See Appendix. □

This proposition identifies a region of parameters in which a full participation cartel is not sustainable but it may emerge a partial cartel between unconstrained firms. Notice, that the sustainability of a partial cartel is a Pareto efficient solution for all the market players. Indeed, the discount factor is not sufficiently high to allow the existence of a full participation cartel and firms would compete if a partial cartel was not sustained. Therefore, the sustainability of a partial cartel allows to increase both Firm 1's profit — because $k \geq \hat{k}_1$ — and Firm 2 and 3's profits — because $k \leq \hat{k}_3$.

Figure 2 illustrates a bi-dimensional non-monotonicity in the cartel formation. The white area below the thick black line identifies the conditions under which either a partial or

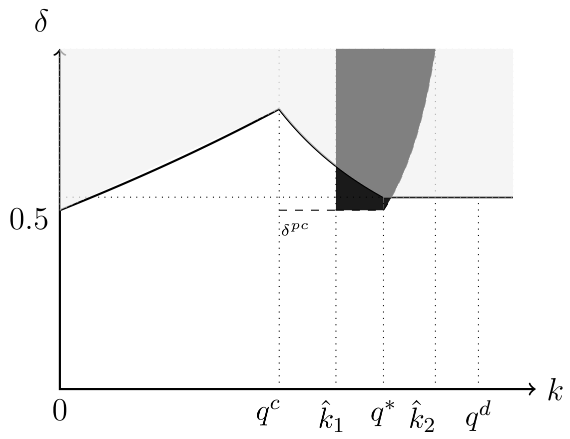

Figure 2: Existence of a full participation cartel (light gray area) and an induced partial cartel (dark gray and black areas). $(\theta = 3, \beta = 1.1, \gamma = 1)$.

a full cartel is not sustainable and, hence, competition results in equilibrium. The light gray area, instead, identifies the region of parameters in which a full participation cartel is sustainable. Finally, there exists an intermediate region of δ and k, in which collusion is sustained only between the larger firms. This area is the dark gray area when a partial cartel arises but a full participation cartel is still feasible, whereas it is black when a partial cartel arises but a full participation cartel is not sustainable.

6.2 Sustained Partial Cartel

In what follows, we study the incentives of the unconstrained firms to orchestrate a conspiracy by excluding the capacity constrained firm. This implies the inclusion of an additional stage in every period τ in which large firms decide whether (or not) to exclude the small firm from the cartel. Indeed, the inclusion of a third yet small firm in the conspiracy exerts a negative externality on cartel quantities and profits and, hence, the unconstrained firms might find it optimal to form a partial cartel while leaving the small firm to operate at full capacity. Once again, we consider both the case in which a full cartel is sustainable and when it is not.

Comparing unconstrained firms' profits under full and partial cartel, it is possible to show that unconstrained firms prefer not to engage in a full participation cartel if the capacity of Firm 1 is sufficiently small. This comparative statics leads us to the following result.

Lemma 3. *There exists a threshold $\tilde{k} \in (q^c, \hat{k}_1)$ such that the unconstrained firms have incentive to create a partial cartel only if $k \leq \tilde{k}$, otherwise they prefer to include the small firm in a full participation cartel when it is sustainable.*

The above lemma shows that a partial cartel is more profitable than a full participation cartel if the competitive pressure of Firm 1 is sufficiently small — i.e., making available its entire capacity, Firm 1 does not steal large market shares from the colluding rivals. In this case, unconstrained firms can sustain a partial cartel without considering the negative externality of their quantities on Firm 1's profit. This allows unconstrained firms to increase their joint profits more than under a full participation cartel.

Because collusive, deviation, and punishment profits have the same functional forms whether capacity constraint is stringent or intermediate, the critical discount factor coincides with the one in (A-6). Therefore, a partial cartel can also emerge when a full participation cartel is not sustainable. This is because a cartel with fewer and symmetric participants is more sustainable than a larger one. Hence, for k sufficiently small, excluding the small firm from the agreement is more profitable and facilitates its sustainability. This discussion is summarized by the following proposition.

Proposition 5. *There exists a threshold $\tilde{k} \in [q^c, q^*)$ such that, for $k \leq \tilde{k}$, unconstrained firms prefer to enforce a partial cartel if and only if $\delta \geq \delta^{pc}$.*

Figure 3 represents graphically the region of parameters in which unconstrained firms sustain a partial cartel. The dark gray area identifies the region of parameters in which a full cartel would be sustainable but a partial cartel emerges because it is even more profitable. The black area, instead, shows that collusion via a partial agreement, being easier to sustain, represents the equilibrium outcome in the repeated game.

6.3 Cartel Formation

In this section, we summarize the main results of our analysis. We can now characterize the optimal collusive schemes for unconstrained and capacity constrained firms. Assume

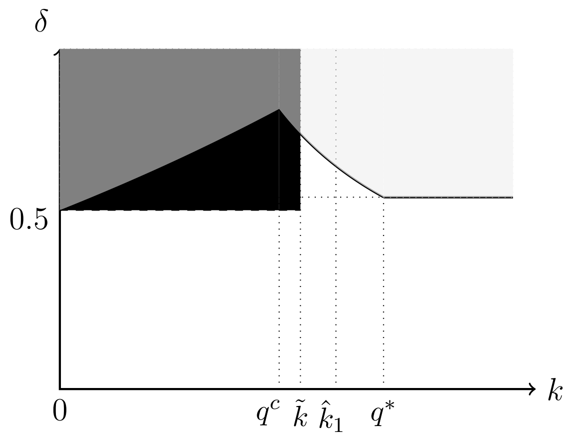

Figure 3: Existence of a full participation cartel (light gray area) and a sustained partial cartel (dark gray and black areas) ($\theta = 3, \beta = 1.1, \gamma = 1$).

that the small firm decides whether to participate in the cartel and the large firms decide whether to exclude it. Such a decision occurs in each period τ before firms set quantities. Using results in propositions 1, 3, 4 and 5, we state the following.

Proposition 6. *Suppose that Firm 1 can credibly commit to make available its entire capacity. A full participation cartel emerges, when it is feasible, if $k \in (\tilde{k}, \hat{k}_1)$ or $k > \hat{k}_3$ and a partial cartel is not sustainable. Otherwise a partial cartel arises when it is feasible.*

Figure 4 represents graphically firms' optimal strategies conditional on the discount factor in the industry and the capacity of the small firm.

Consider a sufficiently high discount factor that allows firms to collude (at least partially). When the capacity of Firm 1 is small, a partial cartel is both more profitable and easier to sustain for the unconstrained firms. Hence, they prefer not to involve the small firm in their cartel, which represents the equilibrium of the repeated game. By increasing the capacity of the small firm, an all-inclusive cartel emerges in equilibrium if the discount factor is sufficiently high. Indeed, in this parameter region, sustaining a partial cartel is easier than a full participation cartel, but less profitable for the unconstrained firms. Notably, a full participation cartel is more sustainable the larger k, up to the point in which the small firm finds it optimal to induce a partial cartel. When Firm 1's capacity is sufficiently large, the rationale for which a partial cartel emerges is starkly different and relates to the incentive the small firm has by credibly committing to a deviation quantity

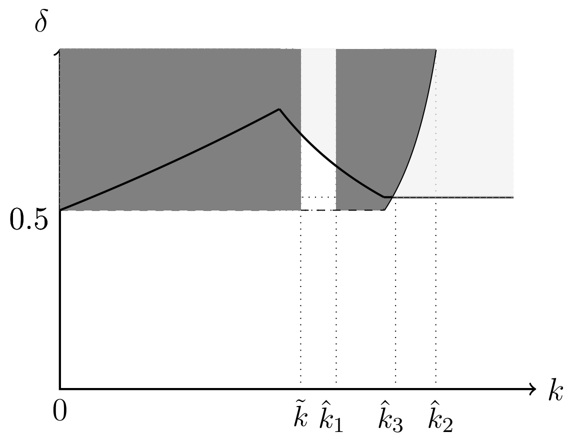

Figure 4: Cartel Formation ($\theta = 3, \beta = 1.1, \gamma = 1$).

