MPRA

Munich Personal RePEc Archive

Collusion, Firm Numbers and Asymmetries Revisited

Luke Garrod and Matthew Olczak

School of Business and Economics, Loughborough University, Aston Business School, Aston University

30 September 2016

Collusion, Firm Numbers and Asymmetries Revisited[*]

Luke Garrod[†] and Matthew Olczak[‡]

September 30, 2016

Abstract

Despite the fact that competition law prohibits explicit cartels but not tacit collusion, theories of collusion often do not distinguish between the two. In this paper, we address this issue and ask: under which types of market structures are cartels likely to arise when firms can alternatively collude tacitly? To answer this question, we analyse an infinitely repeated game where firms with (possibly asymmetric) capacity constraints can make secret price cuts. Tacit collusion can involve price wars on the equilibrium path. Explicit collusion involves firms secretly sharing their private information in an illegal cartel to avoid such price wars. However, this runs the risk of sanctions. We find that, in contrast to the conventional wisdom but consistent with the available empirical evidence, cartels are least likely to arise in markets with a few symmetric firms, because tacit collusion is relatively more appealing in such markets. We discuss the implications for anti-cartel enforcement policy.

JEL classification: D43, D82, K21, L44

Key words: cartel, tacit collusion, imperfect monitoring, capacity constraints

[†]School of Business and Economics, Loughborough University, LE11 3TU, UK, email: l.garrod@lboro.ac.uk

[‡]Aston Business School, Aston University, Birmingham, B4 7ET, UK, email: m.olczak@aston.ac.uk

1   Introduction

Under competition law, there is an important distinction between explicit and tacit collusion. Explicit collusion is where a group of firms directly communicate with each other, usually with the intention of coordinating and/or monitoring their actions to raise profits above competitive levels. This is prohibited in the US by the Sherman Act and in the European Union by Article 101 of the Treaty on the Functioning of the European Union. Consequently, detected cartels usually face hefty sanctions to punish and deter such conduct, and undetected cartels are incentivised to reveal themselves through leniency programmes that grant immunity to the first cartel member to turn themselves in. In contrast, tacit collusion is where firms collude without such explicit communication, and this is not usually considered illegal. Therefore, firms guilty of tacit collusion should not face any penalties, despite their conduct causing similar effects as explicit collusion. This approach to the law is arguably desirable because it ensures legal certainty over prohibited conduct, and it prevents competitive behaviour from being punished erroneously (see Motta, 2004, p.185-190).

Despite its importance in terms of the law, the role of communication has often been absent in theories of collusion (see Harrington, 2008). This has meant that such theories have been presumed to be applicable to both forms of collusion, with the same models often used to derive predictions in applied work that relate specifically to tacit collusion (see Ivaldi *et al.*, 2003) or to explicit collusion (see Grout and Sonderegger, 2005). While this approach has undoubtedly played a key role in developing an understanding of the circumstances in which tacit or explicit collusion is likely to arise, it has the potential to lead to incorrect predictions, especially those related to explicit collusion. The reason is that "parties might be more likely to engage in overtly collusive practices specifically in those circumstances that are predicted by the theory as being adverse to collusion" because "the need for cartel members to communicate intensifies precisely when collusion is harder to sustain" (Grout and Sonderegger, 2005, p.36).

Our aim in this paper is to develop a simple framework that captures the benefits and costs of explicit collusion compared with tacit collusion so we can revisit the question: under which types of market structures are cartels most likely to arise? There is a need to revisit this question due to the mounting evidence that some cartels do not seem to arise in markets that standard theory predicts. For instance, a well-established result of collusion theory is that when there are a large number of firms or asymmetries among them, it is more difficult to prevent some members from deviating from the collusive agreement. Consequently, the conventional wisdom is that cartels

are most likely to arise in markets with a few firms, often considered to be no more than three or four, that are relatively symmetric. However, there is a long established literature into the features of cartels prosecuted in the US which has uncovered some puzzling results regarding the number of firms in each cartel (see Levenstein and Suslow, 2006, for a review). In particular, across a number of different samples from throughout the 20th century the median number of cartel members varies between six and ten, and there is even evidence that cartels are more likely to be sustained for longer periods of time when the cartel has more members (see Posner, 1970; Hay and Kelley, 1974; Fraas and Greer, 1977). More recently, Davies and Olczak (2008) analysed the market shares of cartels prosecuted by the European Commission between 1990 and 2006. They found that the medium number of firms was five and that the extent of the asymmetries among the cartel members were so large that it "calls into question whether symmetry of market shares is a pervasive feature of real world cartels" (p.198).

One possible explanation of this puzzling evidence, first raised by Posner (1970), is that it could be a result of a sample selection bias: cartels with few relatively symmetric firms may be harder to detect and are consequently under-represented in any sample of prosecuted cartels. However, evidence that is inconsistent with the conventional wisdom is still observed for samples of legal cartels where no such selection bias should exist (see Dick, 1996; and Symeonidis, 2003). An alternative explanation raised by Hay and Kelly (1974, p.14) is that: "firms in highly concentrated industries often do not need to collude explicitly, but can rely on tacit collusion or 'conscious paralellism' to achieve a near monopoly price". This is the conjecture that we are interested in exploring in this paper, and it is interesting to note that the (admittedly limited) evidence on tacit collusion does seem to be more consistent with the predictions of standard theory. For example, Davies et al. (2011) show that the European Commission's interventions in mergers due to the increased likelihood of tacit collusion have almost always been confined to cases where there would have been only two relatively symmetric players post-merger. Similarly, numerous laboratory experiments have shown that collusion without communication is unsuccessful at raising prices above competitive prices if there are more than two firms or if there are asymmetries among the firms (see for example, Fonseca and Normann, 2008 and 2012).

We explore this issue by analysing an infinitely repeated game with asymmetrically capacity constrained firms that have the potential to make secret price cuts. We draw on earlier work (Garrod and Olczak, 2016), where we analysed tacit collusion in setting, similar to that first discussed by Stigler (1964), where market demand is unobservable and fluctuates over time, and firms never directly observe their rivals' prices or sales. In such a setting, the firms may need to

initiate costly price wars on the equilibrium path when at least one firm receives sufficiently low sales, similar to other models of imperfect monitoring (see Green and Porter, 1984; and Tirole, 1988). In this paper, we build on our previous analysis by also modelling explicit collusion where each firm can secretly share its private information with their rivals to improve their ability to monitoring each other's actions. Consequently, explicit collusion can raise profits relative to tacit collusion by avoiding price wars, but this is illegal and runs the risk of sanctions. Thus, this tradeoff is likely to capture the incentives of cartels that are set up with the main purpose of monitoring each others sales.[1]

We analyse how this tradeoff changes as asymmetries in capacities between the firms become greater. Consistent with the previous literature, we find that asymmetries in capacities hinder both forms of collusion. In particular, both explicit and tacit collusion are more difficult to sustain as the largest firm gets larger, and tacit collusion can be more difficult to sustain as the smallest firm gets smaller. However, we also show that when firms can alternatively collude tacitly the marginal effect of communication on firm profits is smallest in a symmetric duopoly and it increases as the smallest firm gets smaller. The reason is that the profits from tacit collusion are lower as the smallest firm gets smaller. Consequently, in contrast to the conventional wisdom yet consistent with the empirical evidence, firms have less incentive to form a cartel in markets with a few symmetric firms, because tacit collusion is relatively more appealing in such markets.

Our results have two main policy implications. First, while it is known that anti-cartel enforcement is most effective when leniency programmes are complemented by other more proactive methods of cartel detection (see Harrington and Chang, 2015), our results question the effectiveness of structural screens. These are used in practice to identify industry characteristics associated with collusion to search proactively for as yet undetected cartels (see for example Petit, 2012). Our results suggest that if such devices are constructed according to the conventional wisdom, then they have the potential to overlook markets where cartels are most likely to occur and where cartels cause the greatest marginal detriment to consumer surplus. Second, our results question whether it is appropriate to base anti-cartel enforcement fines on the revenue of the cartel members, as is currently the case in Europe and the US (see Bageri *et al.*, 2013). For instance, we show that, holding the total revenue of the cartel constant, a given fine is least effective at deterring cartels when the marginal effect of communication on firm profit is greatest (or analogously in our model when the marginal detriment to consumer surplus is greatest).

---

[1]For example, see the discussion by Harrington (2006) of the following cartels: carbonless paper, choline chloride, copper plumbing tubes, graphite electrodes, plasterboard, vitamins, and zinc phosphate.

Instead, a fine regime that is based on damages would be more effective, as this would allow cartels that cause the greatest marginal detriment to consumer surplus to receive higher fines.

Our paper is related to a small number of papers and working papers that model the difference between explicit and tacit collusion (see Athey and Bagwell, 2001, 2008; Martin, 2006; Mouraviev, 2013; Awaya and Krishna, 2014; and Spector, 2015). While the underlying competition game and the information exchanged among the firms varies from model to model, the general approach of this literature is to analyse how exchanging private information facilitates collusion by improving the ability of firms to monitor each other's actions. This is also our approach, but the novelty of our model is that it is the only one in which firms can be *ex ante* asymmetric. Consequently, our model is ideal to consider how the effects of communication on collusion vary with firm numbers and asymmetries, which has not been adequately addressed by the previous literature. Furthermore, the focus of much of the previous literature has been on the role of communication in facilitating collusion when the information is non-verifiable, which follows research into this question that solely focuses on explicit collusion (see for example Harrington and Skrzypacz, 2011).[2] While we assume in the main body of paper that sales are verifiable for simplicity, we show in a robustness section that firms cannot gain under explicit collusion by undercutting the collusive price and submitting false sales reports when such information is non-verifiable.

Finally, our paper is also related to the theoretical literature that analyses collusion when firms are asymmetrically capacity constrained (see Compte *et al.* 2002, Vasconcelos, 2005; and Bos and Harrington, 2010 and 2015). In particular, our framework is an extension of Compte *et al.* (2002) to a setting of imperfect monitoring and, consistent with their results, we find that explicit collusion is facilitated as the number of symmetric firms increase, holding the total capacity constant.[3] While this contrasts with the conventional wisdom, it is not the sole driver of our results regarding firm numbers, because we also show that there can be a greater incentive for explicit collusion relative to tacit collusion as the number of symmetric firms rises. This is due

---

[2]There are two exceptions to this. Martin (2006) analyses the difference between explicit and tacit collusion when firms exchange verifiable sales information in a Cournot model with symmetric firms in a framework similar to Green and Porter (1984). His main result is that higher fines for collusion can reduce the profit of explicit collusion compared with tacit, provided the probability that tacit collusion is erroneously prosecuted is sufficiently low. In a similar setting to that just described, but without the chance of prosecution for tacit collusion, a recent working paper by Mouraviev (2013) investigates how frequently firms can exchange verifiable information with each other when such conduct causes a positive probability of detection. His simulations show that firms can meet more or less often as the number of symmetric firms rises.

[3]This result is highlighted by Kühn (2012) in a more general Bertrand-Edgeworth framework with perfect observability.

to the industry profits from tacit collusion falling as a result of longer and more frequent price wars on the equilibrium path. In contrast, we also find that dividing a given amount of capacity between more symmetric firms can have an ambiguous effect on tacit collusion. Thus, we can get the result that as the number of firms increase, explicit collusion is easier to sustain but tacit collusion is more difficult to sustain and less profitable. This appears to be more consistent with the evidence on explicit collusion and on tacit collusion (discussed above).

The rest of the paper is organised as follows. Section 2 sets out the assumptions of the model and solves for the static Nash equilibrium. In section 3, we analyse the two forms of collusion separately. In section 4, we analyse firms' incentives to form a cartel when firms can alternatively collude tacitly and analyse how such incentives changes as the capacity distribution changes. Section 5 discusses the implications for anti-cartel enforcement and Section 6 considers the case of non-verifiable sales information. Section 7 concludes. All proofs are relegated to appendix A. In appendix B and C, we analyse two further robustness checks that are not central to our main story. These appendices are best read after section 3.

2　The Model

To model the difference between explicit and tacit collusion for asymmetric firms, we extend the capacity-constrained private monitoring repeated game in Garrod and Olczak (2016). We now allow firms to form a cartel to exchange their private information at the risk of being detected and penalised by a competition agency.

2.1　Basic assumptions

Consider a market in which a fixed number of $n \geq 2$ capacity-constrained firms compete on price to supply a homogeneous product over an infinite number of periods. Firm $i = \{1, \ldots, n\}$ can produce a unit of the product at a constant marginal cost but the maximum it can produce in any period is k_i. We denote the total industry capacity as $K \equiv \sum_i k_i$, the sum of firm i's rivals' capacities as $K_{-i} \equiv \sum_{j \neq i} k_j$, and we let $k_n \geq k_{n-1} \geq \ldots \geq k_1 > 0$, without loss of generality. In any period t, firms set prices simultaneously, where $\mathbf{p}_t = \{p_{it}, \mathbf{p}_{-it}\}$ is the vector of prices, p_{it} is the price of firm i and $\mathbf{p}_{-it}$ is the vector of prices of all of firm i's rivals. Firms have a common discount factor, $\delta \in (0, 1)$, and we normalised their marginal costs to zero.

