Collusion under Private Monitoring with Asymmetric Capacity Constraints*

Luke Garrod† and Matthew Olczak‡

March 5, 2014

Abstract

We explore the effects of asymmetries in capacity constraints on collusion where demand is uncertain and where firms must monitor the agreement through their privately observed sales and prices. We show that deviations will be detected perfectly when demand fluctuations are sufficiently small. Otherwise, monitoring is imperfect and punishment phases must occur on the equilibrium path. Collusion is hindered in both cases when the largest firm has more capacity and when the smallest firm has less. We demonstrate that a merger with a collusive symmetric outcome can have a lower average best equilibrium price than a more asymmetric noncollusive outcome.

*This paper builds on Centre for Competition Policy Working Paper 10-3. We are grateful for comments from Steve Davies, Morten Hviid, Bruce Lyons, Peter Ormosi, Maarten Pieter Schinkel and Chris M. Wilson.

†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

The recent collusion theory literature has developed a clear consensus that greater asymmetries undermine the sustainability of collusion. For example, this result is robust to whether asymmetries are in terms of firms' capacity constraints (see Compte *et al.*, 2002; Vasconcelos, 2005; and Bos and Harrington, 2010) or the number of differentiated products that each firm sells (see Kühn, 2004). These advances have been particularly important for merger policy as they have highlighted which types of mergers can increase the likelihood of tacit collusion. In particular, with respect to capacity constraints, Compte *et al.* (2002) show that collusion is more difficult as the capacity of the largest firm is increased through a merger, and Vasconcelos (2005) finds that collusion is hindered when the largest firm has more capacity and when the smallest firm has less. Bos and Harrington (2010) show that increasing the capacity of medium-sized firms can facilitate collusion, when the collusive agreement does not encompass all firms in the market.

All of these papers analyse collusion under perfect observability where firms can directly observe their rivals' actions, so that they can immediately detect when a rival has deviated from the collusive agreement. In contrast, many mergers occur in markets in which there is imperfect observability, because there is the potential for secret price cuts. This may be the case, for example, in upstream business-to-business markets where transaction prices can be unrelated to posted prices. Consequently, it is inappropriate to consider the competitive effects of such mergers in terms of collusion under perfect observability and the effects should instead be considered in the context of collusion under imperfect monitoring (see Green and Porter, 1984; Harrington and Skrzypacz, 2007 and 2011). However, while the models in this literature provide many interesting insights into the sustainability of collusion, it is difficult to draw implications for merger policy from them, because they analyse collusion with symmetric firms.

In this paper, we begin to fill this gap in the literature by exploring the effects of asymmetries in capacity constraints on collusion under imperfect observability. We achieve this by extending Compte *et al.* (2002) to a setting where there is demand uncertainty and where firms never directly observe their rivals' prices and sales. Thus, similar to the imperfect monitoring setting first discussed by Stigler (1964), each firm must monitor the collusive agreement using their own privately observed sales and prices, and in this regard our model is related to Tirole's (1988, p.262-264) model of private monitoring that captures the results of Green and Porter (1984) in a Bertrand framework. Similar to Tirole (1988), firms may need to solve a non-trivial signal extraction problem in our model, which means that, in contrast to Compte *et al.* (2002), punishment phases can occur on the equilibrium path. Using this model, we consider whether collusion is facilitated or hindered as a given amount of capacity is reallocated among the firms. We find, consistent with Compte *et al.* (2002), that collusion is hindered as the size of the largest firm is increased. Yet, unlike Compte *et al.* (2002) but similar to Vasconcelos (2005), collusion is facilitated when the capacity of the smallest firm is increased.

Our results are closest to Compte *et al.* (2002) when private monitoring is perfect in the sense that all firms can detect any deviation with certainty from their privately observed sales and prices. In this case, punishment phases do not occur on the equilibrium path and the size of the largest firm's capacity affects the critical discount factor in the same manner that it does for Compte *et al.* (2002). However, we demonstrate that this only occurs for sufficiently small fluctuations in market demand and that the critical level is determined by the size of the capacity of the smallest firm. The reason is that there is perfect private monitoring if each firm's set of collusive sales and its set of sales when at least one of its rivals has deviated are mutually exclusive. This is less likely to be true when the capacity of the smallest firm is reduced, holding total capacity constant, because the smallest firm can now supply less of the market demand when it undercuts the collusive price than before, which means its rivals' resultant maximum possible sales will be closer to their minimum possible collusive sales.