(which coincides with its capacity). Therefore, a partial cartel is the equilibrium outcome as long as the unconstrained firms tolerate such a deviation and this happens if Firm 1's capacity is not too large. Finally, a full participation cartel is an equilibrium outcome if the discount factor is not sufficiently high or Firm 1's capacity is sufficiently large.

7 Extensions

In this section, we discuss two natural robustness issues. First, we show that the main features of our analysis developed in the baseline model extend to a framework in which firms adopt an alternative punishment code in the spirit of Abreu (1986), Abreu (1988) and Vasconcelos (2005), which might be harsher than the one considered in the baseline model. Second, we generalize our baseline model to the case in which there are $n \geq 1$ and $m \geq 1$ unconstrained and capacity constrained firms respectively. We show that our main results are with no loss of insights in a more general market structure.

Alternative Punishment Code. In our model, firms adopt a grim trigger strategy according to which any deviation from the cartel is punished by reverting to the Nash equilibrium of the stage game in all the subsequent periods.[20] Here, we relax this assumption by relying on an alternative punishment code that, under some conditions, might be

[20]It is well known that this is not an optimal punishment à la Abreu (1986) and Abreu (1988) when firms compete on quantity as the one-shot competitive profits are larger than zero.

harsher than the one considered in the baseline model. In the Appendix, we show that firms can make a full participation cartel easier to sustain with respect to Nash reversion through a more complex punishment code. Nevertheless, the non-monotonic relationship between the critical discount factor — i.e., the sustainability of a cartel — and the capacity constraint k is robust to changes in the punishment mechanism.

To this end, we adapt the punishment codes described by Abreu (1986), Abreu (1988) and Vasconcelos (2005) to our framework by constraining prices to be non-negative and by punishing more harshly those firms that have more incentive to deviate from a cooperative agreement — i.e., the unconstrained firms.[21] This penal code requires that unconstrained firms set their quantities such that their prices, and consequently, profits become zero during the punishment phase. The punishment here described renders deviation from the collusive path less appealing than simple Nash reversion if and only if $\delta \leq \frac{\pi^*}{\pi^c}$, which is derived from the simple comparison between the net present value of the different punishment profits.[22] When this condition holds, our findings in the baseline model hold qualitatively also in the case in which the penal code is harsher: (i) the unconstrained firms' critical discount factor continues to be pivotal in order to sustain collusion; (ii) the Firm 1's critical discount factor is zero for $k < q^c$ and weakly increasing otherwise; (iii) the effect of the capacity constraint k on collusion sustainability is non-monotonic. If, instead, $\delta > \frac{\pi^*}{\pi^c}$, our results continue to hold qualitatively but the penal code here described is simply more lenient than Nash reversion.

[21] Firms' asymmetry in capacity constraints lead to asymmetric profits during the punishment phase. The non-negative prices constraint prevents to nullify the capacity constrained firm's profits. All the technical details are discussed and explained in the Appendix.

[22] Note that when this condition is not satisfied, a deviation from the collusive path is more appealing than simple Nash reversion. This implies that the punishment phase in which firms earn zero profit must be longer than just one period to make this penal code harsher. However, it is important to notice that the punishment needs to be credible, meaning that firms should not have incentives to deviate from the punishment path. Specifically. as reported by Motta (2004), 'by making a stronger punishment [...], one relaxes the incentive constraint on the collusive path, but tightens the incentive constraint on the punishment path' (p. 170). This suggests that it might be better for the cartel members to resort to a milder punishment, since the harsher the punishment, the higher the incentive to deviate from the punishment path. More details are provided in the Appendix. We thank an anonymous referee for this insightful suggestion.

More than Three Firms. For ease of exposition, in our model we have considered a setting with two unconstrained firms and one constrained firms. Our insights hold qualitatively when generalizing the analysis to multiple firms. To this end, consider the case in which there are $n \geq 1$ unconstrained firms and $m \geq 1$ firms that, for simplicity, have the same capacity constraint. In the Appendix, we show that the sustainability of the cartel always depends on the incentives of the unconstrained firms, for any asymmetry in the number of firms in the market, i.e., the critical discount factor for the constrained firms is always lower than the (non-monotonic) critical discount factor of the unconstrained firms.

Specifically, in the Appendix we verify that the impact of k on the critical discount factor does not change if the number of capacity constrained firms is larger than the number of unconstrained firms. This implies that collusion is sustainable if and only if unconstrained firms (regardless of the number of firms) have no incentive to deviate, because capacity constrained firms are more eager to collude. Nevertheless, as in the baseline model, being capacity constrained implies getting lower profits than an unconstrained firm, which suggests that there is no systematic incentive to be capacity constrained.

We finally note that, when the critical discount factor is strictly positive,[23] it is increasing in the number of firms in the market. This effect is stronger in the stringent and intermediate capacity case if such an increase regards the number of unconstrained firms. The reason is the following. The entrance of an unconstrained firm leads the other firms to lose market shares both in collusion and in competition with a stronger negative effect on collusion sustainability with respect to the entrance of a capacity constrained firms which leads, instead, to a less pronounced effect on those profits.

[23] The critical discount factor for the capacity constrained firms in the stringent capacity case is always equal to zero regardless of the number of firms in the market. Clearly, the larger the number of firms, the lower the threshold identifying the collusive quantity.

8 Discussion and Conclusions

In this paper, we contribute to the intense debate in the antitrust literature on how asymmetries in the capacity constraints affect the sustainability of collusion. Motivated by the nature of competition and the structure of several markets, we investigate the incentive to collude by reducing the levels of production, when one firm is capacity constrained. We find that the incentive to run an all-inclusive cartel is non-monotonic in the capacity of the small firm, and a cartel is more sustainable for very low or sufficiently large capacity. Hence, the first takeaway of our analysis is that public and Antitrust Authorities should monitor carefully markets where firms are either highly asymmetric or almost symmetric in the capacity constraints as these may be prone to non-competitive behaviors. Our analysis also suggests that authorities should address most of their efforts to detect collusive behaviors when a tacit cartel is more difficult to sustain — i.e., the small firm's capacity constraint is intermediate — and an explicit agreement is therefore necessary.

Anti-competitive outcomes may also emerge in a large range of parameters in the form of partial cartelization between large firms. Our results identify a mechanism that hints at the *umbrella pricing* (see e.g., Bos and Harrington 2010) according to which the small firm can benefit from the cartel generated by its rivals. In our case, this situation allows the small firm to raise its profit while ensuring rivals extra-profits with respect to the competitive outcome. Moreover, the small firm can be excluded by the large firms in order to remove the negative externality that an additional cartel member would create. Such a case arises when the capacity constraint of the small firm is so stringent that producing at full capacity would not harm the cartel members. Hence, if a very small firm were included in a cartel, collusion would not be grounded on tacit coordination only but, for instance, it may turn out from the communication between all the market players.