Market demand consists of a mass of m_t (infintesimally small) buyers, each of whom are willing to buy one unit provided the price does not exceed their reservation price, which we

normalise to 1. We assume that firms do not observe m_τ, for all $\tau \in \{0, \ldots, t\}$, but they know that m_t is independently drawn from a distribution $G(m)$, with mean $\widehat{m}$ and density $g(m) > 0$ on the interval $[\underline{m}, \overline{m}]$ where $G(\underline{m}) = 0$ and $G(\overline{m}) = 1$. Furthermore, firm i never observes firm j's prices, $p_{j\tau}$, or sales, $s_{j\tau}$, $j \neq i$, for all $\tau \in \{0, \ldots, t\}$. In contrast, buyers are informed of prices, so they will want to buy from the cheapest firm. Thus, this setting is consistent with a market in which all buyers are willing to check the prices of every firm in each period to find discounts from posted prices, but actual transaction prices are never public information.

2.2   Demand allocation and sales

Consistent with much of the previous literature, we assume demand is allocated according to the following rule:

The proportional allocation rule
Unsupplied buyers want to buy from the firm(s) with the lowest price among those with spare capacity.

- If the joint capacity of such firms is insufficient to supply all of the unsupplied buyers, then such capacity is exhausted, and the remaining unsupplied buyers now want to purchase from the firm(s) with the next lowest price among those with spare capacity, and so on.

- If the joint capacity of such firms suffices to supply all of the unsupplied buyers, then each firm supplies an amount of buyers equal to its proportion of the joint capacity.

Following Garrod and Olczak (2016), we also place the following plausible yet potentially restrictive assumption on the capacity distribution:

Assumption 1. $K_{-1} \leq \underline{m}$.

This says that the joint capacity of the smallest firm's rivals should not exceed the minimum market demand. This is a necessary condition that ensures firm i's sales in period t are strictly positive, for all i and all $m_t > \underline{m}$, even if it is the highest-priced firm. An implication of Assumption 1 is that if $\underline{m} < K$, then there is a restriction on the size of the smallest firm in that it cannot be too small.

Before moving on to the implications of these two assumptions for firm sales, we should take a moment to defend them. First, the proportional allocation rule is not crucial to our main story, because we can generate similar results to those presented below when demand is instead

allocated evenly between the firms when they set a common price. We prefer the proportional allocation rule because it is consistent with the main literature, it is more tractable than this alternative, and there is also anecdotal evidence of cartels allocated the demand in this manner (see the examples in Vasconcelos, 2005, and Bos and Harrington, 2010). Second, Assumption 1 is imposed as it substantially simplifies the mixed strategy Nash equilibrium analysis when there are more than two firms with asymmetric capacities. However, our main results can be generated without Assumption 1 when there are two firms with asymmetries or when there are more than two symmetric firms. Moreover, we believe that Assumption 1 is not very restrictive for the capacity distributions where tacit and explicit collusion are likely to be substitutes. The reason is that the evidence on tacit collusion indicates that it is most likely to occur in markets with two or three relatively symmetric firms, so the smallest firm is likely to be relatively large. Finally, given the smallest firm's capacity can be no larger than for a symmetric duopoly, a necessary condition for Assumption 1 to hold is that the minimum market demand must be greater than 50% of the total capacity, $\underline{m} \geq 0.5K$. We place no restriction on the level of the maximum market demand, $\overline{m}$.

Denoting $\Omega(p_{it})$ as the set of firms that price strictly below p_{it} and $p_t^{\max} \equiv \max\{\mathbf{p}_t\}$, Assumption 1 and the proportional allocation rule together imply that firm i's sales in period t, $s_{it}(p_{it}, \mathbf{p}_{-it}; m_t)$, for any $p_{it} \leq 1$, are:

$$
s_{it}\left(p_{it}, \mathbf{p}_{-it}; m_t\right) = \begin{cases} k_i & \text{if } p_{it} < p_t^{\max} \\ \min\left\{\frac{k_i}{K - \sum_{j \in \Omega(p_{it})} k_j}\left(m_t - \sum_{j \in \Omega(p_{it})} k_j\right), k_i\right\} \geq 0 & \text{if } p_{it} = p_t^{\max}. \end{cases} \tag{1}
$$

This says that a firm will supply its proportion of the residual demand if it is the highest-priced firm in the market and if capacity is not exhausted, otherwise it will supply its full capacity. This implies that firm i's expected per-period profit is $\pi_{it}\left(p_{it}, \mathbf{p}_{-it}\right) = p_{it} \int_{\underline{m}}^{\overline{m}} s_{it}\left(p_{it}, \mathbf{p}_{-it}; m\right) g(m)dm$, where we drop time subscripts if there is no ambiguity. We write $\pi_i\left(p\right) = k_i p S\left(p\right)$ if $p_j = p$ for all j, where $S\left(p\right)$ is the expect sales per unit of capacity, such that:

$$
S\left(p\right) = \begin{cases} 1 & \text{if } \quad K \leq \underline{m} \\ \int_{\underline{m}}^{K} \frac{m}{K} g(m)dm + \int_{K}^{\overline{m}} g(m)dm & \text{if } \quad \underline{m} < K < \overline{m} \\ \frac{\widehat{m}}{K} & \text{if } \quad \overline{m} \leq K. \end{cases} \tag{2}
$$

So, such profits are maximised for $p^m \equiv 1$.

2.3 Static Nash equilibrium

Lemma 1 states the static Nash equilibrium profits, which can result from pure or mixed strategies. An important part of the analysis is firm i's minimax payoff, which is:

$$\underline{\pi}_i \equiv \begin{cases} \widehat{m} - K_{-i} & \text{if} \quad \overline{m} \leq K, \\ \int_{\underline{m}}^{K} (m - K_{-i}) \, g(m) dm + k_i \int_{K}^{\overline{m}} g(m) dm & \text{if} \quad \underline{m} < K < \overline{m} \\ k_i & \text{if} \quad K \leq \underline{m} < \overline{m} \end{cases} \tag{3}$$

for all i. The intuition is that if the realisation of market demand is below total capacity, then a firm that sets the monopoly price expects to supply the residual demand, otherwise it expects to supply its full capacity. The proof is the same as Lemma 1 in Garrod and Olczak (2016), so we only provide the intuition below.

Lemma 1. *For any given $n \geq 2$ and $K_{-1} \leq \underline{m}$:*
i) if $\underline{m} \geq K$, the unique pure strategy Nash equilibrium profits are $\pi_i^N = k_i \; \forall \; i$;
ii) if $\underline{m} < K$, the mixed strategy Nash equilibrium profits are $\pi_i^N (k_n) = k_i \frac{\pi_n}{k_n} \; \forall \; i$.

Competition is not effective if the minimum market demand is above total capacity, $\underline{m} \geq K$, so firms set $p_i = 1$ in equilibrium and receive $\pi_i^N = k_i$ for all i. In contrast, if market demand can be below total capacity, firms are not guaranteed to supply their full capacity for every level of demand, so they have incentives to undercut each other. However, by charging $p_i = 1$, firm i can obtain its minimax payoff, $\underline{\pi}_i$. Assumption 1 is sufficient to ensure that such profits are nonnegative for all i, so competition does not imply price equals marginal cost. Instead, the largest firm will never set a price below $\underline{p} \equiv \underline{\pi}_n / k_n$ in an attempt to be the lowest-priced firm. This implies that the smaller firms $i < n$ can sell their full capacity with certainty by charging a price slightly below $\underline{p}$ to obtain a profit of $k_i \underline{p} \geq \underline{\pi}_i$. Consequently, the mixed strategy Nash equilibrium profits are given by $\pi_i^N (k_n) = k_i \underline{p}$. The lower bound of the support is $\underline{p}$, where it follows from (3) that $\lim_{\underline{m} \to K} \underline{p} = 1$.

3 Two Forms of Collusion

We now move on to analyse the repeated game. In any period, firm i's prices and sales are initially private information to it, but it may have an incentive to share this information with its rivals to facilitate collusion. In this section, we first set out the assumptions regarding the exchange of this information. Then we solve for the collusive equilibria when firms exchange their private

information with each other, and we refer to this as collusion under explicit monitoring. We then briefly restate the collusive equilibria in the absence of this information exchange, which was the focus of Garrod and Olczak (2016), and we refer to this as collusion under tacit monitoring. Henceforth, we impose $\underline{m} < K$, as collusion is unnecessary otherwise from Lemma 1.

We use the term cartel to refer to a group of firms that exchange their private information with each other. We say that a cartel is active in period t, if at the start of the period there is a chance that the firms will exchange their private information. Otherwise, the cartel is inactive. An active cartel is subject to enforcement. We assume that in each period an active cartel may be detected by the authorities with some probability, $\theta \in (0, 1]$. If such a cartel is detected, then firm i is fined $k_i F$ for all i, where $F \geq 0$ is the fine per unit of the industry's capacity. This implies that larger firms receive a larger proportion of the total cartel fine, given by KF. Given each firm's collusive sales are also in proportion to its capacity, this assumption is consistent with most jurisdictions, including Europe and the US, where the fine for each cartel member is initially linked to the size of its sales (see International Competition Network, 2008). In addition, we also allow for the possibility that each firm can inform the authorities of the cartel in return for leniency, in which case the cartel is detected with certainty but the informant is not fined. Consistent with leniency programs in Europe and in the US, we assume that applying for leniency is publicly observable.

The timing of the game in any given period $t > 1$ is as follows:

Stage 0 (pricing stage). The firms set prices simultaneously. The game continues to stage 1.

Stage 1 (communication stage). The firms realise their sales and profits privately:

- If there is not an active cartel, then period t ends and period $t + 1$ begins.

- If there is an active cartel, then each firm chooses whether to share its private information with its rivals secretly and whether to inform the authorities of the cartel publicly in return for leniency. After this, enforcement is realised:

    - If no firm has informed the authorities of the cartel, then the cartel is detected and convicted with a probability $\theta \in (0, 1]$, and all firms are each fined $k_i F$. Otherwise, the cartel is not detected and no firm is fined.

    - If at least one firm has informed the authorities of the cartel, then the cartel is detected and convicted with probability 1. Leniency is given to only one informant and the competition agency selects the informant with the lowest price (or randomly

selects among these informants with equal probability if there is more than one). This selected informant is not fined and all other firms are each fined $k_i F$.

Finally, period t ends and period $t + 1$ begins.

There are three assumptions that are worth discussing here. First, we initially assume that the information that firms exchange is verifiable for simplicity. This implies that firms do not need to question whether the reported information is accurate or not, as is the case in Harrington and Skrypacz (2011). However, we show that our results are robust to when these sales reports are non-verifiable in section 6. Second, we focus on full leniency in the main analysis, where an informant is not fined at all. This is consistent with the current leniency programs that operate in Europe and in the US. However, leniency programs have evolved over time, especially in Europe, so we analyse the effects of partial leniency on the firms' incentives to form a cartel in Appendix B to understand the effects of this evolution. Third, we assume that a cartel does not become inactive if it is detected and convicted by the competition agency. This seems a natural assumption to us, because there is widespread evidence of recidivism among detected cartels (see Connor, 2010) and it is consistent with the approach of a number of papers on leniency (see Motta and Polo, 2003; Spagnolo, 2005; and Chen and Rey, 2013).

3.1 Explicit monitoring

In this subsection, we analyse collusion under explicit monitoring where there is an active cartel in period 0. We consider the following strategy profile which we refer to as explicit monitoring strategies. There are 'active' phases where there is an active cartel, and 'inactive' phases where there is not. In the pricing stage of a period during an active phase, each firm sets the collusive price p^m. Then in the communication stage, sales are realised and each firm secretly shares its private information with its rivals and does not apply for leniency. The active phase continues into period $t+1$ if $p_{jt} = p^m$ and $s_{jt} = k_i S(p^m)$ for all j and if all firms did not apply for leniency. Otherwise, firms enter an inactive phase. Once in the inactive phase, each firm prices according to the static Nash equilibrium forever and never exchanges its private information.

There are three comments to make regarding this strategy profile. First, reversion to the static Nash equilibrium is the harshest possible punishment under our assumptions. The reason is that, as showed by Lambson (1994), the harshest punishments under the proportional allocation rule are such that the largest firm receives the stream of profits from its minimax strategy.

11

In our setting, the per-period minimax payoff of the largest firm is equivalent to its static Nash equilibrium profits, so it is not possible to implement a harsher punishment given the proportional allocation rule. Second, the fact that firms set the monopoly price is without loss of generality, because it is easy to check that if firms set a lower price, then the equilibrium profits are lower and the critical discount factor is higher. Third, the fact that the cartel may still be active after detection and conviction implies that, under certain conditions, the cartel will want to exploit the leniency programme by applying for leniency in every period in an attempt to reduce its expected fines. In Appendix C we show that $\theta < \frac{1}{2}$ is a sufficient condition to ensure that cartel members cannot profit by exploiting the leniency programme. Given the low detection rates of cartels, this seems likely to hold in most jurisdictions, hence the reason why we restrict attention to explicit monitoring strategies in the main analysis.[4]

We now solve for the equilibria in explicit monitoring strategies. Given firms can observe whether or not they share information or apply for leniency, the game is one of observable actions. Thus, it follows from the one-stage deviation principle (see Fudenberg and Tirole, 1991, p.108-110) that the profile of explicit monitoring strategies is a subgame perfect Nash equilibrium (SPNE) if there is no history that leads to a subgame in which a deviant will chose an action that differs to that prescribed by the strategy, then conforms to the strategy thereafter (assuming the deviant believes others will also conform to the strategy). We say that collusion under explicit monitoring is not sustainable if no such equilibrium strategies exist.