When market demand fluctuates to the extent that private monitoring is imperfect in the sense that firms cannot detect a deviation with certainty, we restrict our attention to a strategy profile in which firms can coordinate whether they behave collusively or noncollusively in each period. Thus, similar to Tirole (1988), firms enter a punishment phase in which they play the static Nash equilibrium for a known number of periods, when at least one firm receives collusive sales in the preceding period that are consistent with a deviation.[1] Given such a strategy profile, we find that the critical discount factor is strictly increasing in the capacity of the largest firm and is strictly decreasing in the capacity of the smallest firm. Moreover, we also find that the capacity allocation affects the average price associated with the best collusive equilibrium because, unlike in the models of collusion under perfect observability, it determines the frequency and the optimal

[1]While we use trigger strategies as opposed to penal codes (see Abreu *et al.*, 1986) in the main analysis, we show in the robustness section that our main results when the length of the punishment phase is determined optimally are equivalent to optimal symmetric penal codes, where the punishment phase may last for one period and firms set a low common price that is determined optimally. The reason is that, although the punishment can be shorter for this alternative strategy, the per-period punishment will be harsher to the extent that the critical discount factor and the best collusive profits are the same. Our approach allows us to highlight the fact that these two strategies generate the same results.

length of punishment phases on the equilibrium path. Specifically, we demonstrate that the best average price is strictly increasing in the capacity of the smallest firm but is independent of the capacity of the largest firm. The intuition for why the size of the smallest firm's capacity affects the critical discount factor and the best average price is due to the fact that punishment phases occur less often when it has more capacity, because the larger firms' sales are less likely to be consistent with a deviation by the smallest firm. In contrast, the size of the largest firm is unimportant for the best average price because, although the static Nash equilibrium profits are strictly increasing in the size of the largest firm, which raises profits on the equilibrium path, it also weakens the punishment, so the length of the punishment phase increases to such an extent that it ensures that the effect on the (optimal) punishment is neutral.

After solving the model, we then use it to draw implications for merger policy. In particular, we analyse both the unilateral and coordinated effects of mergers in a unified framework. Unilateral effects arise if the merged entity is likely to have an individual incentive to raise prices post-merger, whereas coordinated effects arise if the merger results in an increased likelihood and sustainability of tacit collusion (see Ivaldi *et al.*, 2003a and 2003b). It is well understood that the former are associated with asymmetric post-merger market structures and the latter with symmetric post-merger market structures. Previously, it has been only possible to study these effects independently of each other. For example, in the framework of Compte *et al.* (2002), either the monopoly price is sustainable, in which case only coordinated effects matter, or collusion is not sustainable at any price, in which case only unilateral effects matter. In contrast, our model allows for a more continuous treatment of such effects, because play alternates between phases of collusion and competition on the equilibrium path.

The conventional wisdom is that coordinated effects are more harmful to welfare than unilateral effects, because the fear is that firms will share the monopoly profit in every future period if collusion is sustainable (for example, see Compte *et al.*, 2002). This logic, based upon collusion under perfect observability, also implies that a merger which disrupts collusion, by enhancing the market power of a single firm, may actually increase consumer surplus post-merger because, as described by Röller and Mano (2006, p.22): "it is preferable that any coordination is by only a subset of firms (i.e. the merging parties) rather than all firms (tacitly)." However, we show, as conjectured by Kühn (2001) and Motta *et al.* (2003), that this logic does not always hold under imperfect monitoring. The reason is that collusion may not enable the firms to share the monopoly profit between them in every future period, because punishment phases can occur on the equilibrium path. Consequently, a merger that facilitates collusion by allocating capacity symmetrically would be less harmful to welfare than one that creates a near monopoly. We demonstrate that a collusive symmetric outcome will have lower prices on average than a noncollusive asymmetric outcome, when fluctuations in demand are sufficiently large.