This paper provides implications for intervention by Antitrust Authorities. As partial cartels emerge when the small firm produces at full capacity, it may be difficult to distinguish competitive conduct from a facilitating practice that confers stability and induces the creation of a partial cartel. This is particularly controversial when the capacity of the small firm is below the quantity set under competition. In that case, it would be hard to prove anti-competitive conduct and more investigations on the large firms' behaviors would be required. By contrast, for a mild capacity constraint, overproduction could hint at a strategic behavior and the presence of a partial cartel between unconstrained firms.

Importantly, while we acknowledge that in most cases small firms might not be able to induce a cartel by large firms and that large firms typically orchestrate conspiracies, we envision such a situation might still be possible in theory. It is therefore possible that the capacity constraint represents a credible commitment for the small firm to increase production at a level that is not too far from the quantity set by the cartel members. In response, the latter might prefer to tolerate and accommodate such a *deviation* rather than fighting back. This suggests that Antitrust Authorities should pay much attention to the hidden effects deriving from public announcements of small firms declaring to produce at full capacity. This strategy might be prone to a partial cartel.

Finally, we are aware that our analysis is not definitive as our model presents some limitations and some issues must be investigated in future works. To isolate any distortion in the incentives to join a cartel, we have assumed that one firm has limited capacity but firms are symmetric in other dimensions (e.g., marginal cost). In the real world, firms can be asymmetric in different dimensions at the same time. Similarly, for tractability, our analysis has been conducted using a simple model with a linear demand function. While we think that the impact of capacity constraints on the sustainability captured by our model is quite general and likely to be present with other demand functions, a more general investigation is left for future research. Moreover, we have assumed that the capacity of the small firm is exogenous and investments in capacity take several years and are sunk. Future research can be devoted to investigating the relationship between investments in capacity and cartel formation and sustainability.

References

Abada, I. and Lambin, X. (2020). Artificial intelligence: Can seemingly collusive outcomes be avoided? *Available at SSRN*.

Abreu, D. (1986). Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory*, 39(1):191–225.

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica*, pages 383–396.

Assad, S., Clark, R., Ershov, D., and Xu, L. (2020). Algorithmic pricing and competition: Empirical evidence from the German retail gasoline market. *CESifo Working Paper*.

Benjamin, R. (2016). Tacit collusion in electricity markets with uncertain demand. *Review of Industrial Organization*, 48:69–93.

Bergler, J., Heim, S., and Huschelrath, K. (2017). Strategic capacity withholding through failures in the German-Austrian electricity market. *Energy Policy*, 102:210 – 221.

Bernheim, B. D. and Whinston, M. D. (1990). Multimarket contact and collusive behavior. *The RAND Journal of Economics*, 21(1):1–26.

Bos, I., Davies, S., Harrington Jr, J. E., and Ormosi, P. L. (2018). Does enforcement deter cartels? a tale of two tails. *International Journal of Industrial Organization*, 59:372–405.

Bos, I. and Harrington, J. E. (2010). Endogenous cartel formation with heterogeneous firms. *The RAND Journal of Economics*, 41(1):92–117.

Bos, I. and Harrington, J. E. (2015). Competition policy and cartel size. *International Economic Review*, 56(1):133–153.

Bos, I. and Marini, M. A. (2019). Cartel stability under quality differentiation. *Economics Letters*, 174:70–73.

Bos, I., Marini, M. A., and Saulle, R. (2020). Cartel formation with quality differentiation. *Mathematical Social Sciences*, 106:36–50.

Brandão, A., Pinho, J., and Vasconcelos, H. (2014). Asymmetric collusion with growing demand. *Journal of Industry, Competition and Trade*, 14(4):429–472.

Brock, W. A. and Scheinkman, J. A. (1985). Price setting supergames with capacity constraints. *The Review of Economic Studies*, 52(3):371–382.

Calvano, E., Calzolari, G., Denicolò, V., Harrington, J. E., and Pastorello, S. (2020a). Protecting consumers from collusive prices due to AI. *Science*, 370(6520):1040–1042.

Calvano, E., Calzolari, G., Denicolo, V., and Pastorello, S. (2020b). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–97.

Chang, M.-H. (1991). The effects of product differentiation on collusive pricing. *International Journal of Industrial Organization*, 9(3):453 – 469.

Choné, P. and Linnemer, L. (2020). Linear demand systems for differentiated goods: Overview and user's guide. *International Journal of Industrial Organization*, 73:102663.

Compte, O., Jenny, F., and Rey, P. (2002). Capacity constraints, mergers and collusion. *European Economic Review*, 46(1):1–29.

Connor, J. M. and Bolotova, Y. (2006). Cartel overcharges: Survey and meta-analysis. *International Journal of Industrial Organization*, 24(6):1109 – 1137. The Economics of Cartels, Cartel Policy, and Collusion.

Davies, S. and Olczak, M. (2008). Tacit versus overt collusion firm asymmetries and numbers: What's the evidence? *ESRC Centre for Competition Policy Working Paper*.

De Roos, N. (2004). A model of collusion timing. *International Journal of Industrial Organization*, 22(3):351–387.

De Roos, N. and Smirnov, V. (2021). Collusion, price dispersion, and fringe competition. *European Economic Review*, 132:103640.

Dechenaux, E. and Kovenock, D. (2007). Tacit collusion and capacity withholding in repeated uniform price auctions. *The RAND Journal of Economics*, 38(4):1044–1069.

Deneckere, R. (1983). Duopoly supergames with product differentiation. *Economics Letters*, 11(1-2):37–42.

Fabra, N. (2003). Tacit collusion in repeated auctions: Uniform versus discriminatory. *The Journal of Industrial Economics*, 51(3):271–293.

Fabra, N. (2006). Collusion with capacity constraints over the business cycle. *International Journal of Industrial Organization*, 24(1):69–81.

Fabra, N. and Toro, J. (2005). Price wars and collusion in the Spanish electricity market. *International Journal of Industrial Organization*, 23(3):155 – 181.

Fogelberg, S. and Lazarczyk, E. (2014). Strategic withholding through production failures. IFN Working Paper 1015, Stockholm.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, 38(1):1–12.

Garrod, L. and Olczak, M. (2018). Explicit vs tacit collusion: The effects of firm numbers and asymmetries. *International Journal of Industrial Organization*, 56:1–25.

Harrington, J. E. (2018). Developing competition law for collusion by autonomous artificial agents. *Journal of Competition Law & Economics*, 14(3):331–363.

Harrington, J. E., Huschelrath, K., and Laitenberger, U. (2018). Rent sharing to control noncartel supply in the german cement market. *Journal of Economics & Management Strategy*, 27(1):149–166.

Hasnas, I. and Wey, C. (2015). Full versus partial collusion among brands and private label producers. *DICE Discussion Papers*.

Ivaldi, M., Jullien, B., Rey, P., Seabright, P., and Tirole, J. (2003). The economics of tacit collusion. *Report for DG Comp, European Commission.*

Khün, K.-U. (2004). The co-ordinated effects of mergers in differentiated products markets. *CEPR Discussion Papers.*

Klein, T. (2021). Autonomous algorithmic collusion: Q-learning under sequential pricing. *RAND Journal of Economics.* Forthcoming.

Kwoka, J. and Sabodash, V. (2011). Price spikes in energy markets: 'business by usual methods' or strategic withholding? *Review of Industrial Organization*, 38:285–310.

Lambertini, L. and Sasaki, D. (1999). Optimal punishments in linear duopoly supergames with product differentiation. *Journal of Economics*, 69(2):173–188.

Levenstein, M. C. and Suslow, V. Y. (2006). What determines cartel success? *Journal of Economic Literature*, 44(1):43–95.