Denoting firm i's expected (normalised) profit in an active phase as $k_i V^e$, if all firms follow explicit monitoring strategies, then $k_i V^e = (1 - \delta) (\pi_i (p^m) - \theta k_i F) + \delta k_i V^e$, where solving yields:

$$k_i V^e = \pi_i (p^m) - \theta k_i F. \tag{4}$$

This says that in a period during an active cartel phase firm i expected profits are the expected per-period profit from setting p^m minus its expected fine. We must find the conditions under which no firm will deviate from its prescribed strategy. Firms play the static Nash equilibrium during each period of an inactive phase, so it is clear that they have no incentive to deviate in any such periods. Consequently, we need only consider deviations during active phases. We begin by considering deviations in the communication stage of a given period t, and then we

---

[4]An important determinant for the analysis in Appendix C is which informant is selected for leniency when there is more than one. We have assumed that the deviant with the lowest price is given leniency, which is consistent with Spagnolo (2005), who assumes that a deviant informant will be given leniency over a colluding informant. This assumption has little effect on the main analysis, however, because firms do not apply for leniency on the equilibrium path, so there is only ever one deviating informant in this case, who will have the lowest price.

move back to the pricing stage. Recall that during an active phase, firm i believes that its rivals will set p^m in the pricing stage, and that they will share their private information and not apply for leniency in the communication stage.

Communication stage

First suppose that all firms abided by their strategies in the pricing stage of period t by setting $p_j = p^m$ for all j and consider a deviation in the communication stage. Firm i can deviate at this point by applying for leniency and/or by not sharing its private information. If firm i applies for leniency, then firms will enter an inactive phase and firm i's expected (normalised) discounted profit from this stage on is $\delta \pi_i^N (k_n)$, regardless of whether firm i shares its information or not. Consequently, this is firm i's optimal deviation at this stage, because if it deviates by not sharing its private information without applying for leniency, then firm i's profits would be lower due to the risk of being fined. Thus, firm i has no incentive to deviate in the communication stage if:

$$k_i V^e \geq (1 - \delta) \pi_i (p^m) + \delta \pi_i^N (k_n). \tag{5}$$

This says that, provided all firms have abided by their strategies up to this stage, firm i will not deviate in the communication stage if it cannot gain by applying for leniency, which eliminates its fine but makes the cartel inactive thereafter. We refer to this as the "communication" incentive compatibility constraint (ICC).

   Next suppose that only firm i had deviated from its prescribed strategy in stage 0 of period t by setting $p_i \neq p^m$. This implies that firms will enter a punishment phase regardless of whether firm i's shares its private information or not. Thus, firm i has a dominant strategy to apply for leniency to raise its expected profits by eliminating its fine.

Pricing Stage

Now consider firm i's incentive to deviate from its prescribed strategy in the pricing stage of period t by setting $p_i \neq p^m$. It follows from the above that if the communication ICC is satisfied, then firm i has no incentive to deviate in the pricing stage if:

$$k_i V^e \geq (1 - \delta) k_i + \delta \pi_i^N (k_n). \tag{6}$$

This says that firm i will not deviate in the pricing stage if it cannot gain by marginally undercutting p^m in the pricing stage to supply its full capacity, k_i, and applying for leniency in

the communication stage, which makes the cartel inactive thereafter. We refer to this as the "pricing" ICC.

Note that it is easy to check using (2) that the right-hand side of (6) is strictly greater than the right-hand side of the above for all $K > \underline{m}$, and the left-hand sides of the two ICCs are the same. Thus, the pricing ICC is more stringent than the communication ICC, so it is the one that binds. Rearranging (6) in terms of the discount factor yields:

$$\delta \geq \frac{1 - V^e}{1 - \underline{p}} \equiv \delta_e^* \left(k_n, F \right), \tag{7}$$

where the fact that the right-hand side of (7) is independent of k_i implies that if the pricing ICC holds for firm i, then it holds for all of its rivals $j \neq i$. This implies that, despite potential asymmetries, each firm has the same incentive to deviate as its rivals.

Thus, it follows from the above that if firms are sufficiently patient, such that $\delta \geq \delta_e^* \left(k_n, F \right)$, then firm i's optimal equilibrium profits are given by (4) for all i. Finally, it is easy to check that there exists a unique fine per unit of capacity, $\overline{F} \left(k_n \right) \equiv \frac{1}{\theta} \left(S \left(p^m \right) - \underline{p} \right) > 0$, that is the level of F that sets the optimal equilibrium profits equal to the static Nash equilibrium profits, $k_i V_e^* = \pi_i^N \left(k_n \right)$, such that if the fine per unit of capacity is sufficiently low, $F \in \left[0, \overline{F} \left(k_n \right) \right)$, then $\delta_e^* \left(k_n, F \right) < 1$ and $k_i V_e^* > \pi_i^N \left(k_n \right)$. The above analysis is summarised more formally by the following proposition.

Proposition 1. *For any given $n \geq 2$, $K_{-1} \leq \underline{m} < K$ and $\overline{m} \geq \underline{m}$, if the fine per unit of capacity is low, such that $F \in \left[0, \overline{F} \left(k_n \right) \right)$, and if firms are sufficiently patient, such that $\delta \geq \delta_e^* \left(k_n, F \right) \in \left[\frac{k_n}{K}, 1 \right)$, then firm i's optimal equilibrium profits under explicit monitoring strategies are:*

$$k_i V_e^* = \pi_i \left(p^m \right) - \theta k_i F \in \left(\pi_i^N \left(k_n \right), \pi_i \left(p^m \right) \right] \quad \forall \ i. \tag{8}$$

Otherwise, collusion under explicit monitoring is not sustainable.

3.2   Tacit monitoring

In this subsection, we analyse tacit collusion where there is never an active cartel in any period. Below we present a simplified version of the analysis in Garrod and Olczak (2016), and refer the reader to our other paper for more details.

In the absence of communication, firms can make inferences about their rivals' actions using information from their privately observed sales. In particular, all firms can always infer when at

least one firm's sales are below some firm-specific "trigger level". These trigger levels are given by $s_i^* \equiv \min\left\{\frac{k_i}{K}m^*(k_1, \overline{m}), k_i\right\}$ for all i, where $m^*(k_1, \overline{m}) \equiv \frac{K(\overline{m}-k_1)}{K-1}$ such that, from (1), they are determined by the largest possible sales firms $i > 1$ can make if all such firms set the same price and firm 1 undercuts to sell its full capacity. Thus, firms can always infer when at least one firm's sales are below its trigger level for the following reasons. First, if all firms set a common price, then all firms' sales will exceed their respective trigger levels when the realisation of market demand is high, otherwise they can all fall below the trigger levels. Second, if all firms do not set a common price, then it follows from (1) that the highest-priced firms receive sales below their trigger levels and the lower-priced firm(s) can infer this from the fact that they supply their full capacity. Consequently, in any given period it is common knowledge among the firms whether all firms' sales are greater than their trigger levels or not, and this information over time constitutes a public history that firms can condition their play on. Thus, we restrict attention to equilibria in public strategies, known as perfect public equilibria (PPE).

This information can be sufficient to ensure that tacit monitoring is perfect in the sense that a deviation will be detected with certainty. This is the case if fluctuations in market demand are small, because then firms will only ever receive sales below their trigger levels if they are undercut. Formally, the condition for which tacit monitoring is perfect is:

$$\overline{m} < k_1\left(\frac{K - \underline{m}}{K}\right) + \underline{m} \equiv \underline{x}(k_1),$$

such that each firm i's minimum realised sales if all firms set a common price, $k_i\frac{\underline{m}}{K}$, exceed its trigger level, s_i^*. However, if fluctuations in market demand are larger than this, such that $\overline{m} \geq \underline{x}(k_1)$, then tacit monitoring is imperfect, because collusive sales will also fall below the trigger levels when the realisation of market demand is low. Thus, in this case, colluding firms face a non-trivial signal extraction problem: each firm does not know whether the realisation of market demand was unluckily low or whether at least one rival has undercut them. Consequently, similar to Green and Porter (1984), punishment phases must occur on the equilibrium path to provide firms with the correct incentives to collude.

We want to find the optimal PPE profits. In Garrod and Olczak (2016), we used the techniques of Abreu *et al.* (1986, 1990) to show that the maximal PPE payoffs under imperfect monitoring can be generated from a simple strategy profile, which henceforth we refer to as tacit monitoring strategies. We also showed that this strategy profile generates the optimal equilibrium under perfect monitoring if the punishment duration lasts an infinite number of periods. Under tacit monitoring strategies, there can be 'collusive phases' and 'punishment phases'. Sup-

pose period t is in a collusive phase. In any such period, a firm should set the monopoly price, p^m. If firms receive a good signal in that all firms' sales are above their trigger levels, then the collusive phase continues into the next period $t+1$. If firms receive a bad signal in that at least one firm's sales are below its trigger level, then firms enter a punishment phase in the next period $t+1$. In the punishment phase, each firm should play the static Nash equilibrium for T periods, after which a new collusive phase begins. This sequence repeats in any future collusive phase.

Denoting firm i's expected (normalised) profit in a collusive phase as $k_i V^c$ and its expected (normalised) profit at the start of a punishment phase as $k_i V^p$, if all firms follow tacit monitoring strategies, then:

$$k_i V^c = (1 - \delta) \pi_i (p^m) + \delta \left[(1 - G (m^* (k_1, \overline{m}))) k_i V^c + G (m^* (k_1, \overline{m})) k_i V^p \right]$$
$$k_i V^p = (1 - \delta) \sum_{t=0}^{T-1} \delta^t \pi_i^N (k_n) + \delta^T k_i V^c,$$

for all i, where $G (m^* (k_1, \overline{m}))$ is the probability that at least one firm's sales are below their trigger level when firms set a common price. Substituting $k_i V^p$ into $k_i V^c$ and solving yields:

$$k_i V^c = \pi_i^N (k_n) + \frac{(1 - \delta)}{1 - \delta + G (m^* (k_1, \overline{m})) \delta (1 - \delta^T)} \left(\pi_i (p^m) - \pi_i^N (k_n) \right), \qquad (9)$$

where it is then straightforward to check that $\pi_i (p^m) \geq k_i V^c > k_i V^p$ for any positive punishment phase duration $T > 0$ and that $k_i V^p > \pi_i^N (k_n)$ for any $T < \infty$.

The profile of tacit monitoring strategies is a PPE if, for each date t and any history h^t, the strategies yield a Nash equilibrium from that date on. We say that collusion under tacit monitoring is not sustainable if no such equilibrium strategies exist. It is clear that firms have no incentive to deviate during the punishment phase, so we need only consider deviations during collusive phases. The ICC for firm i is as follows:

$$k_i V^c \geq (1 - \delta) k_i + \delta k_i V^p, \ \forall \ i. \qquad (10)$$

This says that firm i will not deviate in a collusive phase if it cannot gain by marginally undercutting p^m to supply its full capacity k_i in which case firms will enter the punishment phase with certainty. This ICC is never satisfied if the maximum market demand is greater than total capacity, $\overline{m} \geq K$, because then firms will receive a bad signal with certainty, since $G (m^* (k_1, \overline{m})) = 1$ for all $\overline{m} \geq K$, from $m^* (k_1, \overline{m}) \geq \overline{m}$.

Thus, focussing on the case where $\overline{m} < K$, substituting $k_i V^p$ and $k_i V^c$ into (10), then rearranging yields:

$$(1 - G (m^* (k_1, \overline{m}))) \frac{K}{k_n} - \frac{1}{\delta} \geq \delta^T \left[(1 - G (m^* (k_1, \overline{m}))) \frac{K}{k_n} - 1 \right]. \qquad (11)$$

16

This ICC is independent of k_i which, similar to explicit monitoring, implies that if the ICC holds for firm i, then it also holds for all other firms $j \neq i$. Furthermore, (11) also implies that there are three necessary conditions for the profile of tacit monitoring strategies to be a PPE. First, the length of the punishment phase must be sufficiently long, where the critical duration is implicitly defined by the level of T where (11) holds with equality. Second, firms must also be sufficiently patient, such that:

$$\delta \geq \frac{1}{(1 - G(m^*(k_1, \overline{m})))} \frac{k_n}{K} \equiv \delta_c^*(k_1, k_n), \tag{12}$$

in which case the left-hand side of (11) is positive, such that the ICC holds as $T \to \infty$. Third, the probability of receiving a bad signal must be sufficiently low, where:

$$G(m^*(k_1, \overline{m})) < \frac{K_{-n}}{K}, \tag{13}$$

such that the expression in square brackets in (11) is positive. This also ensures that the critical discount factor $\delta_c^*(k_1, k_n) < 1$. The inequality in (13) holds if fluctuations in market demand are sufficiently small such that $\overline{m} < \overline{x}(k_1, k_n)$, where $\overline{x}(k_1, k_n)$ denotes the level of $\overline{m}$ that solves (13) with equality. Furthermore, since the right-hand side of (13) is strictly between zero and 1, it follows that $\underline{x}(k_1) < \overline{x}(k_1, k_n) < K$.