The rest of the paper is organised as follows. Section 2 discusses how our paper relates to other models of collusion under imperfect monitoring. Section 3 sets out the assumptions of the model and solves for the static Nash equilibrium. In section 4, we first show that private monitoring is perfect when demand fluctuates to a small extent and we demonstrate how the capacity distribution affects firms' ability to monitor the collusive agreement perfectly. We then analyse larger fluctuations in market demand for which there is imperfect private monitoring and find the conditions for which collusion is sustainable, given that firms' strategies allow them to coordinate when they should behave collusively and noncollusively. Then we consider how the capacity allocation affects the critical discount factor and the best average price, drawing implications for mergers. In section 5, we analyse an example to compare the unilateral and coordinated effects of mergers on consumer surplus. Section 6 explores the robustness of our results, and section 7 concludes. All proofs are relegated to the appendix.

2 Related Literature

Imperfect monitoring was first discussed by Stigler (1964) in a collusive setting, similar to ours, in which each firm's prices and sales are private information. In contrast, more recent applied game theoretic models of imperfect monitoring have tended to have a public signal that ensures that firms can coordinate whether they behave collusively or noncollusively in each period, for any history. The benefit of such settings is that the public signals ensure that the analysis is relatively simple, because equilibrium strategies have a recursive structure. For example, in the classic framework developed by Green and Porter (1984), there are fluctuations in unobservable market demand over time and firms compete in quantities, where each firm's sales are private information but the resulting market price is publically observable. Consequently, when the market price is low, firms are unsure whether this is due to low demand or that a rival deviated. Thus, firms condition their play on the level of the market price, entering a punishment phase whenever price is below some trigger level. Similarly, in other papers in which price is the strategic variable and firm demand is stochastic, prices are not publically observable but play can be conditioned on sales because either each firm's sales are assumed to be public information (see Skrzypacz and Hopenhayn, 2004; and Harrington and Skrzypacz, 2007) or firms can provide a credible signal of the size of their sales to their rivals (see Harrington and Skrzypacz, 2011).

While there is no such directly observable public signal in our model, we restrict attention to a strategy profile where firms can coordinate whether they should behave collusively or noncollusively in each period. This aspect of coordination is similar to Tirole's (1988) model of private monitoring, in which there are two symmetric firms selling a homogeneous product, without capacity constraints, and where there is a chance that market demand is either high or low (see also Campbell *et al.*, 2005). Under these assumptions, there is imperfect monitoring when demand in the low state is zero, because a firm cannot be sure that making no sales is due to low demand or due to a deviation by its rival. Nevertheless, firms can sustain collusion by competing for a known number of periods, when at least one firm receives collusive sales in the preceding period that are consistent with a rival deviating from the strategy, and by colluding in any period otherwise. This strategy implies that firms should enter a punishment phase whenever at least one firm makes zero sales, and this coordinates behaviour because this information is always common knowledge. For instance, if a firm makes zero sales because market demand is zero, its rival will also make zero sales, regardless of the prices chosen. In contrast, if a firm makes zero sales because its rival deviated and demand is high, then the deviant makes twice the sales it would have had it set the collusive price, and from this information it can infer that its rival made no sales.

To get a basic understanding of how our model compares with Tirole (1988), consider adding symmetric capacity constraints to firms in the setting above and suppose that they can always supply the market demand collectively but never individually. The first feature to notice is that we do not require the unrealistic assumption that market demand is zero in some periods to generate imperfect monitoring. This is because a firm will supply the residual demand in the event of a deviation by its rival (i.e. market demand minus the deviant's capacity). Thus, there is imperfect private monitoring when each firm's collusive sales in the low demand state are the same as the residual demand of the high state. Moreover, when firms follow the same strategy as above, a punishment phase is triggered whenever at least one firm's collusive sales in the preceding period are less than or equal to the residual demand in a high demand state. Again, this is common knowledge because if a firm's sales equal the residual demand in the high state due only to low market demand, its rival will also make low sales. Yet, if a firm makes such sales because its rival deviated and demand is high, then the deviant sells its capacity, and from this information it can infer that its rival made low sales. The difference here is that if a firm's sales equal the residual demand in the low state, it will know for sure that its rival deviated, because such sales are only consistent with a deviation. However, to ensure coordination, such deviations (that will not occur on the equilibrium path) cannot be punished differently because, while the deviant can infer that its rival will receive sales that are low enough to trigger a punishment phase, it has no way of telling how low its rival's sales are, since it supplies its capacity regardless of the level of demand.[2]