Levitan, R. and Shubik, M. (1978). Duopoly with price and quantity as strategic variables. *International Journal of Game Theory*, 7(1):1–11.

Marshall, R. C., Marx, L., and Samkharadze, L. (2019). Monopolization conduct by cartels. *Mimeo.*

Miklos-Thal, J. (2011). Optimal collusion under cost asymmetry. *Economic Theory*, 46(1):99–125.

Motta, M. (2004). *Competition Policy: Theory and Practice.* Cambridge University Press.

OECD (2018). Algorithms and collusion: Competition policy in the digital age.

Rausser, G. and Stuermer, M. (2020). A dynamic analysis of collusive action: The case of the world copper market, 1882-2016. *MPRA Working Paper.*

Ross, T. W. (1992). Cartel stability and product differentiation. *International Journal of Industrial Organization*, 10(1):1 – 13.

Rothschild, R. (1999). Cartel stability when costs are heterogeneous. *International Journal of Industrial Organization*, 17(5):717 – 734.

Shaffer, S. (1995). Stable cartels with a Cournot fringe. *Southern Economic Journal*, 61(3):744–754.

Singh, N. and Vives, X. (1984). Price and quantity competition in a differentiated duopoly. *The RAND Journal of Economics*, 15(4):546–554.

Smuda, F. (2013). Cartel Overcharges and the Deterrent Effect of EU Competition Law. *Journal of Competition Law & Economics*, 10(1):63–86.

Song, R. and Wang, L. F. (2017). Collusion in a differentiated duopoly with network externalities. *Economics Letters*, 152:23–26.

Tyagi, R. K. (1999). On the relationship between product substitutability and tacit collusion. *Managerial and Decision Economics*, 20(6):293–298.

Vasconcelos, H. (2005). Tacit collusion, cost asymmetries, and mergers. *RAND Journal of Economics*, 36(1):39–62.

Wernerfelt, B. (1989). Tacit collusion in differentiated cournot games. *Economics Letters*, 29(4):303–306.

Wolfram, C. D. (1998). Strategic bidding in a multiunit auction: An empirical analysis of bids to supply electricity in England and Wales. *The RAND Journal of Economics*, 29(4):703–725.

Wolfram, C. D. (1999). Measuring duopoly power in the British electricity spot market. *American Economic Review*, 89(4):805–826.

Appendix

Proof of Proposition 1

In what follows, we identify the critical discount factors in the different cases.

Stringent Capacity Constraint. As discussed in the main paper, the sustainability of the cartel entirely depends on the incentive of the unconstrained firms.

In what follows we determine the deviation profit of the large firms, who maximize the following payoff function

$$\pi_2^{d,stringent}(k) = q_2^d\left(\theta - \beta q_2^d - \gamma\left(k + \frac{\theta - 2k\gamma}{2(\beta+\gamma)}\right)\right),$$

This leads to the following deviation quantity and profit

$$q_2^{d,stringent}(k) = \frac{\theta(2\beta+\gamma) - 2k\beta\gamma}{4\beta(\gamma+\beta)}, \qquad \pi_2^{d,stringent}(k) = \frac{(\theta(2\beta+\gamma) - 2k\beta\gamma)^2}{16\beta(\gamma+\beta)^2}. \tag{A-1}$$

We observe that in the relevant interval, $[0, q^c)$,

$$\frac{\partial \pi_2^{d,stringent}}{\partial k}(k) = -\frac{\gamma(\theta(2\beta+\gamma) - 2\beta\gamma k)}{4(\beta+\gamma)^2} < 0, \quad \frac{\partial \pi_2^*}{\partial k}(k) = -\frac{2\beta\gamma(\theta - \gamma k)}{(2\beta+\gamma)^2} < 0, \quad \frac{\partial \pi^c}{\partial k} = -\frac{\theta\gamma}{2(\beta+\gamma)} < 0, \tag{A-2}$$

Using results presented in the main text, a conspiracy is sustainable for the unconstrained firms if, and only if,

$$\delta \geq \frac{\pi_2^{d,stringent} - \pi_2^c}{\pi_2^{d,stringent} - \pi_2^*} = \frac{(2\beta+\gamma)^2(\theta + 2\beta k)}{8\beta(\theta\gamma + \theta\beta - k\beta\gamma) + \gamma^2(\theta - 6k\beta)} \equiv \delta_2^{stringent}(k)$$

which is increasing in k and in γ. Therefore, we can conclude that for $k \in [0, q^c)$we have

$$\left.\frac{\partial\tilde{\delta}}{\partial k}\right|_{k\in[0,q^c)} = \frac{1}{(\pi_2^d(k) - \pi_2^*(k))^2}\Bigg(\underbrace{(\pi_2^d(k) - \pi_2^c(k))\frac{\partial \pi_2^*(k)}{\partial k}}_{(-)} + \underbrace{(\pi_2^c(k) - \pi_2^*(k))\frac{\partial \pi_2^d(k)}{\partial k}}_{(-)} + \underbrace{(\pi_2^*(k) - \pi_2^d(k))\frac{\partial \pi_2^c}{\partial k}}_{(+)}\Bigg) > 0$$

and the last term dominates the sum of the two negative terms.

Intermediate Capacity Constraint. Consider first the deviation strategy of Firm 1, which consists in making available the entire capacity k. In turn, the deviation profit would be

$$\pi_1^{d,intermediate}(k) = k\left(\theta - \beta k - 2\gamma\frac{\theta}{2(\beta+2\gamma)}\right). \tag{A-3}$$

Using (4) to determine the critical discount factor, Firm 1 has no incentive to defect from the cartel if

$$\delta \geq \frac{\pi_1^{d,intermediate} - \pi^c}{\pi_1^{d,intermediate} \pi_1^*} = \frac{(2\beta+\gamma)(2k\beta-\theta)(2k(\beta+2\gamma)-\theta))}{4(2k(\beta+2\gamma)-3\theta)k\gamma^2} \equiv \delta_1^{intermediate}(k), \quad \text{(A-4)}$$

which is increasing in γ.

Consider now the incentives of the unconstrained firms. Using (4) to determine the critical discount factor, unconstrained firms have incentive to join an all-inclusive cartel only if

$$\delta \geq \frac{\pi^d - \pi^c}{\pi^d - \pi_2^*} = \frac{(2\beta+\gamma)^2\theta^2\gamma}{(\theta(4\beta^2+7\beta\gamma+\gamma^2)-2k\beta\gamma(2\gamma+\beta))(\theta(\gamma-\beta)+2k\beta(2\gamma+\beta))} \equiv \delta_2^{intermediate}(k). \quad \text{(A-5)}$$

which is decreasing in γ

Thus, for any $k \in [q^c, q^*)$, a cartel is sustainable if and only if $\delta \geq \max[\delta_1^{intermediate}, \ \delta_2^{intermediate}] = \delta_2^{intermediate}$.

Soft Capacity Constraint . In this range of parameters, the critical discount factor of Firm 1 is denoted by

$$\frac{\pi_1^{d,soft} - \pi^c}{\pi_1^{d,soft} - \pi^*} = \frac{(\beta+\gamma)^2(\theta-2\beta k)(\theta-2k(\beta+2\gamma))}{\beta\theta^2(\beta+2\gamma)+4\beta k^2(\beta+2\gamma)(\beta+\gamma)^2-4\theta k(\beta+\gamma)^3} \equiv \delta_1^{soft}.$$

which is an increasing function of γ and k and, for $k \to q^d$, we have $\delta_1^{soft} \to \delta^*$.