Finally, if tacit monitoring is perfect, where $\overline{m} < \underline{x}(k_1)$ such that $G(m^*(k_1, \overline{m})) = 0$, then it follows from (9) that the optimal equilibrium profits are $k_i V_c^* = \pi_i(p^m)$ for all i. The reason is that punishment phases do not occur on the equilibrium path, so firms are able share the monopoly profits between them, if the threat of reverting to the static Nash equilibrium forever is sufficiently harsh, such that $\delta \geq \frac{k_n}{K} \equiv \underline{\delta}_c^*(k_n)$. In contrast, if tacit monitoring is imperfect, such that $\overline{m} \geq \underline{x}(k_1)$ and $G(m^*(k_1, \overline{m})) > 0$, then the optimal equilibrium payoffs can be found by evaluating $k_i V^c$ at level of T where (11) holds with equality. Thus, rearranging (11) in terms of δ^T, then substituting into (9) we can obtain the optimal equilibrium profits:

$$k_i V^c = \frac{k_i}{K} \left(\frac{\widehat{m} - G(m^*(k_1, \overline{m})) K}{1 - G(m^*(k_1, \overline{m}))} \right) \equiv k_i V_c^*,$$

for all i. The sum of the optimal equilibrium profits is below the monopoly profit for any $\overline{m} \geq \underline{x}(k_1)$, due to the fact that punishment phases occur on the equilibrium path, but such profits are above the static Nash equilibrium profits for any $\overline{m} < \overline{x}(k_1, k_n)$.

The above analysis is more formally summarised by the following proposition.

Proposition 2. *For any given $n \geq 2$ and $K_{-1} \leq \underline{m} < K$:*

i) if fluctuations in market demand are small, such that $\overline{m} < \underline{x}(k_1)$, then tacit monitoring is perfect, and if firms are sufficiently patient, such that $\delta \geq \frac{k_n}{K}$, then firm i's optimal equilibrium profits are $k_i V_c^* = \pi_i(p^m) \; \forall \; i$;

ii) if fluctuations in market demand are in an intermediate range, such that $\underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n)$, then tacit monitoring is imperfect, and if firms are sufficiently patient, such that $\delta \geq \delta_c^*(k_1, k_n) = \frac{1}{1 - G(m^*(k_1, \overline{m}))} \frac{k_n}{K} \in \left(\frac{k_n}{K}, 1 \right)$, then firm i's optimal equilibrium profits are $k_i V_c^* = \frac{k_i}{K} \left(\frac{\widehat{m} - G(m^*(k_1, \overline{m})) K}{1 - G(m^*(k_1, \overline{m}))} \right) \in \left(\pi_i^N(k_n), \pi_i(p^m) \right) \forall \; i$;

iii) otherwise, collusion under tacit monitoring is not sustainable.

3.3   Comparative statics

We want to analyse the effects of asymmetries on the incentive to form a cartel when firms could alternatively collude tacitly. Before doing so, it is helpful to get a clear understanding of how the capacity distribution affects both forms of collusion in isolation. In this subsection, we discuss the effects of changes to the capacity distribution under the assumption that the total capacity and the number of firms constant are held constant. This implies that any increase in the capacity of a given firm will require capacity to be reallocated from a rival. Thus, when the capacity of firm j changes by a small amount, other things equal, the capacities of the other firms change to the extent that $\frac{\partial k_i}{\partial k_j} \in [-1, 0]$ for all $i \neq j$, where $\sum_{i \neq j} \frac{\partial k_i}{\partial k_j} = -1$. However, in what follows we restrict the discussion to capacity reallocations that directly affect the equilibrium analysis, and this is the case for changes to the capacity of the smallest firm or the largest firm.

We begin by analysing sustainability of collusion.

Proposition 3. *For any given $n \geq 2$ and $K_{-1} \leq \underline{m} < K$:*

i) the critical discount factor under explicit monitoring, $\delta_e^*(k_n, F)$, is strictly increasing in the capacity of the largest firm, k_n;

ii) the critical discount factor under tacit monitoring, $\delta_c^*(k_1, k_n)$, is strictly increasing in the capacity of the largest firm, k_n, and, if $\underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n)$, is strictly decreasing in the capacity of the smallest firm, k_1.

Increasing the capacity of the largest firm raises the static Nash equilibrium profits, because competition is less intense than before. Consequently, under explicit monitoring the punishment is weaker than before and the pricing ICC is tighter, so there is a rise in the critical discount

factor of explicit monitoring, $\delta_e^*(k_n, F)$. Similarly, under tacit monitoring a punishment phase that lasts an infinite number of periods is weaker than before, which increases the critical discount factor under tacit monitoring, $\delta_c^*(k_1, k_n)$. In contrast, increasing the capacity of the smallest firm does not affect the sustainability of collusion under explicit monitoring but it can make collusion under tacit monitoring easier to sustain. For instance, when the smallest firm is larger, it is less likely that firms will receive sales below their trigger levels when they set a common price. An implication of this is that if tacit monitoring is imperfect, such that $\overline{m} \in [\underline{x}(k_1), \overline{x}(k_1, k_n))$, then it is less likely that a collusive phase will switch to a punishment phase on the equilibrium path than before, and the expected future profits from collusion are higher. As a result, a punishment phase that lasts an infinite number of periods is relatively harsher, so the critical discount factor falls. A further implication of this is that the range of fluctuations in market demand for which monitoring is perfect is larger, since $\underline{x}(k_1)$ is strictly increases with the capacity of the smallest firm, k_1.

Next, we turn our attention to the profitability of collusion. It is easy to see that the capacity distribution does not affect the optimal equilibrium profits when firms set the monopoly price in every period. This is always the case for collusion under explicit monitoring and it is the case for collusion under tacit monitoring when such monitoring is perfect, such that $\overline{m} < \underline{x}(k_1)$. Thus, Proposition 4 restricts attention to when tacit monitoring is imperfect.

Proposition 4. *For any given $n \geq 2$ and $K_{-1} \leq \underline{m} < \underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n) < K$, the optimal equilibrium profits under tacit monitoring, V_c^*, are strictly increasing in the capacity of the smallest firm, k_1.*

When tacit monitoring is imperfect, such that $\underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n)$, increasing the capacity of the smallest firm makes it is less likely that firms will receive sales below their trigger levels when they set a common price, as described above. Consequently, a collusive phase is less likely to switch to a punishment phase than before, so profits rise on the equilibrium path, other things equal. This increase in profits also introduces slack into the ICC for tacit monitoring, so the optimal punishment phase duration shortens to the extent that the ICC binds with no slack. Both effects imply that firms expect there to be fewer and shorter punishment phases on the equilibrium path when the smallest firm has more capacity, so collusion under tacit monitoring is more profitable. In contrast, increasing the capacity of the largest firm causes two equal offsetting effects on the optimal equilibrium profits under tacit monitoring. The first effect is that the static Nash equilibrium profits rise, which increases the profits under tacit monitoring

on the equilibrium path, other things equal. The second effect results from the fact that the ICC under tacit monitoring is now tighter than before, so the optimal punishment phase duration lengthens to ensure that the ICC is binding with no slack. This second effect cancels out the first such that the profits under tacit monitoring are independent of the capacity of the largest firm.

The above implies that symmetry is ideal for both forms of collusion. This is due to the fact that when firms are symmetric, such that $k_i = K/n$ for all i, the largest firm is as small as possible and the smallest firm is as large as possible. Thus, it follows from Proposition 3 that the critical discount factors of both forms of collusion are as low as possible and from Proposition 4 that the profits under tacit monitoring are as high as possible. In contrast, increasing the number of symmetric firms can have different effects on the two forms of collusion. For instance, it follows from Proposition 3 that increasing the number of symmetric firms actually facilitates collusion under explicit monitoring because, holding the industry capacity constant, the 'largest' firm will then be smaller. This is consistent with the results of Compte *et al.* (2002), yet it contrasts with the conventional wisdom as highlighted by Kühn (2012). However, the effect of increasing the number of symmetric firms on collusion under tacit monitoring is more consistent with the conventional wisdom when such monitoring is imperfect. In this case, it follows from Proposition 4 that the optimal equilibrium profits under tacit monitoring fall as the number of symmetric firms rises because, holding the industry capacity constant, the 'smallest' firm will then be smaller. Furthermore, the effect on the stability of collusion under tacit monitoring is ambiguous, because dividing a fixed industry capacity between more symmetric firms would require the capacity of the largest firm and the smallest firm to decrease, which has countervailing effects from Proposition 3. Moreover, assuming demand is drawn from a uniform distribution, then the critical discount factor under tacit monitoring is $\delta_c^* (k_1, k_n) = \left(\frac{\overline{m} - \underline{m}}{K - \overline{m}} \right) \left(1 - \frac{1}{n} \right)$ if $\overline{m} > \underline{x} (k_1) = \left(\frac{K - \underline{m}}{n} \right) + \underline{m}$, such that, consistent with the conventional wisdom, it is strictly increasing in the number of symmetric firms, n.

4   Cartels, Firm Numbers and Asymmetries

In the previous section, we showed that symmetry is ideal for both forms of collusion. In this section, we now use our equilibrium analysis to investigate the effects of asymmetries on the incentives of firms to form a cartel when firms can alternatively collude tacitly. In what follows, we say that it is privately optimal for a firm to be a cartel member if its profits from being a

member are *strictly* greater than its profits from not being a member. Furthermore, we define a cartel phase as a sequence of periods that begins the period after detection and ends the period of the next detection. Thus, the expected duration of a cartel phase is $\sum_{t=1}^{\infty} t (1-\theta)^{t-1} = 1/\theta$, which implies that the lower the probability of detection, the longer the expected duration of a cartel phase. Before focussing on the comparison between explicit and tacit monitoring, we first consider the incentives of firms to form a cartel when the alternative to explicit monitoring is the static Nash equilibrium, and use this as a benchmark for our analysis.

For the benchmark case, it follows from Proposition 1 that joining a cartel is privately optimal for each firm i if collusion under explicit monitoring is sustainable, such that $\delta \geq \delta_e^* (k_n, F)$ for any $F \in \left[0, \overline{F} (k_n)\right)$. This is due to the fact that $k_i V_e^* > \pi_i^N (k_n)$ for such conditions, since rewriting this inequality yields:

$$F < \frac{1}{\theta} \left(S \left(p^m \right) - \underline{p} \right), \tag{14}$$

where the right-hand side is equivalent to $\overline{F} (k_n)$. Note that we can interpret the left-hand side of (14) as, from the cartel's perspective, the expected cost of a cartel phase per unit of capacity. Moreover, the right-hand side is the expected benefit of a cartel phase per unit of capacity, which is the multiplication of the expected duration of a cartel phase, $1/\theta$, and the per-period difference between the profits per unit of capacity of explicit monitoring and of the static Nash equilibrium, $S \left(p^m \right) - \underline{p}$. Thus, the above implies that not only is collusion under explicit monitoring more difficult to sustain in our framework as the largest firm gets larger (as we demonstrated in section 3.3), but the incentive to join a cartel is reduced, because expected benefit of a cartel is lower as the largest firm gets larger, because the static Nash equilibrium profits are higher.

4.1   Explicit versus tacit monitoring

We now investigate the incentives to form a cartel when firms could alternatively collude tacitly. Henceforth we restrict attention to $\overline{m} < \overline{x} (k_1, k_n)$, because collusion under tacit monitoring is not sustainable otherwise. For the case where $\overline{m} \geq \overline{x} (k_1, k_n)$, the appropriate comparison is that of the benchmark case discussed above.

Proposition 5 shows when collusion under explicit monitoring is more profitable than collusion under tacit monitoring, and vice versa.

21

Proposition 5. *For any given $n \geq 2$ and $K_{-1} \leq \underline{m} \leq \overline{m} < \overline{x}(k_1, k_n)$, there exists a unique fine per unit of capacity given by:*

$$F^*(k_1) \equiv \begin{cases} 0 & if \quad \overline{m} < \underline{x}(k_1) \\ \frac{1}{\theta}\left(\frac{\widehat{m}}{K} - V_c^*\right) \in \left(0, \overline{F}(k_n)\right) & if \quad \underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n), \end{cases}$$

such that if and only if $F \in [0, F^*(k_1))$, then the optimal profits that could be sustained as an equilibrium under explicit monitoring strategies are greater than under tacit monitoring strategies, $V_e^* > V_c^*$.

Collusion under explicit monitoring is more profitable than collusion under tacit monitoring, $V_e^* > V_c^*$, if the fine per unit of capacity is sufficiently low, such that:

$$F < \frac{1}{\theta}\left(S(p^m) - V_c^*\right), \tag{15}$$

where the right-hand side is equivalent to $F^*(k_1)$. Similar to (14), the above inequality compares, from the cartel's perspective, the expected cost of a cartel phase per unit of capacity on the left-hand side against the expected benefit per unit of capacity on the right-hand side. The only difference compared with (14) is that the term in brackets on the right-hand side of (15) is the per-period difference between the profits per unit of capacity of explicit monitoring and of tacit monitoring, where $S(p^m) = \frac{\widehat{m}}{K}$ for all $\overline{m} < K$ from (2). If tacit monitoring is perfect, such that $\overline{m} < \underline{x}(k_1)$, then being a cartel member is never privately optimal for any firm. The reason is that firms are able to extract the monopoly profit through tacit monitoring, so $V_c^* = \frac{\widehat{m}}{K}$ such that the expected benefit of a cartel phase equals zero. If tacit monitoring is imperfect, such that $\overline{m} \in [\underline{x}(k_1), \overline{x}(k_1, k_n))$, then it can be privately optimal for each firm to join a cartel. In this case, the expected benefit of a cartel phase is positive but less than compared with the benchmark case. This is due to the fact that the optimal equilibrium profits from tacit monitoring are below the monopoly level but are above the static Nash equilibrium profits, $\frac{\widehat{m}}{K} > V_c^* > \underline{p}$.