As well as being consistent with Tirole (1988), this feature that firms condition their play on information that is common knowledge, rather than acting upon their private information, is also consistent with the concept of public strategies, which are used by the models of imperfect public monitoring discussed above.[3] Furthermore, we show that the information that at least one firm's sales are consistent with a deviation is always common knowledge, even if play is off the equilibrium path. Thus, consistent with one of the two necessary conditions for a strategy profile to be perfect public equilibrium (PPE) (see Fudenberg and Tirole, 1994, p.188), we can check, for any date and any such common knowledge history, whether our strategies yield a Nash equilibrium from that date on. Our equilibrium stratgies are not PPE strategies, however, because they do not satisfy the other necessary condition: that each firm's strategy is a public strategy, in which firms ignore their private information in choosing their actions. Instead, firms use their private information to infer the information that is common knowledge in our equilibrium strategies, so they are not technically public strategies.[4]

Finally, focusing on a strategy profile in which firms can coordinate whether they behave collusively or noncollusively in all periods allows us to avoid the challenges that face other models

[2] Another paper that is related to ours is Marshall and Marx (2008), who consider the incentives to communicate for unconstrained and symmetric firms, when each firm's prices and sales are private information. In their setting without communication, they assume that any deviation is perfectly detected by all firms so that, similar to our model, they can coordinate whether they behave collusively or noncollusively in each period. However, our approach differs to theirs because the condition that allows the (possibly asymmetric) firms to coordinate is determined endogenously in our model.

[3] A strategy σ_i is a public strategy if $\sigma_i^t\left(h^t, z_i^t\right) = \sigma_i^t\left(h^t, \widetilde{z}_i^t\right)$ for all periods t, public histories h^t, and private histories z_i^t and $\widetilde{z}_i^t$ (See Fudenberg and Tirole, 1994, p.187).

[4] Put differently, if there were a mechanism that made publically available the information of whether at least one firm's sales are consistent with a deviation, such as a trade association that verified each firm's sales and announced this information, then our equilibrium strategies would indeed be PPE strategies, because firms then could condition their play on the public signal and ignore their private information. Of course, such a mechanism would actually be superfluous because the information provided by the trade association could already be inferred by each firm from its private information (and even if such a mechanism existed, the firms would want it to provide better information to eliminate the non-trivial signal extraction problem, removing the need for punishment phases on the equilibrium path).

of collusion under private monitoring (for example, see Fudenberg and Levine, 1991; Mailath and Morris, 2002). In such models, players cannot always coordinate their behaviour, so a player may enter the punishment phase unsure of whether a rival will do so as well. These models are more complicated than when coordination is possible, because they tend not to have a recursive structure. While future research should address whether firms have incentives to follow strategies that coordinate their play or not, it seems to us that analysing the situation where firms can coordinate their behaviour is the most sensible place to start to analyse imperfect monitoring and asymmetries in our applied oligopoly setting.

3 The Model

3.1 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. Firms' costs are normalised to zero and they have a common discount factor, $\delta \in (0, 1)$. In any period t, firms set prices simultaneously where $\mathbf{p}_t = \{p_{it}, \mathbf{p}_{-it}\}$ is the vector of prices set in period t, p_{it} is the price of firm $i = \{1, \ldots, n\}$ and $\mathbf{p}_{-it}$ is the vector of prices of all of firm i's rivals. Market demand consists of a mass of m_t buyers, each of whom are willing to buy one unit provided the price, without loss of generality, does not exceed 1. We assume that firms are uncertain of the level of market demand 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 $[\underline{m}, \overline{m}]$.[5]