Proof of Lemma 1

To study the incentive of the capacity constrained firm to induce a partial cartel, we need to compare its profit when taking part in the full participation cartel, (8), and when inducing a partial cartel, (16). The proof proceeds by establishing: (i) the optimal deviation quantity for Firm 1 and the relative profit; (ii) the size of k above which inducing a partial cartel maximizes Firm 1's profit.

(i) If Firm 1 can credibly commit not to participate to a full participation cartel and to offer a deviation quantity q_1^d, Firm 2 and 3 maximize their joint profits as follows

$$\max_{q_2,q_3} q_2(\theta - \beta q_2 - \gamma q_3 - \gamma q_1^d) + q_3(\theta - \beta q_3 - \gamma q_2 - \gamma q_1^d).$$

From the first-order conditions, it is possible to derive the equilibrium quantities for Firm 2 and 3 under partial cartel, i.e.,

$$q_i^{pc} = \frac{\theta - q_1^d\gamma}{2(\beta+\gamma)}, \qquad i = 2, 3.$$

Provided the equilibrium quantities for Firm 2 and 3, Firm 1 maximizes its profit as follows

$$\max_{q_1^d} q_1^d \left(\theta - \beta q_1^d - 2\gamma \frac{\theta - q_1^d\gamma}{2(\beta+\gamma)}\right).$$

Differentiating with respect to q_1^d yields the following optimal deviation quantity for Firm 1

$$q_1^d = \frac{\theta\beta}{2(\beta^2 - \gamma^2 + \beta\gamma)},$$

which is always larger than q^d — i.e., the deviation quantity if firms are not capacity constrained. This implies that the capacity constraint of Firm 1 is binding and offering q_1^d is not feasible. As Firm 1's profit is increasing in quantity until $q = q_1^d$, the optimal deviation quantity coincides with its capacity k.

(ii) Comparing Firm 1's profit under a partial cartel between unconstrained firms — i.e., π_1^{pc} — and under a full participation cartel — i.e., π^c —, it can be shown that $\pi_1^{pc} \geq \pi^c$ if and only if the following condition holds

$$k \geq \hat{k}_1 \equiv \frac{\theta}{2\left(\beta^2 + \beta\gamma - \gamma^2\right)}\left(\beta - \gamma\sqrt{\frac{\gamma}{\beta + 2\gamma}}\right).$$

Note that such a threshold is larger q^c and lower than q^* for any $k \in (q^c, q^d)$.

Proof of Lemma 2

Consider the unconstrained firms under partial cartel. The following inequality identifies conditions under which an induced partial cartel is sustainable:

$$\frac{\pi_i^{pc}}{1-\delta} > \pi_i^{pc,d} + \pi^* \frac{\delta}{1-\delta}.$$

where $\pi_i^{pc,d}$ represents the short-run profit that Firm i, participating in the partial cartel, would obtain if deviating from the conspiracy, whereas π^* identifies the punishment payoff.

The critical discount factor in for intermediate capacity is

$$\delta \geq \frac{\pi_i^{pc,d} - \pi_i^{pc}}{\pi_i^{pc,d} - \pi_i^*} = \frac{(2\beta+\gamma)^2}{8\beta^2 + 8\beta\gamma + \gamma^2} \equiv \delta^{pc,intermediate} \tag{A-6}$$

which is independent of k.

In the presence of soft capacity constraints, the critical discount factor is as follows

$$\delta \geq \frac{\pi_i^{pc,d} - \pi_i^{pc}}{\pi_i^{pc,d} - \pi^*} = \frac{\gamma(\theta - \gamma k)^2}{(\theta - k(2\beta+\gamma))(\theta(4\beta+\gamma) - \gamma k(2\beta+\gamma))} \equiv \delta^{pc,soft}(k),$$

which is always increasing in k.

Optimal deviation under partial cartel

In what follows, we present the optimal deviation under partial cartel for an unconstrained firm (say Firm 2)[24] Given the (partially) collusive quantity q_3^{pc} set by Firm 3 and the

[24]The following results are symmetric for Firm 3.

capacity k offered entirely by Firm 1, Firm 2 maximizes the following short-run profit

$$\pi^{pc,d}(k) = q^d \left(\theta - \frac{\gamma(\theta - \gamma k)}{2\beta + 2\gamma} - \gamma k - \beta q^d\right).$$

Differentiating with respect to q^d yields the following deviation quantity

$$q^{pc,d}(k) = \frac{(2\beta + \gamma)(\theta - \gamma k)}{4\beta(\beta + \gamma)},$$

and the resulting deviation profits

$$\pi^{pc,d}(k) = \frac{(2\beta + \gamma)^2(\theta - \gamma k)^2}{16\beta(\beta + \gamma)^2}.$$

Proof of Proposition 3

In this proof, we show that a partial cartel exists in an intermediate region of k if the discount factor is sufficiently high. First, by Lemma 1 there exists a threshold of k, such that Firm 1 prefers to induce a cartel for any $k \geq \hat{k}_1$ and such a threshold belongs to the interval (q^c, q^*).

Second, we identify the critical value of δ above which a partial cartel is sustainable. Suppose $k \in (q^c, q^*]$. As $\delta_2^{intermediate} > \delta^{pc,intermediate}$, it follows that, for $k \in [\hat{k}_1, q^*]$, a partial cartel is induced by Firm 1 only if $\delta \geq \delta_2^{intermediate}$.

Suppose, instead, $k \in (q^*, q^d)$. The critical discount factor for a full participation cartel, δ^*, is constant and independent of k, while the critical discount factor for a partial cartel, $\delta^{pc,soft}$, is strictly increasing in k. Specifically, $\delta^{pc,soft}$ equals 1 if[25]

$$k = \hat{k}_2 := \frac{\theta(\beta + \gamma) - \sqrt{\beta\theta^2(\beta + \gamma)}}{\gamma(\beta + \gamma)}.$$

Hence, combining the above results, the small firm induces a partial cartel if $k \in [\hat{k}_1, \hat{k}_2]$ and δ is sufficiently high to sustain both a partial and a full participation cartel.

Proof of Proposition 4

In this proof, we show that there exists an intermediate region of k in which Firm 1 has incentive to induce a partial cartel and a full participation cartel is not sustainable. It follows immediately from Proof of Lemma 1 and Proof of Proposition 3 that Firm 1 induces a partial cartel only if $k \geq \hat{k}_1$ and a partial cartel is sustainable if and only if $\delta \geq \delta^{pc,intermediate}$ for an intermediate capacity constraint and $\delta \geq \delta^{pc,soft}$ for a soft capacity constraint. The critical discount to sustain a partial cartel is always lower than

[25] As $\delta^{pc,soft}$ is a quadratic function, there exist two solutions for $\delta^{pc,soft} = 1$. However, we rule out the negative solution of k.

the critical discount factor to sustain a full participation cartel if the capacity constraint is intermediate. Instead, when the capacity constraint is soft, $\delta^{pc,soft} \leq \delta^*$ if and only if

$$k \leq \hat{k}_3 \equiv \frac{\theta}{\gamma} - \frac{2\beta\theta^2(\beta+\gamma)^2}{\gamma\sqrt{\beta\theta^2(\beta+\gamma)^2\left(12\beta^2\gamma + 4\beta^3 + 11\beta\gamma^2 + 2\gamma^3\right)}}.$$

where $\hat{k}_3$ is larger than $\hat{k}_1$.

Proof of Proposition 5

In this proof, we show the incentive for the unconstrained firms to run a partial cartel although a full participation cartel would be sustainable.