The above implies that forming a cartel is not privately optimal if the expected fine is sufficiently high such that collusion under tacit monitoring is more profitable. However, this is only true if collusion under tacit monitoring is sustainable as an equilibrium. Proposition 6 shows that if collusion under tacit monitoring is more profitable than explicit monitoring, then it is sustainable at lower discount factors, and vice versa.

Proposition 6. *For any given $n \geq 2$ and $K_{-1} \leq \underline{m} \leq \overline{m} < \overline{x}(k_1, k_n)$, if and only if $F \in [0, F^*(k_1))$, then the critical discount factor under explicit monitoring is less than under tacit monitoring, $\delta_e^*(k_n, F) < \delta_c^*(k_1, k_n)$.*
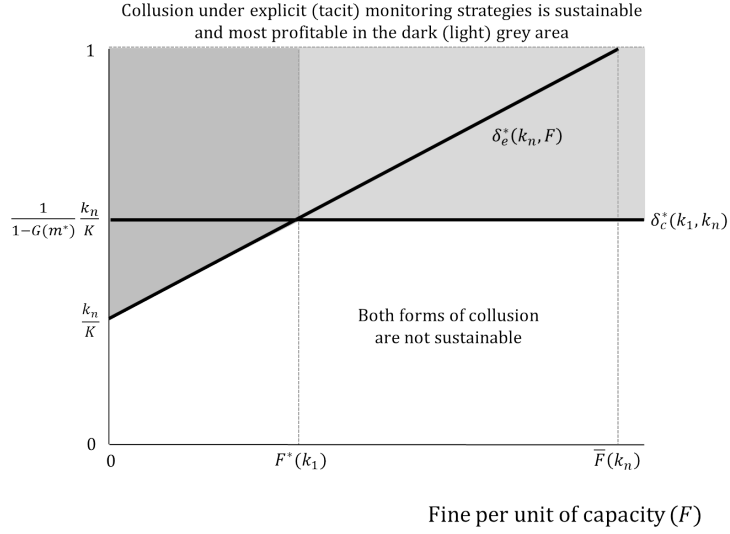
Figure 1: The parameter space of explicit and tacit monitoring

The above results are brought together in Figure 1, which depicts the critical discount factors of both forms of collusion for any given fine per unit of capacity, given tacit monitoring is imperfect, such that $\overline{m} \in [\underline{x}(k_1), \overline{x}(k_1, k_n))$. Note that the critical discount factor under tacit monitoring, $\delta_c^*(k_1, k_n)$, is a horizontal line because it is independent of the fine per unit of capacity F, and it lies between k_n/K and 1 since it equals the former at $\overline{m} = \underline{x}(k_1)$ and it tends to the latter as $\overline{m} \to \overline{x}(k_1, k_n)$. In contrast, the critical discount factor under explicit monitoring, $\delta_e^*(k_n, F)$, is linear and strictly increasing in the fine per unit of capacity F. These two critical discount factors are equal at the critical fine $F^*(k_1)$. If the fine per unit of capacity is below this critical fine, $F < F^*(k_1)$, then collusion under explicit monitoring is more profitable and easier to sustain than tacit monitoring. This is where it is privately optimal for each firm to join a cartel and is signified by the dark grey area in Figure 1. We refer to this as the parameter space of explicit monitoring. In contrast, if the fine per unit of capacity is above the critical fine $F > F^*(k_1)$, then collusion under explicit monitoring is less profitable and more difficult to sustain than tacit monitoring. Thus, the light grey area in Figure 1 highlights the area where collusion under tacit monitoring is most profitable, and we refer to it as the parameter space of tacit monitoring. Outside of these two areas, both forms of collusion are not sustainable. Finally, assuming a mean preserving spread, as the size of the fluctuations in market demand increases the parameter space of explicit monitoring expands at the expense of the parameter space of tacit

23

monitoring. The reason is that the critical discount factor under tacit monitoring, $\delta_c^*(k_1, k_n)$, is strictly increasing in the maximum market demand, $\overline{m}$. Thus, the parameter space of explicit collusion is greatest when there are large fluctuations in market demand where monitoring is *most* difficult.

4.2 Comparing capacity distributions

In this subsection, we analyse the effects of asymmetries in capacities on the incentives for firms to form a cartel when they can alternatively collude tacitly. Holding total capacity constant, we first analyse changes to the capacity distribution that affect the equilibrium analysis either by only increasing the size of the smallest firm or by only increasing the size of the largest firm. We discuss each in turn. Figure 2 builds on the illustration in Figure 1 to depict the effects of such changes on the parameter space of explicit monitoring. Changes to the capacities of the smallest and the largest firms at the same time will have a mix of the two effects discussed here, and we discuss two special cases below.

(a) Increasing the largest firm's capacity
$(k_n' > k_n)$

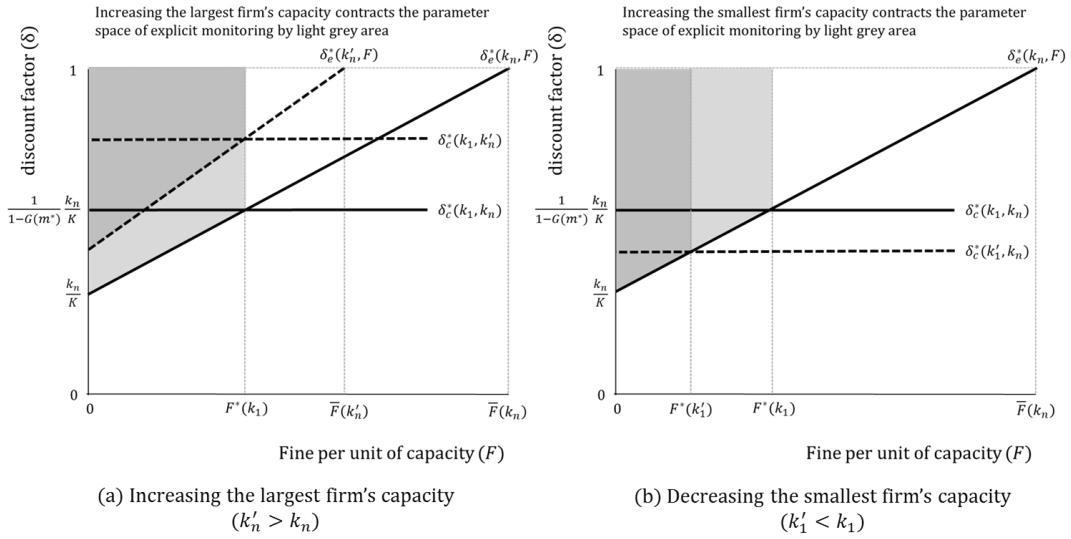(b) Decreasing the smallest firm's capacity
$(k_1' < k_1)$

Figure 2: Changes to the capacity distribution

Increasing the capacity of the largest firm contracts the parameter space of explicit monitoring, when the capacity of the smallest firm is held constant. It follows from Proposition 3 that both critical discount factors rise, because both forms of collusion are more difficult to sustain. Yet, in contrast to the benchmark case, it follows from (15) that there is no effect on the critical

24

fine at which each firm is indifferent between joining a cartel or not, $F^*(k_1)$. This is due to the fact that the expected benefit of a cartel phase is independent of the capacity of the largest firm because, as discussed in section 3.3, the optimal equilibrium profits under tacit monitoring does not vary with the capacity of the largest firm. Thus, the reduction of the parameter space of explicit monitoring is illustrated by the light grey area of Figure 2(a).

Increasing the capacity of the smallest firm can contract the parameter space of explicit monitoring, when the capacity of the largest firm is held constant. This occurs only if tacit monitoring is imperfect, such that $\overline{m} \in [\underline{x}(k_1), \overline{x}(k_1, k_n))$. Under such conditions, it follows from Proposition 3 that the critical discount factor under tacit monitoring falls and that the critical discount factor under explicit monitoring is unaffected. Furthermore, in contrast to the benchmark case, the critical fine $F^*(k_1)$ falls if tacit monitoring is imperfect, which implies that firms have less incentive to join a cartel as the smallest firm gets larger. The reason is that the expected benefit of a cartel phase decreases, because collusion under tacit monitoring is more profitable from Proposition 4. Thus, the reduction of the parameter space of explicit monitoring is illustrated by the light grey area of Figure 2(b).

Now consider two special cases where the capacities of the smallest and largest firm change at the same time. The first is for duopoly, where $k_2 = K_{-1}$ and $k_1 \leq \frac{K}{2}$, such that increasing the largest firm causes the smallest firm to decrease, when the total capacity is held constant. For this case, the two effects discussed above imply that increasing asymmetries between the firms has an ambiguous effect on the parameter space of explicit monitoring when tacit monitoring is imperfect. The reason is that increasing the capacity of the largest firm makes collusion under explicit and tacit monitoring more difficult to sustain, which contracts the parameter space of explicit monitoring at low levels of delta, as illustrated in Figure 2(a). However, this will also decrease the capacity of the smallest firm, which will make collusion under tacit monitoring more difficult to sustain and less profitable than before. This makes explicit monitoring more appealing than before, so $F^*(k_1)$ increases and the parameter space of explicit monitoring expands to the right, which is the reverse of the effect illustrated in Figure 2(b).

The second case is the number of symmetric firms, where $k_1 = k_n = K/n$, such that a decrease in the number of symmetric firms increases each firm's capacity by $\frac{1}{n(n+1)}$, when the total capacity is held constant. Thus, decreasing the number of firms contracts the parameter space of explicit monitoring due to the two effects illustrated in Figure 2(a) and 2(b). First, the 'largest' firm is larger, so both forms of collusion are more difficult to sustain, which contracts the parameter space of explicit monitoring at low levels of delta. Second, the 'smallest' firm is

larger, so collusion under tacit monitoring is easier to sustain and more profitable than before. This makes explicit monitoring relatively less appealing than before, so $F^*(k_1)$ decreases and the parameter space of explicit monitoring contracts to the left.

Bringing together the preceding analysis, if we assume that the parameter space of explicit monitoring is a good proxy for the likelihood that a cartel will arise in a market, then our results imply that the conventional wisdom that cartels should arise in markets with a few symmetric firms is not always correct. Instead, firms can have less incentive to form a cartel in such markets because, holding the size of the largest firm constant, collusion under tacit monitoring becomes can become easier to sustain and more profitable as the smallest firm gets larger, and this makes forming a cartel relatively less appealing. There are a number of important policy implications that follow from these findings and we will discuss these in section 5. Before we do this it is useful to complement our previous analysis by quantifying the extent to which cartels can be deterred more in markets with symmetric firms.

4.3 Example

In this subsection, we analyse an example that shows that the fine necessary and sufficient to deter a cartel can be substantially lower for markets with a symmetric capacity distribution than for market with an asymmetric capacity distribution. In what follows, we restrict attention to the case of $\delta \to 1$, such that collusion under tacit monitoring is not sustainable only if $\overline{m} \geq \overline{x}(k_1, k_n)$. Thus, it then follows from our earlier analysis that the fine that is necessary and sufficient to deter firms from forming a cartel is $F^*(k_1)$ if collusion under tacit monitoring is sustainable, otherwise it is $\overline{F}(k_n)$. For expositional purposes, we transform these fines into the fine per unit sold per period of a cartel phase, which is given by $f \equiv \theta \frac{K}{\overline{m}} F$ for any $\overline{m} < K$. Denoting the level of f at the critical fine level as f^*, it follows from (14) and (15) that:

$$
f^* = \begin{cases}
0 & \text{if} \quad \overline{m} < \underline{x}(k_1) \\
1 - \widehat{p}^c(k_1, \overline{m}) & \text{if} \quad \underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n) \\
1 - \widehat{p}^N(k_n) & \text{if} \quad \overline{x}(k_1, k_n) \leq \overline{m},
\end{cases}
\tag{16}
$$

where $\widehat{p}^c(k_1, \overline{m}) \equiv \frac{K}{\overline{m}} V_c^*$ is the (optimal) average price of collusion under tacit monitoring and $\widehat{p}^N(k_n) \equiv \frac{K}{\overline{m}} \underline{p}$ is the average price of the static Nash equilibrium. Note that f^* has a second interpretation due to the fact that it is equivalent to the expected benefit of the cartel per unit sold per period of a cartel phase. This is the difference between setting the monopoly price p^m and the average price of the most profitable alternative, $\widehat{p}$. Consequently, given demand is
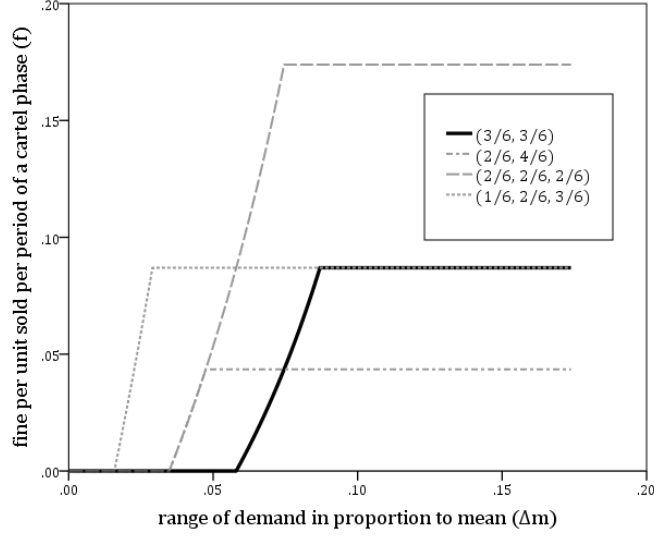
26

Figure 3: $G(m) = \frac{m - \underline{m}}{\overline{m} - \underline{m}}$, $\widehat{m} = 92$, $K_{-1} \leq \frac{5}{6}(100) < 100 = K$, and $\delta \to 1$

perfectly inelastic, f^* also measures the consumer surplus per unit sold of the most profitable alternative to collusion under explicit monitoring, $CS(\widehat{p}) = 1 - \widehat{p}$.