Buyers are informed of prices, so they will want to buy from the cheapest firm. However, the maximum that firm i can supply in any period is k_i, where without loss of generality let $k_n \geq k_{n-1} \geq \ldots\ldots \geq k_1 > 0$. We denote total capacity as $K \equiv \sum_i k_i$ and the maximum that firm i's rivals can supply in each period as $K_{-i} \equiv \sum_{j \neq i} k_j$. In contrast to the buyers, firm i never observes firm j's prices, $p_{j\tau}$, or sales, $s_{j\tau}$, $j \neq i$, for all $\tau \in \{0, \ldots, t-1\}$. This implies that each firm's past prices and sales are its private information in period t, denoted $z_i^t \equiv \{p_{i0}, s_{i0}; \ldots; p_{it-1}, s_{it-1}\}$. Thus, similar to Tirole (1988), our setting has the feature that all buyers are fully aware of prices, yet all firms are only aware of their own prices. Such a 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

[5]We consider the implications of downward-sloping demand in section 6.

information.[6]

3.2 Demand rationing and sales

Following the other papers in the literature (for example, Vasconcelos, 2005; and Bos and Harrington, 2010), we assume that demand is allocated using the proportional rationing rule, which can be described as follows:

The proportional rationing rule

- Demand is allocated to the firm with the lowest price first. If this firm's capacity is exhausted, demand is then allocated to the firm with the second lowest price, and so on.
- If two or more firms set the same price and if their joint capacity is sufficient to supply the remaining buyers, then the (residual) demand is allocated proportionally to capacity.

To ensure that firm i's sales in period t, $s_{it}(p_{it}, \mathbf{p}_{-it}; m_t)$, are positive for all i and m_t, even when it is the highest-priced firm, we place the following plausible yet potentially restrictive assumption on the capacity allocation:

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

For a given level of $\underline{m}$, Assumption 1 is not restrictive with respect to the size of the smallest firm's capacity if all firms can only ever collectively supply as much as the minimum market demand, $\underline{m} \geq K$. Otherwise, there is a restriction on the size of the smallest firm in that it cannot be too small, but this is less restrictive the closer $\underline{m}$ is to K. The smallest firm's capacity can be no larger than for a symmetric duopoly, so a necessary (but not sufficient) condition to satisfy Assumption 1 is $\underline{m} \geq \frac{K}{2}$. We place no restriction on the level of the maximum market demand, $\overline{m}$.

Denoting $\Omega(p_i)$ as the set of firms that price below p_i and $p_{-it}^{\max} = \max\{\mathbf{p}_{-it}\}$, the proportional demand rationing rule and Assumption 1 imply that firm i's sales in period t are:

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

[6] As we discuss further in section 6, our main results simply require that enough buyers are informed of prices to remain robust.

It follows from the above that firm i's per-period profit is $\pi_{it} = p_{it}s_{it}(p_{it}, \mathbf{p}_{-it}; m_t)$, where we write $s_{it}(p_t; m_t)$ when $p_{it} = p_t$ for all i and we drop time subscripts when there is no ambiguity. Notice that firm i's per-period profit is strictly increasing in m, when it does not supply its full capacity, and that in expectation $\pi_i = p\frac{k_i}{K}\widehat{m}$ for any $p \leq 1$ for all i, so such profits are maximised for $p^m \equiv 1$.

3.3 Static Nash equilibrium

In this subsection, we derive the static Nash equilibrium that can be in pure strategies or mixed strategies. While the proof of the former is trivial, we extend the equilibrium analysis in Fonseca and Normann (2008) to our setting of demand uncertainty to solve for the latter. The mixed strategy Nash equilibrium also converges to the equilibrium analysis of Gal-Or (1984) when firms' capacity constraints are symmetric.

Lemma 1. *For any given $n \geq 2$ and $\underline{m} \geq K_{-1}$, there exists:*

i) a unique pure strategy Nash equilibrium, such that $\pi_i^N = k_i$ for all i, if $\underline{m} \geq K$,

ii) a mixed strategy Nash equilibrium, such that:

$$\pi_i^N = \begin{cases} \frac{k_i}{k_n}\left(\int_{\underline{m}}^{K}(m - K_{-n})\,g(m)dm + k_n\int_{K}^{\overline{m}} g(m)dm\right) & \text{if} \quad \underline{m} < K < \overline{m} \\ \frac{k_i}{k_n}\left(\widehat{m} - K_{-n}\right) & \text{if} \quad \overline{m} \leq K, \end{cases} \tag{2}$$

for all i.