Suppose $k \in [0, q^c)$. Comparing unconstrained firms' profits under full participation cartel in (12) and under a partial cartel in (15), one can verify that the latter are always larger than the former.

Suppose, instead, $k \in [q^c, q^*)$. Comparing unconstrained firms' profits under full participation cartel in (8) and under a partial cartel in (15), it can be shown that there exists a critical value

$$\tilde{k} \equiv \frac{\theta}{\gamma} - \sqrt{\frac{\theta^2(\beta+\gamma)}{\gamma^2(\beta+2\gamma)}},$$

which is lower than $\hat{k}_1$, such that profits under full participation cartel are larger than those under partial cartel if, and only if, $k > \tilde{k}$. The opposite holds otherwise.

Alternative Punishment Code

In this section, we present the proof of our extension on an alternative punishment system, which might be harsher than the one presented in the baseline model. Before proceeding, it is important to notice that, when the capacity constraint matters, firms' asymmetry leads to different punishment profits. Assuming that $p \geq 0$, the harshest punishment code implies setting q such that unconstrained firms — i.e., those firms have greater incentive to deviate from the collusive agreement — get zero profits. Instead, the capacity constrained firm, during the punishment phase, makes available its entire capacity, so that it earns positive profits.

Let us define the punishment profits. Suppose all firms are symmetric and the capacity constraint is never binding. They set q such that $\pi = 0$, i.e.,

$$q_i\left(\theta - \beta q_i - \gamma\sum_{j\neq i} q_j\right) = 0$$

and, by symmetry,

$$q^{hp} = \frac{\theta}{\beta + 2\gamma}$$

Consider the case in which the capacity constraint is binding — i.e., $k < q^{hp}$. The capacity constrained firm makes available its entire capacity, while unconstrained firms set q such that they get $\pi^{hp} = 0$, i.e.,

$$q_i\left(\theta - \beta q_i - \gamma q_j - \gamma k\right) = 0 \qquad i = 2, 3.$$

and, by symmetry,

$$\hat{q}^{hp}(k) = \frac{\theta - \gamma k}{\beta + \gamma}.$$

In turn, the capacity constrained firm gets

$$\pi_1^{hp}(k) = k\frac{\left(\beta - \gamma\right)\left(\theta - k\beta - 2k\gamma\right)}{\beta + \gamma},$$

which is larger than zero for any $k < q^{hp}$.

To make our point in the simplest possible way, we consider the following penal code, based on a *stick and carrot* strategy, that firms use to sustain collusion:

(i) if a firm deviates from the collusive agreement in period τ, then in period $\tau + 1$ all firms set q in order to enforce the punishment code. If all firms obey the punishment code, they return to collusion for the rest of the game.[26]

(ii) if a firm deviates during the punishment phase in period τ, in period $\tau + 1$ there is another round of punishment. If all firms obey the punishment code in $\tau + 1$, they return to collusion in period $\tau + 2$ and for the rest of the game. Otherwise, the punishment phase continues.

The proof will proceed by showing that firms have no incentive to deviate neither from the collusive path not from the punishment code.

We start studying the incentives not to deviate during the collusive phase (see (i)). If the capacity constraint is not binding, an unconstrained firm has no incentive to deviate from the agreement if and only if

$$\frac{\pi^c}{1 - \delta} \geq \pi^d + \pi^c\frac{\delta^2}{1 - \delta} \tag{A-7}$$

where π^c and π^d are the collusive and deviation profits in the main text. This implies

$$\delta \geq \delta^{dc} \equiv \frac{\gamma^2}{\beta\left(\beta + 2\gamma\right)}. \tag{A-8}$$

Instead, in the stringent capacity case, firm 1 has no incentive to deviate from the collusive path as deviation profits are never larger than collusive profits (as in the baseline model), while unconstrained firms have no incentive to deviate from the collusive path if and only if (A-7) holds, which implies

$$\delta \geq \hat{\delta}^{dc1}(k) \equiv \frac{\gamma^2(\theta + 2k\beta)^2}{4\theta\beta(\beta + \gamma)(\theta - 2k\gamma)}. \tag{A-9}$$

[26] This penal code might be harsher than Nash reversion if and only if $\frac{\delta\pi^c}{1-\delta} \leq \frac{\pi^*}{1-\delta}$, which implies that $\delta \leq \frac{\pi^*}{\pi^c}$. If this penal code is not sufficiently harsh with respect to simple Nash reversion, the punishment phase in which firms earn zero profit can be extended. Of course, there exists a finite T such that $\delta^T \leq \frac{\pi^*}{\pi^c}$. In the following of the proof, we will show the impact of making the punishment harsher on the conditions not to deviate from the punishment phase and why this is a enough to sustain collusion.

In any other interval, the unconstrained firms have no incentive to deviate from the collusive path if and only if (A-8) holds, as k is sufficiently large not to affect the collusive profits. By contrast, the capacity constrained firm has no incentive to deviate from the collusive path if and only if

$$\frac{\pi^c}{1-\delta} \geq \pi^{dc} + \delta\pi_1^{hp} + \frac{\delta^2}{1-\delta}\pi^c$$

which implies

$$\delta \geq \hat{\delta}^{dc2}(k) \equiv \frac{(\beta+\gamma)(\theta-2\beta k)(\theta k(\beta-\gamma)(\beta+2\gamma)-2k(\beta+2\gamma)-\theta)}{\theta^2(\beta+\gamma)-4k^2(\beta-\gamma)(\beta+2\gamma)^2} \tag{A-10}$$

for any $k \in [q^c, q^d]$, and

$$\delta \geq \hat{\delta}^{dc3}(k) \equiv \frac{2\gamma\theta^2(\beta+\gamma)}{(\beta+2\gamma)\left(4\theta k(\beta-\gamma)(\beta+2\gamma)-\theta^2(\beta+\gamma)-4k^2(\beta-\gamma)(\beta+2\gamma)^2\right)} \tag{A-11}$$

for $k \in (q^d, q^{hp})$. Note that the critical discount factors for the capacity constrained firm are increasing in k.

Now, let us study the the incentives not to deviate during the punishment phase (see (ii)). We start by considering the case in which the capacity constraint is not binding. The deviating firm solves the following problem

$$\max q_i \left(\theta - \beta q_i - \gamma \sum_{j\neq i} q_j^{hp}\right)$$

which leads to the following deviation quantity and profits

$$q^{dp} = \frac{\theta}{2\left(\beta+2\gamma\right)}, \qquad \pi^{dp} = \frac{\theta^2\beta}{4\left(\beta+2\gamma\right)^2}.$$

Firms have no incentive to deviate from the punishment phase if and only if

$$\frac{\delta}{1-\delta}\pi^c \geq \pi^{dp} + \frac{\delta^2}{1-\delta}\pi^c$$

which implies

$$\delta \geq \delta^{dp} \equiv \frac{\beta}{(\beta+2\gamma)}. \tag{A-12}$$

Instead, when the capacity constraint is binding, the deviating unconstrained firm solves the following problem

$$\max q_i \left(\theta - \beta q_i - \gamma k - \gamma q j^{hp}\right)$$

which leads to the following deviation quantity and profits

$$\hat{q}^{dp}(k) = \frac{(\theta-k\gamma)}{2\left(\beta+\gamma\right)}, \qquad \hat{\pi}^{dp} = \frac{\beta\left(k\gamma-\theta\right)^2}{4\left(\beta+\gamma\right)^2}.$$

Unconstrained firms have no incentive to deviate from the punishment phase if

$$\frac{\delta}{1-\delta}\pi^c \geq \hat{\pi}^{dp} + \frac{\delta^2}{1-\delta}\pi^c.$$