Figure 3 plots f^* as a function of $\Delta m \equiv \frac{\overline{m} - \underline{m}}{\widehat{m}}$ for four capacity distributions where total capacity is $K = 100$. The first two distributions are a symmetric duopoly and a symmetric triopoly, denoted (3/6, 3/6) and (2/6, 2/6, 2/6), respectively, where the first element of the vector relates to firm 1's proportion of the total capacity, the second is firm 2's proportion, and so on. The other two distributions are an asymmetric duopoly, (2/6, 4/6), and an asymmetric triopoly, (1/6, 2/6, 3/6). Note that comparing (2/6, 2/6, 2/6) with (2/6, 4/6) for any given Δm is consistent with moving horizontally from left to right on Figure 2(a) as $\delta \to 1$, because only the capacity of the largest firm increases. Similarly, comparing (1/6, 2/6, 3/6) with (3/6, 3/6) for any given Δm is consistent with moving horizontally from left to right on Figure 2(b) as $\delta \to 1$, because only the capacity of the smallest firm changes. It is assumed that market demand is drawn from a uniform distribution where parameter values are chosen such that $\widehat{m} = 92$ for all Δm and that $K_{-1} \leq \frac{5}{6}(100) \leq \underline{m} \leq \overline{m} \leq K = 100$, so Assumption 1 holds.

Each of the plotted lines in Figure 3 has a similar shape. If f^* equals zero, then tacit monitoring is perfect and forming a cartel has no benefit, since $\widehat{p}^c(k_1, \overline{m}) = p^m$. If f^* is upward-sloping, then tacit monitoring is imperfect and the critical fine to deter the cartel increases with Δm, because collusion under tacit monitoring is less profitable such that $\widehat{p}^c(k_1, \overline{m})$ falls. If f^*

is positive and constant, then collusion under tacit monitoring is not sustainable, such that the expected benefit per unit sold per period of a cartel phase is the difference between p^m and $p^N(k_n)$. This shape implies that, for a given capacity distribution, if a fine is sufficient to deter a cartel for a certain level of Δm, then it will also deter the same cartel at lower levels of Δm. Thus, firms have the greatest incentive to form a cartel when there are large fluctuations in market demand where monitoring is *most* difficult.[5]

Figure 3 illustrates that the different capacity distributions often require substantially different fines to deter firms from forming a cartel. Of particular interest is that the fine required to deter the firms from forming a cartel in the symmetric duopoly (3/6, 3/6) can be lower than that of the asymmetric duopoly (2/6, 4/6). This occurs when fluctuations in market demand are low, such that $\Delta m < 0.07$, and at its largest the difference between the two fines is about 4% of total welfare. Similarly, the fine required to deter the firms from forming a cartel in the symmetric triopoly (2/6, 2/6, 2/6) is lower than that of the asymmetric triopoly (1/6, 2/6, 3/6) at around $\Delta m < 0.06$ and the difference can be as much as 8% of total welfare. Finally, the fine required to deter the firms from forming a cartel in the symmetric duopoly (3/6, 3/6) can be lower than that of the symmetric triopoly (2/6, 2/6, 2/6). This occurs when the fluctuations are large, where the difference between the two fines begins to emerge at around $\Delta m = 0.04$ and after approximately $\Delta m > 0.06$ the difference is in excess of 8% of total welfare.

5   Policy implications

We have shown that, in contrast to the conventional wisdom, firms can have the greatest incentive to form a cartel when there are a large number of firms or asymmetries among them - that is, when tacit collusion is difficult to sustain and not very profitable. This provides an explanation for the puzzling evidence that prosecuted cartels do not tend to be in markets with a few, relatively symmetric firms. Of most relevance is the analysis of Davies and Olczak (2008), who analysed the asymmetries in market shares of members of cartels prosecuted by the European Commission between 1990 and 2006. They found that only 5% of their sample were "tacit-collusive compatible" in that they had market shares consistent with post-merger outcomes that would expect to lead to concerns of tacit collusion in a merger investigation by the European

---

[5]If $\delta < 1$, then there would be a discontinuity in each of the f lines at the threshold of Δm where the outcome is noncollusive. At this thresholds, a line would jump up to the level of the horizontal line, such that this level of fine extends for lower levels of Δm than in Figure 3.

Commission. Our results suggest that cartel activity may be low in such markets due to the fact that tacit collusion is relatively more appealing when firms are symmetric. Furthermore, our analysis suggests that we may expect greater cartel activity in unconcentrated market structures where both the largest and the smallest firm are small, because the parameter space of explicit collusion is greatest under such conditions. Again, this is not inconsistent with the results of Davies and Olczak (2008) as they find that 61% of their sample fell within a group of "unconcentrated" cartels where the largest firm was relatively small and the number of firms in the cartel was relatively high.

Our results have implications for the effectiveness of structural screens that use market characteristics to search proactively for as yet undetected cartels. Despite being credited with some success in practice, scholars are often sceptical over the effectiveness of such screens. For example, Harrington (2006) argues that, even if they correctly indicate the markets in which cartel activity is likely to occur, there is still a high chance of false positives because a collusive outcome is just one of the possible multiple equilibria. Our analysis provides a further reason to be sceptical, as it suggests that structural indicators will not even consistently flag up the markets in which cartel activity is more likely to occur. This is because if the structural indicators are constructed according to the conventional wisdom, then such a screening device will highlight markets in which firms prefer tacit collusion over explicit, and it may overlook markets where explicit collusion is most likely to arise. This problem would be compounded further if proactive investigations were mainly to rely on such methods. The reason being that this would likely lead to the probability of detection being negatively correlated with the degree of asymmetries, which would make explicit collusion even less appealing in symmetric markets and more appealing in more asymmetric markets.

Our analysis also sheds light upon the associated benefit to society of anti-cartel enforcement policy. For instance, in many jurisdictions, including Europe and the US, fines are initially based on the revenue of the cartel members in the cartelised markets (see Bageri *et al.*, 2013). In our model, the total revenue of the cartel is constant for all capacity distributions, so it seems reasonable to assume that the fine per unit of capacity will not vary across cartels (although how the total fine is distributed among its members does vary with each firm's revenue, as consistent with the current approach in the EU and US). Under such an assumption, our results suggest that any given fine is most effective at deterring cartels that cause the smallest marginal detriment to consumer surplus and least effective at deterring those that cause the most. To see this, consider the symmetric duopoly (3/6, 3/6) in Figure 3, and recall that f^* is equivalent to the consumer

surplus per unit sold of the most profitable alternative to collusion under explicit monitoring. Consequently, for any cartel that is deterred from forming, f^* also measures the resultant benefit to society per unit sold from deterrence. Suppose that the fine per unit sold per period of a cartel phase is 5% of total welfare, $f^* = 0.05$ (which is also 5% of the revenue per unit sold). Such a fine will deter a symmetric duopoly from forming a cartel if fluctuations in market demand are small, with the threshold at around $\Delta m < 0.075$. However, for most of this range (that is, around $\Delta m < 0.06$), tacit monitoring is perfect and so there is no increase in consumer surplus associated with deterrence due to the fact that firms can acquire the monopoly profit through tacit collusion. Instead, it is for large fluctuations in market demand, such that $\Delta m > 0.09$, where a cartel causes the greatest marginal detriment to consumer surplus (approximately 8% of total welfare), but a fine of 5% of total welfare will not deter such a cartel.

Finally, the fact that a given fine is less effective at deterring cartels that cause the greatest marginal detriment to consumer surplus raises the question whether there is a more effective manner to calculate fines than basing the fines on revenue.[6] Assuming that fines cannot be set at an arbitrarily high level, it follows from our model that basing fines on the extent to which the cartel enables firms to raise price would be more effective at deterring the most harmful cartels. To see this, again let us return to Figure 3 and note that f^* is the difference between setting the monopoly price, p^m, and the average price of the most profitable alternative, $\widehat{p}$. Thus, it represents the extent to which the cartel is able to raise prices beyond the most profitable alternative - that is, it represents the cartel overcharge. Consequently, basing the fine on the cartel overcharge instead of the revenue could allow fines to follow the shape of f^* in Figure 3 such that fines can be higher for cartels that cause the greatest marginal detriment to consumer surplus, even when the total revenue of the cartel is held constant. In practice, this would involve estimating the difference between the cartel price and some hypothetical 'but-for' price that buyers would have been charged had the cartel not existed. While this difference is often calculated in practice with the but-for price estimated as some competitive benchmark price, the difference here is that the counterfactual could be the (average) price set under tacit collusion. Of course, estimating damages based on the successfulness of tacit collusion is likely to be even more controversial than calculating damages simply based on a competitive benchmark. However, even if the overcharge were based on the static Nash equilibrium, such that the cartel overcharge is overestimated if firms would have alternatively colluded tacitly, then it would be possible to

---

[6]A related body of theoretical literature considers how alternative penalty regimes affect cartel formation and pricing (for example see Katsoulacos *et al.*, 2015).

achieve greater deterrence than fines based on revenue. For example, such a fine would be about 8% of total welfare for the symmetric duopoly (3/6, 3/6) in Figure 3, which would deter such a cartel for all Δm, yet such a fine for the symmetric triopoly (2/6, 2/6, 2/6) would need to be about 17% of total welfare.

6   Non-Verifiable Sales Information

Up to this point, we have assumed that the private information that firms exchange in a cartel is verifiable such that they do not need to consider the truthfulness of each other's information. This simplifies the analysis because we have not needed to consider whether a firm could deviate in an active phase by undercutting the collusive price and under-reporting the level of their sales. This could be a profitable deviation from collusion under explicit monitoring if an active phase continued into the next period rather than switching to an inactive phase. In this subsection, we wish to establish that it is not possible for a firm to gain from such a deviation when its rivals tell the truth.

We assume that in the communication stage of each period firms must simultaneously submit a sales report, r_{it}, which then becomes common knowledge among the firms. Denote the total reported sales in period t as $R_t \equiv \sum_i r_{it}$ and the sum of firm i's rivals' reported sales in period t as $R_{-it} \equiv \sum_{j \neq i} r_{jt}$. We assume that firms follow explicit monitoring strategies with the addition that the strategy prescribes each firm must submit a truthful sales report in the communication stage, $r_{jt} = s_{jt}$, and that an active phase continues into the next period if and only if:

$$ r_{jt} = \frac{k_j}{K} R_t \ \forall \ j \qquad \text{and} \qquad R_t \in [\underline{m}, \overline{m}] . \tag{17} $$

Thus, the difference between this section and the main analysis is that in the main analysis we effectively restricted attention to $r_{it} = s_{it}$, but here we allow $r_{it} \neq s_{it}$. Following the main analysis, we maintain Assumption 1 and for simplicity impose $\overline{m} < K$ to restrict attention to where collusion under tacit monitoring may be sustainable.

Proposition 7. *For any given $n \geq 2$ and $K_{-1} \leq \underline{m} \leq \overline{m} < K$, if firm i's rivals set the collusive price and report their sales truthfully, such that $p_j = p^m$ and $r_{jt} = s_{jt}$ for all $j \neq i$, then an active phase will continue into the next period if and only if firm i set the collusive price and reports its sales truthfully, $p_i = p^m$ and $r_{it} = s_{it}$.*

The intuition is that a non-deviant firm can infer from its own sales what other firms will report if they have set the collusive price. In contrast, since a deviant with a lower price sells

31

its full capacity, k_i, it has no information on the level of the market demand. Consequently, it must guess at the level of sales that will ensure that the active phase continues into the next period. However, given market demand is drawn from a continuous distribution, the probability of guessing correctly is infinitesimally small. Thus, firms cannot gain by deviating and submitting an untruthful report.

7 Concluding remarks

We have analysed the effects of asymmetries in capacity constraints on the incentives of firms to collude explicitly when they can alternatively collude tacitly. In our setting, market demand is uncertain and firms never directly observe their rivals' prices and sales, so firms have the potential to make secret price cuts. Tacit monitoring involves price wars on the equilibrium path when at least one firm receives sufficiently low sales. In contrast, explicit monitoring involves firms secretly sharing their private information in an illegal cartel to improve their ability to monitor each other. Consequently, explicit monitoring can raise profits relative to tacit monitoring by avoiding price wars, but this runs the risk of sanctions. Consistent with other models of collusion with asymmetric capacity constraints, we showed that both forms of collusion are easiest to sustain when capacities are distributed symmetrically. However, in contrast to the conventional wisdom and consistent with some evidence, we found that the incentives for firms to form an illegal cartel can be smallest in markets with a few symmetric firms, because tacit monitoring is relatively more appealing compared to explicit monitoring in such markets.