Competition is not effective when the minimum market demand is above total capacity, $\underline{m} \geq K$, so firms set $p_i = 1$ and receive $\pi_i^N = k_i$ for all i. In contrast, when 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 $p_n = 1$, the largest firm can ensure that its expected per-period profit is at least:

$$\overline{\pi}_n \equiv \begin{cases} \int_{\underline{m}}^{K}(m - K_{-n})\,g(m)dm + k_n\int_{K}^{\overline{m}} g(m)dm & \text{if} \quad \underline{m} < K < \overline{m} \\ \widehat{m} - K_{-n} & \text{if} \quad \overline{m} \leq K. \end{cases}$$

The intuition is that the largest firm with strictly the highest price expects to supply its full capacity when the market demand exceeds total capacity, but it expects to supply the residual demand otherwise. It follows from this that the largest firm will never set a price below $\underline{p} \equiv \overline{\pi}_n/k_n$ 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 $k_i\overline{\pi}_n/k_n$. Consequently, there exists a mixed strategy Nash equilibrium where $\pi_i^N = \frac{k_i}{k_n}\overline{\pi}_n > 0$ for all i. This is equivalent to (2) and Assumption 1 is sufficient to ensure that it is positive. The lower bound of the support of the mixed strategy Nash equilibrium is $\underline{p}$. We provide a complete characterisation of the mixed strategy Nash equilibrium in the proof of Lemma 1.

4 Collusion under Private Monitoring

In this section, we first find the conditions for which firms can monitor a collusive agreement perfectly in that any deviation will be detected with certainty by all firms from their private information. Then we consider the case of imperfect monitoring in which there is some uncertainty over whether there has been a deviation. We henceforth focus on the case where $\underline{m} < K$, as collusion is unnecessary otherwise.

4.1 Perfect private monitoring

We wish to find the conditions for which all firms can perfectly monitor an agreement to sustain a collusive price p^c, given their private information, z_i^t. Given that firms can be asymmetric, the inferences that each firm can draw about their rivals' actions may differ. However, under certain conditions, all firms will be able to infer with certainty when any rival has not set p^c in period $t-1$, for any level of demand. This information then ensures that firms can follow trigger strategies, where they set p^c until they detect that all firms have not set the same price, in which case they play the static Nash equilibrium forever.[7] Following the standard terminology, we refer to this as perfect private monitoring.

To begin, suppose firm i set $p^c \leq 1$. From (1), firm i will be able to detect with certainty from its resultant sales when at least one rival has not set p^c if:

$$\frac{k_i}{K - \sum_{j\in\Omega(p^c)} k_j}\left(\overline{m} - \sum_{j\in\Omega(p^c)} k_j\right) < \frac{k_i}{K}\underline{m} \leq \frac{k_i}{K}\overline{m} < k_i,\ \forall\ j \neq i \tag{3}$$

where $\Omega(p^c)$ is nonempty in (3). The reason is that firm i's set of collusive sales, $\left[\frac{k_i}{K}\underline{m}, \frac{k_i}{K}\overline{m}\right]$, and the sales it can receive if at least one rival does not charge p^c are mutually exclusive. It

[7] As we discuss further below, trigger strategies generate the lowest critical discount factor given the proportional demand rationing rule.

follows from the left-hand side of (3) that the inferences that a firm can draw depend upon the size of its smallest rival's capacity. This is due to the fact that if it is possible for a firm to infer with certainty that its smallest rival has not set a lower price than p^c, then it follows that other larger rivals (or any set of rivals) have also not done so. Consequently, the larger firms $i > 1$ each have the same ability to infer whether all rivals set p^c, yet it is easier for firm 1 to infer whether all of its rivals set p^c compared with any other firm, if firm 1 is strictly the smallest firm, $k_1 < k_2$. The reason is that the smallest rival of the larger firms $i > 1$ is firm 1, and firm 1's smallest rival is larger than its competitiors' smallest rival when $k_1 < k_2$. Thus, a necessary condition for perfect private monitoring is that the largest firms can detect with certainty when firm 1 has not set p^c from their private information.