If the capacity constraint is stringent, then the critical discount factor is equal to

$$\delta \geq \hat{\delta}^{dp1}(k) \equiv \frac{\beta\left(k\gamma-\theta\right)^2}{\left(\beta+\gamma\right)\theta\left(\theta-2k\gamma\right)}. \tag{A-13}$$

Otherwise, the critical discount factor is equal to

$$\delta \geq \hat{\delta}^{dp2}(k) \equiv \frac{\beta\left(\beta+2\gamma\right)\left(\theta-k\gamma\right)^2}{\theta^2\left(\beta+\gamma\right)^2}. \tag{A-14}$$

Instead, the deviating capacity constrained firm sets $q = q^{dp}$ if k is sufficiently large, otherwise it produces at full capacity.[27] It is clear that, in the latter case, the capacity constrained firm has never incentive to deviate from the punishment phase as there are no gains from deviating during the punishment phase. In the former case, instead, the deviation profits for the capacity constrained firm are

$$\hat{\pi}_1^{dp}(k) = \theta\frac{8k\gamma^3+\theta\beta^2-4\theta\gamma^2+\theta\beta\gamma+4k\beta\gamma^2}{4\left(\beta+2\gamma\right)^2\left(\beta+\gamma\right)}.$$

Hence, Firm 1 has no incentive to deviate from the punishment phase if and only if

$$\pi_1^{hp} + \frac{\delta}{1-\delta}\pi^c \geq \hat{\pi}_1^{dp} + \delta\pi_1^{hp} + \frac{\delta^2}{1-\delta}\pi^c$$

namely, if

$$\delta \geq \hat{\delta}^{dp3}(k) \equiv \frac{(\theta-2k(\beta+2\gamma))(\theta(\beta^2+\beta\gamma-4\gamma^2)-2k(\beta-\gamma)(\beta+2\gamma)^2)}{\theta^2(\beta+\gamma)(\beta+2\gamma)}. \tag{A-15}$$

Finally, we compare the critical discount factors for constrained and unconstrained firms in any relevant interval of k, by considering the following results:

a) Capacity constrained firm: the critical discount factor is zero (as in the baseline model) when the capacity constraint is stringent. In any other interval, it is increasing in k and tends to the critical discount factor in the absence of capacity constraints (as in the baseline model). We also note that, for k sufficiently small, when firm 1 has no incentive to deviate during the collusive phase, the same holds during the punishment phase. The opposite happens for k sufficiently large. Specifically, there exists a threshold $k = \tilde{k}^{hp} \in (q^c, q^d)$, denoted by

$$\tilde{k}^{hp} \equiv \frac{\theta\left(\beta^2+2\beta\gamma-\gamma^2\right)}{2(\beta+2\gamma)\left(\beta^2+\beta\gamma-\gamma^2\right)}$$

such that (A-10) is larger than (A-15) if and only if $k \leq \tilde{k}^{hp}$.[28]

[27] Note that q^{dp} coincides to the collusive quantity when the capacity constraint is not binding.
[28] It is intuitive to prove that (A-11) is always lower than (A-15) in the relevant parameters' range.

b) Unconstrained firm: the critical discount factor for the punishment phase is always larger than the critical discount factor for the collusive phase in any interval of k.

It is easy to show that the critical discount factor for the unconstrained firms is always larger than the critical discount factor for the capacity constrained firm (as in the baseline model). Moreover, looking at the effect of k on the relevant critical discount factor, we can easily verify that $\hat{\delta}^{dp1}$ is increasing in k, while $\hat{\delta}^{dp2}$ is decreasing in k and δ^{dp} is independent of k. This confirms our results described in Proposition 1.

Longer punishment phase. In this subsection, we study how increasing the length of the punishment phase, in which deviating firms earn zero profits, affects the incentive not to deviate from the punishment phase. We show that increasing the periods of punishment does not make collusion easier for the cartel members, for any k.

To this end, we modify the structure of the punishment code as follows: if a firm deviates from the collusive agreement in period τ, then all firms enforce a punishment in periods $\tau + 1$ and $\tau + 2$. If all firms obey the punishment code, they return to collusion for the rest of the game.

Specifically, the unconstrained firms have no incentive to deviate from the punishment phase if and only if

$$\frac{\delta^2}{1-\delta}\pi^c \geq \pi^{dp} + \frac{\delta^2}{1-\delta}\pi^c,$$

or, alternatively, solving for the discount factor, if

$$\delta \geq \sqrt{\frac{\beta}{(\beta + 2\gamma)}}.$$

if the capacity constraint is not binding.

If the capacity constraint is binding, the unconstrained firm have no incentive to deviate from the punishment phase if and only if

$$\frac{\delta^2}{1-\delta}\pi^c \geq \hat{\pi}^{dp} + \frac{\delta^2}{1-\delta}\pi^c.$$

If the capacity constraint is stringent, there is no incentive to deviate if

$$\delta \geq (\theta - k\gamma)\sqrt{\frac{\beta}{(\beta+\gamma)\,\theta\,(\theta - 2k\gamma)}}.$$

Otherwise, in all other cases, there is no incentive to deviate if

$$\delta \geq \frac{(\theta - k\gamma)}{\theta\,(\beta+\gamma)}\sqrt{\beta\,(\beta + 2\gamma)}.$$

We note that the conditions identified above are more stringent than those derived for the penal code in which a deviating firm is punished with only one period of zero profits. Since the unconstrained firms are pivotal in order to sustain a collusive agreement and being the critical discount factor for the punishment phase always larger than the critical discount factor for the collusive phase in any interval of k, a *stick and carrot* strategy with only one period of punishment might be enough to sustain collusion through a penal code. This implies that, in our setting, increasing the length of the punishment does not make it easier for firms to engage in collusive behaviors.

Multiple Firms

Consider an industry with $n \geq 1$ unconstrained firms and $m \geq 1$ firms with a limited capacity $k > 0$, which is assumed to be the same for the sake of simplicity. In the following, we show that the results of the baseline model hold with a more general setup. Note that this extension is equivalent to the baseline setting if $n = 2$ and $m = 1$.

We start by defining firms' profits in the stage game, that will characterize the punishment phase.[29] If the capacity of the m firms is not binding, each firm solves the following maximization problem

$$\max_{q_i} q_i \left(\theta - \beta q_i - (n + m - 1)\gamma q_j\right)$$

which implies

$$q^* = \frac{\theta}{2\beta + \gamma(m + n - 1)} \qquad \pi^* = \frac{\beta\theta^2}{\left(2\beta + \gamma(m + n - 1)\right)^2}$$

Instead, if the capacity constraint is binding — i.e., m firms produce at full capacity — the unconstrained firms solves the following maximization problem

$$\max_{q_i} q_i \left(\theta - \beta q_i - (n - 1)\gamma q_j - m\gamma k\right)$$

which implies

$$q_n^* = \frac{\theta - km\gamma}{2\beta + (n - 1)\gamma} \quad \pi_n^* = \frac{\beta(\theta - km\gamma)^2}{\left(2\beta + (n - 1)\gamma\right)^2} \quad \pi_m^* = \left(\theta - \beta k - n\gamma\frac{\theta - km\gamma}{2\beta + (n - 1)\gamma} - (m - 1)\gamma k\right)$$

where the subscripts "n" and "m" denote the individual outcomes of unconstrained and capacity constrained firms respectively.