Our results have two main policy implications. First, regarding the detection of cartels, it raises further questions about the effectiveness of screening devices that use structural indicators to screen for cartels proactively. If such devices are constructed according to the conventional wisdom, then they have the potential to highlight markets conducive to tacit collusion, where explicit collusion is unnecessary, and to overlook markets where cartels can lead to the largest marginal detriment to the buyers. Second, regarding the deterrence of cartels, our analysis highlights that a given fine is least effective at deterring cartels that cause the greatest marginal detriment to buyers. This is true whether the counterfactual is tacit collusion or the static Nash equilibrium. An implication of this is that basing fines on damages has the potential to have greater deterrence effect than basing the fine only on the revenue of the cartel, because then a cartel that causes greater marginal detriment to buyers can receive a higher fine.

References

[1] Abreu, D., Pearce, D. and Stacchetti, E. (1986) "Optimal Cartel Equilibria with Imperfect Monitoring," *Journal of Economic Theory*, 39(1), 251-269

[2] Abreu, D., Pearce, D. and Stacchetti, E. (1990) "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58(5), 1041-1063

[3] Athey, S. and Bagwell, K. (2001) "Optimal Collusion with Private Information," *RAND Journal of Economics*, 32(3), 428–465

[4] Athey, S. and Bagwell, K. (2008) "Collusion with Persistent Cost Shocks," *Econometrica*, 76(3), 493–540

[5] Awaya, Y. and Krishna, V. (2014) "On Tacit versus Explicit Collusion," mimeo, Penn State University

[6] Bageri, V., Katsoulacos, Y. and Spagnolo, G. (2013) "The Distortive Effects of Antitrust Fines Based on Revenue," *The Economic Journal*, 123 (572), 545-557

[7] Bos, I. and Harrington, J. (2010) "Endogenous Cartel Formation with Heterogeneous Firms," *RAND Journal of Economics*, 41(1), 92–117

[8] Bos, I. and Harrington, J. (2015) "Competition Policy and Cartel Size," *International Economic Review*, 56(1), 133-153

[9] Chen, Z. and Rey, P. (2013) "On the Design of Leniency Programs," *Journal of Law and Economics*, 56(4), 917 - 957

[10] Compte, O., Jenny, F. and Rey, P. (2002) "Capacity Constraints, Mergers and Collusion," *European Economic Review*, 46(1), 1-29

[11] Connor, J. (2010) "Recidivism Revealed: Private International Cartels 1991-2009" *Competition Policy International*, 6(2), 101-127

[12] Davies, S. and Olczak, M. (2008) "Tacit versus Overt Collusion Firm Asymmetries and Numbers: What's the Evidence?" *Competition Policy International*, 4(2), 175-200

[13] Davies, S., Olczak, M., and Coles, H. (2011) "Tacit Collusion, Firm Asymmetries and Numbers: Evidence from EC Merger Cases," *International Journal of Industrial Organization*, 29(2), 221-31

[14] Dick, A. (1996) "When Are Cartels Stable Contracts?" *Journal of Law and Economics*, 39(1), 241–8

[15] Fonseca, M. and Normann, H.-T. (2008) "Mergers, Asymmetries and Collusion: Experimental Evidence," *Economic Journal*, 118(1), 287-400

[16] Fonseca, M. and Normann, H.-T. (2012) "Explicit vs. Tacit Collusion – The Impact of Communication in Oligopoly Experiments," *European Economic Review*, 56(8), 1759-1772

[17] Fraas, A., and D. Greer (1977) "Market Structure and Price Collusion: An Empirical Analysis," *Journal of Industrial Economics*, 26(1), 21–44

[18] Fudenberg, D. and Tirole, J. (1991), *Game Theory*, MIT Press

[19] Garrod, L. and Olczak, M. (2016) "Collusion under Imperfect Monitoring with Asymmetric Firms," MPRA Paper No. 70647, https://mpra.ub.uni-muenchen.de/70647/

[20] Green, E. and Porter, R. (1984) "Non-Cooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87-100

[21] Grout, P. and Sonderegger, S. (2005) "Predicting Cartels," A report for the Office of Fair Trading, Economic Discussion Paper 773

[22] Harrington, J. (2006) "How do Cartels Operate?" *Foundations and Trends in Microeconomics*, 2(1), 1-105

[23] Harrington, J. (2008) "Detecting Cartels," in Buccirossi, P., (ed.), *Handbook of Antitrust Economics,* MIT Press

[24] Harrington, J. and Chang, M.-H. (2015) "When Can We Expect a Corporate Leniency Program to Result in Fewer Cartels?," *Journal of Law and Economics*, 58(2), 417-449

[25] Harrington, J. and Skrzypacz, A. (2011) "Private Monitoring and Communication in Cartels: Explaining Recent Collusive Practices," *American Economic Review*, 101(6), 2425-2449

[26] Hay, G. and Kelley, D. (1974) "An Empirical Survey of Price Fixing Conspiracies" *Journal of Law and Economics*, 17(1), 13-38

[27] International Competition Network (2008) "Setting of Fines for Cartels in ICN Jurisdictions," Report to the 7th ICN Annual Conference, Kyoto

[28] Ivaldi, M., Jullien, B., Rey, P., Seabright, P., and Tirole, J. (2003) "The Economics of Tacit Collusion," Report for DG Comp, European Commission

[29] Y. Katsoulacos, Y., Motchenkova, E. and Ulph, D. (2015) "Penalizing Cartels: The Case for Basing Penalties on Price Overcharge," *International Journal of Industrial Organization*, 42, 70-80

[30] Kühn, K.-U., (2012) "How Market Fragmentation Can Facilitate Collusion," *Journal of the European Economic Association*, 10(5), 1116-1140

[31] Lambson, E. (1994) "Some Results on Optimal Penal Codes in Asymmetric Bertrand Supergames," *Journal of Economic Theory*, 62(2), 444-468

[32] Levenstein, M. and Suslow, V. (2006) "What Determines Cartel Success?" *Journal of Economic Literature*, 44(1), 43-95

[33] Martin, S. (2006) "Competition Policy, Collusion and Tacit Collusion," *International Journal of Industrial Organization*, 24(6), 1299-1332

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

[35] Mouraviev, I. (2013) "Explicit Collusion under Antitrust Enforcement," Center for Mathematical Economics Working Paper, no. 494

[36] Motta, M. and Polo, M. (2003) "Leniency Programs and Cartel Prosecution," *International Journal of Industrial Organization*, 21(3), 347-379

[37] Petit, L. (2012) "The Economic Detection Instrument of the Netherlands Competition Authority: The Competition Index," NMa Working Papers, No. 6

[38] Posner, R. (1970) "A Statistical Study of Antitrust Enforcement," *Journal of Law and Economics*, 13(2), 365–419

[39] Spagnolo, G. (2005) "Divide et Impera: Optimal Leniency Programs," CEPR. Discussion Paper No. 4840

[40] Spector, D. (2015) "Facilitating Collusion by Exchanging Non-Verifiable Sales Reports," PSE Working Papers no. 2015-07

[41] Stigler, G. (1964) "A Theory of Oligopoly," *Journal of Political Economy*, 72, 44–61

[42] Symeonidis, G. (2003) "In Which Industries is Collusion More Likely? Evidence from the UK," *Journal of Industrial Economics*, 51(1), 45–74

[43] Tirole, J. (1988) *The Theory of Industrial Organisation*, MIT Press

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

Appendix A: Proofs

Proof of Proposition 3. Differentiating $\delta_e^* (k_n, F) = \frac{1 - S(p^m) + \theta F}{1 - \underline{p}}$ with respect to k_j yields:

$$\frac{\partial \delta_e^*}{\partial k_j} = \frac{1 - S(p^m) + \theta F}{(1 - \underline{p})^2} \frac{\partial \underline{p}}{\partial k_j}.$$

Thus, $\frac{\partial \delta_e^*}{\partial k_n} > 0$ from $\frac{\partial \underline{p}}{\partial k_n} = \frac{1}{k_n^2} \int_{\underline{m}}^{\max\{\overline{m}, K\}} (K - m) g(m) \, dm > 0$ and $S(p^m) \in (0,1)$ for all $\underline{m} < K$ from (2).

For any $\underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n)$, differentiating $\delta_c^* (k_1, k_n) = \frac{1}{(1 - G(m^*(k_1, \overline{m})))} \frac{k_n}{K}$ with respect to k_j yields:

$$\frac{\partial \delta_c^*}{\partial k_j} = \frac{1}{K[1 - G(m^*)]} \left[\frac{\partial k_n}{\partial k_j} + k_n \frac{g(m^*)}{1 - G(m^*)} \frac{\partial m^*}{\partial k_1} \frac{\partial k_1}{\partial k_j} \right].$$

Thus, $\frac{\partial \delta_c^*}{\partial k_1} < 0$ from $\frac{\partial k_n}{\partial k_1} \in [-1, 0]$, $\frac{\partial m^*}{\partial k_1} < 0$ and $\frac{\partial k_1}{\partial k_1} = 1$. Furthermore, $\frac{\partial \delta_c^*}{\partial k_n} > 0$ from $\frac{\partial k_n}{\partial k_n} = 1$, $\frac{\partial m^*}{\partial k_1} < 0$ and $\frac{\partial k_1}{\partial k_n} \in [-1, 0]$. Finally, for any $\overline{m} \leq \underline{x}(k_1)$, $\delta_c^* (k_1, k_n) = \frac{k_n}{K}$ implies $\frac{\partial \delta_c^*}{\partial k_n} > 0$. ∎

Proof of Proposition 4. For any $\underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n)$, differentiating V_c^* with respect to k_j yields:

$$\frac{\partial V_c^*}{\partial k_j} = - \frac{(K - \widehat{m}) g(m^*)}{K (1 - G(m^*))^2} \frac{\partial m^*}{\partial k_1} \frac{\partial k_1}{\partial k_j}.$$

Thus, $\frac{\partial V_c^*}{\partial k_1} > 0$ from $\frac{\partial m^*}{\partial k_1} < 0$, $\widehat{m} < \overline{m} < K$ and $\frac{\partial k_1}{\partial k_1} = 1$. ∎

Proof of Proposition 5. Note that $k_i V_e^* > k_i V_c^*$ if and only if $F < \frac{1}{\theta} \left(\frac{\widehat{m}}{K} - V_c^* \right) \equiv F^* (k_1)$, where $V_c^* = \frac{1}{K} \left(\frac{\widehat{m} - G(m^*) K}{1 - G(m^*)} \right)$ from Proposition 2. For any $\overline{m} < \underline{x}(k_1)$, such that $G(m^*) = 0$, then $F^* = 0$. It also follows from $\lim_{\overline{m} \to \overline{x}(k_1, k_n)} V_c^* = \underline{p}$ that $\lim_{\overline{m} \to \overline{x}(k_1, k_n)} F^* (k_1) = \overline{F}(k_n)$. Finally, if $\underline{x}(k_1) \leq \overline{m} < \overline{x}(k_1, k_n)$, then $F^* \in \left(0, \overline{F}(k_n) \right)$ from $V_c^* \in \left(\underline{p}, \frac{\widehat{m}}{K} \right)$. ∎

Proof of Proposition 6. For any $\overline{m} < K$, $\delta_e^* < \delta_c^*$ if and only if:

$$\frac{1 - \frac{\widehat{m}}{K} + \theta F}{1 - \underline{p}} < \frac{1}{1 - G(m^* (k_1, \overline{m}))} \frac{k_n}{K}.$$

Rearranging the above yields:

$$F \;<\; \frac{1}{\theta}\left[\frac{1-p}{1-G(m^*(k_1,\overline{m}))}\frac{k_n}{K} - \left(1-\frac{\widehat{m}}{K}\right)\right]$$
$$<\; \frac{1}{\theta}\left(\frac{\widehat{m}}{K} - \frac{1}{K}\left(\frac{\widehat{m}-G(m^*)K}{1-G(m^*)}\right)\right),$$

where the right-hand side can be rewritten as $\frac{1}{\theta}\left(\frac{\widehat{m}}{K} - V_c^*\right) = F^*(k_1)$. ∎

Proof of Proposition 7. We wish to show that if firm i's rivals set $p_j = p^m$ and report $r_{jt} = s_{jt}$ for all $j \neq i$, then firm i cannot gain from deviating from its prescribed report of $r_{it} = s_{it}$, for any arbitrary price for firm i, and for any feasible level of sales for firm i, s_{it}.

To begin, consider the case where $p_i = p^m$. If all firm i's rivals set p^m and report truthful sales, then $r_{jt} = s_{jt} = k_j\frac{m_t}{K}$ for all $j \neq i$ such that $R_{-it} = \frac{K_{-i}}{K}m_t$. It follows from this that if firm i reports truthfully such that $r_{it} = s_{it} = k_i\frac{m_t}{K}$, then the active phase continues into the next period, since $r_{jt} = k_j\frac{m_t}{K}\ \forall\ j$ and $R_t = m_t \in [\underline{m},\overline{m}]$; otherwise, firms enter an inactive phase. Thus, firm i will report its sales truthfully such that the active phase continues if:

$$k_i V^e \geq (1-\delta)\left(\pi_i(p^m) - \theta k_i F\right) + \delta\pi_i^N(k_n),$$

which is true for all $F < \overline{F}(k_n)$.