Now, let us analyze the repeated game when the capacity constraint is slack. Firms maximize their joint profits as follows

$$\max_{q_i} \sum_{i=1,\dots,n,1,\dots,m} q_i \left(\theta - \beta q_i - \gamma \sum_{j \neq i} q_j\right)$$

which leads to the following collusive outcomes

$$q^c = \frac{\theta}{2\beta + 2\gamma(n + m - 1)} \qquad \pi^c = \frac{\theta^2}{4\left(\beta + \gamma(n + m - 1)\right)}.$$

A deviating firm solves the following maximization problem

$$\max_{q^d} q^d \left(\theta - \beta q^d - \gamma(n + m - 1) q^c\right)$$

[29] We analyze the case in which firms adopt a grim trigger strategy but, as shown in the previous section, all results are robust to an alternative punishment code.

which leads to the following deviation quantity and profits

$$q^d = \frac{\theta\left(2\beta + \gamma\left(n+m-1\right)\right)}{4\beta\left(\beta + \gamma\left(n+m-1\right)\right)} \qquad \pi^d = \frac{\theta^2\left(2\beta + \gamma\left(n+m-1\right)\right)^2}{16\beta\left(\beta + \gamma\left(n+m-1\right)\right)^2}.$$

Hence, in the absence of capacity constraints, collusion is sustainable if and only if

$$\delta \geq \frac{\left(2\beta + \gamma\left(n+m-1\right)\right)^2}{m^2\gamma^2 + 2mn\gamma^2 + 8m\beta\gamma - 2m\gamma^2 + n^2\gamma^2 + 8n\beta\gamma - 2n\gamma^2 + 8\beta^2 - 8\beta\gamma + \gamma^2}. \quad \text{(A-16)}$$

Analogously to the baseline model, let us analyze the sustainability of a full cartel for the stringent, intermediate and soft capacity case. Clearly, the thresholds identifying the three cases are different with respect to the baseline model if we consider both $n \neq 2$ and $m \neq 1$.

In the stringent capacity case, m firms might collude by producing at full capacity, while unconstrained firms might collude by solving the following maximization problem

$$\max_{q_i}\left(\sum_{i=1,\ldots,n} q_i\left(\theta - \beta q_i - \gamma\left(n-1\right)q_j - \gamma mk\right) + \sum_{i=1,\ldots,m} k\left(\theta - \beta k - \gamma q_i - \gamma\left(n-1\right)q_j - \gamma\left(m-1\right)k\right)\right)$$

which leads to

$$q_n^c = \frac{\theta - 2km\gamma}{2\left(\beta + \left(n-1\right)\gamma\right)} \quad \pi_n^c = \frac{\theta\left(\theta - 2km\gamma\right)}{4\left(\beta + \left(n-1\right)\gamma\right)} \quad \pi_m^c = k\left(\theta - \beta k - n\gamma\frac{\theta - 2km\gamma}{2\left(\beta + \left(n-1\right)\gamma\right)} - \gamma\left(m-1\right)k\right)$$

With the same logic of the baseline model, the capacity constrained firms have no incentive to deviate from the cartel. Instead, an unconstrained firm might deviate by setting a quantity that solves the following maximization problem

$$\max_{q^d} q^d\left(\theta - \beta q^d - \gamma\left(n-1\right)\frac{\theta - 2km\gamma}{2\left(\beta + \left(n-1\right)\gamma\right)} - \gamma mk\right)$$

which leads to

$$q^d = \frac{2\theta\beta + \left(n-1\right)\theta\gamma - 2km\beta\gamma}{4\beta\left(\beta + \left(n-1\right)\gamma\right)} \qquad \pi^d = \frac{\left(2\theta\beta + \left(n-1\right)\theta\gamma - 2km\beta\gamma\right)^2}{16\beta\left(\beta + \left(n-1\right)\gamma\right)^2}.$$

Using (4) to determine the critical discount factor, collusion is sustainable if and only if

$$\delta \geq \delta_n^{stringent} \equiv \frac{\left(\theta - n\theta - 2km\beta\right)\left(2\beta - \gamma + n\gamma\right)^2}{\left(n-1\right)\left(-8\theta\beta^2 - \theta\gamma^2 - n^2\theta\gamma^2 + 8\theta\beta\gamma + 2n\theta\gamma^2 - 8n\theta\beta\gamma - 6km\beta\gamma^2 + 8km\beta^2\gamma + 6kmn\beta\gamma^2\right)} \quad \text{(A-17)}$$

which is increasing in k, n and m. Note that the marginal increase of $\delta_n^{stringent}$ deriving from an increase in n is larger than that deriving from an increase in m.

Finally, it is worth noting that, if $n = 1$ collusion is never sustainable. Indeed, an unconstrained firms get the same profits from deviating from the cartel and competing with m firms, each producing a limited capacity k (lower than q^c).

In the intermediate capacity case, all firms set the same quantity under collusion, while they differ in their incentives to deviate. A deviating capacity constrained firm gets

$$\pi_m^d = \frac{1}{2} \frac{k\left(\left(m+n-1\right)\theta\gamma + 2\theta\beta - 2k\beta^2 - \left(m+n-1\right)2k\beta\gamma\right)}{\beta + \left(m+n-1\right)\gamma}$$

while a deviating unconstrained firm gets π^d. Using (4), we derive the critical discount factors for the capacity constrained firms

$$\delta \geq \delta_m^{intermediate} \equiv \frac{\pi_m^d - \pi^c}{\pi_m^d - \pi_m^*} \tag{A-18}$$

and for the unconstrained firms

$$\delta \geq \delta_n^{intermediate} \equiv \frac{\pi^d - \pi^c}{\pi^d - \pi_n^*}. \tag{A-19}$$

By comparing the above critical discount factors, it is easy to prove that, in the relevant space of parameters, (A-18), which is increasing in k, is always lower than (A-19), which is instead decreasing in k.[30]

Moreover, it is worth noting that the marginal increases of $\delta_n^{intermediate}$ and $\delta_m^{intermediate}$ deriving from an increase in n are larger than those deriving from an increase in m.

Finally, let us consider the soft capacity case. The unconstrained firms have no incentive to deviate if (A-16) holds. Instead, the capacity constrained firms have no incentive to deviate if and only if

$$\delta \geq \delta_m^{soft} \equiv \frac{\pi_m^d - \pi^c}{\pi_m^d - \pi^*} \tag{A-20}$$

which is increasing in k and lower than (A-16). As in the baseline model, as k tends to q^d, the above critical discount factor tends to (A-16).

[30] The extended equations are quite complex but available upon request.

Figure 1 in the presence of perfect substitutability ($\beta = \gamma = 1.1$)

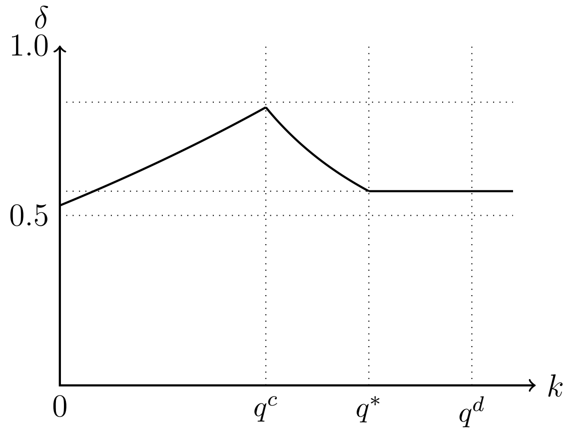

Figure 5: Critical discount factors for the unconstrained firms as a function of k ($\theta = 3, \beta = 1.1, \gamma = 1.1$).