Next, suppose $p_i < p^m$. If all firm i's rivals set p^m and report sales truthfully, then $r_{jt} = s_{jt} = k_j\left(\frac{m_t - k_i}{K_{-i}}\right)$ for $j \neq i$ such that $R_{-it} = m_t - k_i$. Thus, the active phase continues into the next period if firm i reports its sales untruthfully, such that:

$$r_{it} = k_i\left(\frac{m_t - k_i}{K_{-i}}\right) < s_{it} = k_i,$$

and if $R_t \in [\underline{m},\overline{m}]$; otherwise firms enter an inactive phase. Given firm i does not know m_t, suppose it reports $r_{it} = k_i\left(\frac{m-k_i}{K_{-i}}\right)$ for some $m \in [\underline{m},\overline{m}]$. Then its present discounted value of submitting an untruthful report of $r_{it} = k_i\left(\frac{m-k_i}{K_{-i}}\right)$ is:

$$(1-\delta)(k_i - \theta k_i F) + \delta\left[\Pr(m = m_t)k_i V^e + (1 - \Pr(m = m_t))\pi_i^N(k_n)\right],$$

for all $m \in [\underline{m},\overline{m}]$. Furthermore, given $\Pr(m = m_t) = 0$ for all $m \in [\underline{m},\overline{m}]$ due to the continuous distribution, these profits are the same as it would get if it submitted a truthful report: $(1-\delta)(k_i - \theta k_i F) + \delta\pi_i^N(k_n)$. Therefore, firm i, for all i, cannot gain by submitting an untruthful report, $r_{it} \neq k_i$ for any $p_i < p^m$.

Finally, suppose $p_i > p^m$. If all firm i's rivals set p^m and report sales truthfully, then $r_{jt} = s_{jt} = k_j$ for $j \neq i$ such that $R_{-it} = K_{-i}$. Thus, the active phase continues into the next

period if firm i reports its sales untruthfully, such that:

$$r_{it} = k_i > s_{it} = m_t - K_{-i},$$

and if $R_t = K \in [\underline{m}, \overline{m}]$; otherwise firms enter an inactive phase. Therefore, $\overline{m} < K$ suffices to ensure that firm i cannot gain by submitting an untruthful report, $r_{it} = k_i > s_{it}$. ∎

Appendix B: The extent of leniency

In the main text, we assumed that an informant receives full leniency. While this is consistent with contemporary leniency programs in Europe and the US, this has not always been the case. For example, leniency was introduced to Europe in 1996, but full leniency was not guaranteed to the first firm to come forward until the program was adjusted in 2002. In this subsection, we consider our results when an informant receives less than full leniency to analyse the effect of this change. More specifically, we now consider a leniency programme which reduces the informant's fine from $k_i F$ to $k_i F (1 - \lambda)$, such that the informant receives less than full leniency if $\lambda < 1$. For brevity, we assume that leniency is sufficiently generous that $\lambda > 1 - \theta$. This implies that applying for leniency will reduce the informant's expected fine, since $k_i F (1 - \lambda) < \theta k_i F$. Nevertheless, to obtain the results for the case of $1 - \lambda \geq \theta$, one simply needs to substitute $\lambda = 1 - \theta$ into the below, because then no firm will apply for leniency on or off the equilibrium path. All other assumptions are unchanged.

This change to the extent of leniency only affects explicit collusion. However, it does not change the profits under explicit collusion, because firms do not apply for leniency on the equilibrium path. Hence, $k_i V^e$ is the same as in (8), and as a result $F^* (k_1)$ is the same as in (15). Instead it only changes the critical discount factor, because it affects the deviation profits. Following similar arguments as in section 4.1, firm i has no incentive to deviate in the communication stage if:

$$k_i V^e \geq (1 - \delta) (\pi_i (p^m) - k_i F (1 - \lambda)) + \delta \pi_i^N (k_n), \tag{18}$$

and firm i has no incentive to deviate in the pricing stage if

$$k_i V^e \geq (1 - \delta) (k_i - k_i F (1 - \lambda)) + \delta \pi_i^N (k_n). \tag{19}$$

These ICCs are the same as (5) and (6), respectively, except that deviating by applying for leniency only reduces the fine by $k_i F \lambda$. Furthermore, consistent with the main analysis, it is easy to check that the pricing ICC (6) is more stringent and hence determines the critical discount
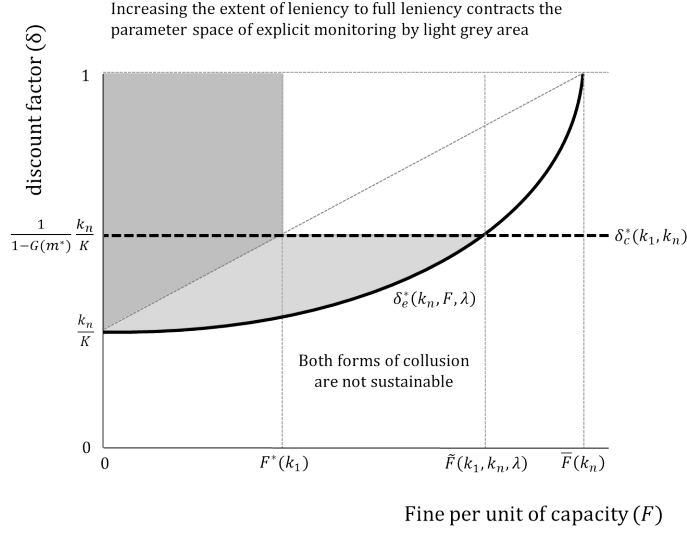
Figure 4: The parameter space of explicit collusion

factor. Thus, rearranging (19) in terms of the discount factor yields:

$$\delta \geq \delta_e^* (k_n, F, \lambda) \equiv \frac{1 - V^e - (1 - \lambda) F}{1 - \underline{p} - (1 - \lambda) F}.$$

As in the main analysis, the critical discount factor $\delta_e^* (k_n, F, \lambda)$ is strictly increasing in fine per unit of capacity, F, but the difference is that it is convex in F for any $\lambda < 1$. This implies that increasing the fine raises the critical discount factor to a greater extent for higher fines.

Figure 4 illustrates the effect of leniency under the assumption that tacit monitoring is imperfect such that $\overline{m} \in [\underline{x}(k_1), \overline{x}(k_1, k_n))$. The critical discount factor of full leniency, where $\lambda = 1$, that we analysed in the main text, is given by the linear upward sloping line. In contrast, if $\lambda < 1$, then the critical discount factor $\delta_e^* (k_n, F, \lambda)$ is nonlinear and as a consequence it no longer intersects with $\delta_c^* (k_1, k_n)$ at $F^* (k_1)$. Instead, it does so at a higher fine, denoted $\widetilde{F} (k_1, k_n, \lambda)$, and this implies that the parameter space of explicit monitoring expands if there is less than full leniency. The extent of this expansion compared to the main text is illustrated by the light grey area in Figure 4. Thus, assuming that the parameter space of explicit monitoring is a good proxy of cartel activity, the above implies that raising the extent of leniency from $\lambda < 1$ to full leniency is likely to make cartel activity less likely. Furthermore, note that the light grey area in Figure will contract if the smallest firm gets larger. This is due to the fact that, when such monitoring is imperfect, collusion under tacit monitoring is easier to sustain (such that $\delta_c^* (k_1, k_n)$ falls) and is more profitable (such that $F^* (k_1)$ falls). Consequently, this implies that increasing the extent

39

of leniency is likely to be less successful at deterring cartels in markets with symmetric firms, where the smallest firm is as large as possible, than compared with markets where the smallest firm is smaller, other things equal.

Appendix C: Exploiting the leniency programme

In this appendix, we analyse an alternative strategy in which the firms in the cartel apply for leniency in every period in an attempt to reduce their expected fines. We refer to this as strategic leniency strategies. Formally, under this strategy profile, in the pricing stage of a period during an active cartel phase, each firm sets the collusive price p^m. Then, in the communication stage sales are realised and each firm secretly shares its private information with its rivals and always applies for leniency. The active phase continues into the next period $t + 1$ if $p_{jt} = p^m$ and $s_{jt} = k_i S(p^m)$ for all j. Otherwise, firms enter an inactive phase and each firm prices according to the static Nash equilibrium forever and never exchanges its private information.[7]

Recall from section 3 that, consistent with Spagnolo (2005), we assume that the firm with the lowest price is given leniency when there is more than one informant. This contrasts with Chen and Rey (2013), who assume that any informant is randomly selected with equal probabilities, regardless of the prices they set. We follow the approach of Spagnolo (2005) for two reasons. First, it makes collusion more difficult in that the critical discount factor is higher than if a firm is randomly selected from all of the informants. Consequently, the competition agency has an incentive to favour a deviant informant over colluding informants. Second, in reality the first informant to apply for leniency is usually selected, and it seems to us that a deviating informant will be able to plan its deviation such that it beats its rivals in the race for leniency.

Under this alternative strategy profile all firms apply for leniency and the cartel is detected every period, so the per-period probability that firm i's leniency application is successful and it avoids paying a fine is $\frac{1}{n}$. Therefore, given conviction does not lead to the breakdown of the cartel, firm i's expected (normalised) discounted profit if each firm abides by its prescribed strategy is:

$$k_i V^s = \pi_i(p^m) - \frac{n-1}{n} k_i F. \tag{20}$$

As in the main text, this says that in a period during an active cartel phase that firm i expects

---

[7]As in the main analysis the fact that firms set the monopoly price is without loss of generality, because it is easy to check that if firms set a lower price, then the equilibrium profits are lower and the critical discount factor is higher.

to receive the expected per-period profit from setting p^m minus the expected fine. The only difference is that the expected fine now depends upon the probability that a strategic leniency application is successful rather than the probability the cartel is detected.

We next consider whether firm i has an incentive to deviate from its prescribed strategy. First suppose firm i has abided by its strategy in the pricing stage of period t by setting $p_i = p^m$ and consider deviations in the communication stage. Firm i can deviate here by not applying for leniency and/or not sharing its private information. However, not applying for leniency increases the expected fine a firm faces and not sharing its private information means that firms will enter an inactive phase. Therefore, for any $k_i V^s \geq \pi_i^N (k_n)$ there is no gain from deviating in the communication stage. Thus, we only need to consider the pricing ICC.

Now, consider firm i's incentive to deviate from its prescribed strategy in the pricing stage of period t. It follows from above that if firm i deviates in the pricing stage, then in the communication stage it will apply for leniency and firms will enter a punishment phase. In contrast, if a firm does not deviate in the pricing stage, then it will share its private information and apply for leniency. Thus, firm i has no incentive to deviate in the pricing stage if:

$$k_i V^s \geq (1 - \delta) k_i + \delta \pi_i^N (k_n) \tag{21}$$

where the left-hand side is firm i's profit from abiding by its strategy, and the right-hand side is the profit from firm i's optimal deviation. The first term on the right-hand side is firm i's profit from undercutting p^m marginally to supply its full capacity, k_i, which is firm i's optimal deviation from $p^m > V^s > \underline{p}$. Rearranging (21) yields:

$$\delta \geq \frac{1 - V^s}{1 - \underline{p}} \equiv \delta_s^* (k_n, n, F)$$

where the fact that the right-hand side of (7) is independent of k_i implies that if the pricing ICC holds for firm i, then it holds for all of its rivals $j \neq i$. This implies that, despite potential asymmetries, each firm has the same incentive to deviate as its rivals.

Thus, it follows from the above that if firms are sufficiently patient, such that $\delta \geq \delta_s^* (k_n, n, F)$, then firm i's optimal equilibrium profits are given by (20) for all i. Finally, it is easy to check that there exists a unique fine per unit of capacity, $\overline{F} (k_n, n) \equiv \frac{n}{n-1} \left(S(p^m) - \underline{p} \right) > 0$, that is the level of F that sets the optimal equilibrium profits equal to the static Nash equilibrium profits, $k_i V_s^* = \pi_i^N (k_n)$, such that if the fine per unit of capacity is sufficiently low, $F \in \left[0, \overline{F} (k_n, n) \right)$, then $\delta_s^* (k_n, n, F) < 1$ and $k_i V_s^* > \pi_i^N (k_n)$.

Comparing the optimal SPNE profits under strategic leniency strategies with those under explicit monitoring strategies in Proposition 1, shows that it is more profitable for the cartel to

not apply for leniency, such that $V_e^* > V_s^*$, if $\theta < \frac{n-1}{n}$. The reason is that in this case informing the competition agency raises the expected fine firms face. The most profitable collusive strategy profile is then as analysed in the main text. Furthermore, comparing the critical discount factor under strategic leniency strategies with the equivalent under explicit monitoring strategies as given in Proposition 1, shows that $\delta_e^* (k_n, F) \leq \delta_s^* (k_n, n, F)$ if and only if $\theta \leq \frac{n-1}{n}$. This implies that if collusion under explicit monitoring strategies is more profitable than under strategic leniency strategies, then it is also sustainable at lower discount factors. The above implies that $\theta < \frac{1}{2}$ is a sufficient condition for firms to never inform the competition agency about the cartel as part of the collusive strategy. Given the low detection rates of cartels, this seems likely to hold in most jurisdictions. Hence, the focus on explicit monitoring strategies in our main analysis.