Proposition 1. *For any given $n \geq 2$, $\delta \in (0,1)$ and $\underline{m} \geq K_{-1}$, there exists a unique level of market demand, $\overline{m}^* (k_1) < K$, such that there is perfect private monitoring, if and only if $\overline{m} \in [\underline{m}, \overline{m}^* (k_1))$.*

There is perfect private monitoring when fluctuations in market demand are small for two reasons. First, the maximum market demand must be below total capacity, $\overline{m} < K$. Otherwise, if firm i set $p^c \leq 1$, it will supply its full capacity for any $m_t \geq K$, in which case it will be uncertain as to whether market demand is high and all of its rivals set p^c or whether market demand is low and at least one of its rivals charged more than p^c (and whether other rivals also undercut p^c). Second, the maximum market demand must be sufficiently close to the minimum, where the critical threshold, $\overline{m}^* (k_1)$, is defined by the level of $\overline{m}$ that solves $\frac{k_i}{K_{-1}} (\overline{m} - k_1) = \frac{k_i}{K} \underline{m}$, from (3). Thus, low market demand implies that each firm's lowest collusive sales will exceed the maximum sales that they will receive when its smallest rival undercuts p^c. The critical level $\overline{m}^* (k_1)$ is strictly below K, because the maximum sales of nondeviating firms equal their full capacity for any $m_t \geq K$, which cannot be strictly lower than their minimum collusive sales. Consequently, the necessary condition that guarantees that all firms know when any rival has undercut p^c is also a sufficient condition to ensure that they will know when any rival has set a price above p^c.

Given deviations are perfectly detected when fluctuations in market demand are sufficiently small, it follows that the standard results of collusion under perfect observability also apply under perfect private monitoring. In particular, the trigger strategy profile defines a subgame perfect

Nash equilibrium for $p^c = p^m$ if:

$$\delta \geq \frac{\pi_i^d - \pi_i^m}{\pi_i^d - \pi_i^N} = \frac{k_n}{K} \equiv \underline{\delta}^* (k_n), \tag{4}$$

where $\pi_i^m = \frac{k_i}{K}\widehat{m}$ is firm i's per-period collusive profits and $\pi_i^d = k_i$ is firm i's optimal deviation profits given $p^m = 1$.[8] This implies that all firms' incentives to collude are the same, despite possible asymmetries, and that increasing the capacity of the largest firm makes collusion more difficult to sustain. The intuition is that the static Nash equilibrium profits increase with the capacity of the largest firm, so the punishment is weaker. This is the same as the lowest critical discount factor in Compte *et al.* (2002), who assume that firms follow an alternative strategy in which their market shares remain constant during collusive and punishment phases, and it also coincides with the lowest possible discount factor that sustains collusion given the proportional rationing rule.[9]

In contrast to Compte *et al.* (2002), the size of the smallest firm's capacity is also important in our setting, because the critical level of market demand, $\overline{m}^* (k_1)$, is strictly increasing with the capacity of the smallest firm, k_1, holding total capacity constant. This is due to the fact that if it is just possible for a firm to infer that the smallest firm has not deviated for a given level of $\overline{m}$, then it is also possible for the same level of $\overline{m}$ when the smallest firm has more capacity. Thus, as k_1 increases, there is perfect private monitoring for a slightly higher level of $\overline{m}$, because the set of collusive sales can be wider and yet still not overlap with the set of sales for which at least one firm does not set p^c. These results can be summarised in the following corollary.

Corollary 1. *For any given $n \geq 2$ and $\underline{m} \geq K_{-1}$, the critical discount factor, $\underline{\delta}^* (k_n)$, is strictly increasing in the capacity of the largest firm, k_n, and the critical level of market demand, $\overline{m}^* (k_1)$, is strictly increasing in the capacity of the smallest firm, k_1.*

An implication of the above is that the parameter space of collusion under perfect private monitoring is greatest when firms' capacities are symmetric, $k_i = k$ for all i, holding total capacity and the number of firms constant. The reason is that the critical discount factor, $\underline{\delta}^* (k_n)$, is at

[8] It is easy to check that the critical discount factor is higher and collusive profits are lower for $p^c < 1$.

[9] The reason is that, as showed by Lambson (1994), the harshest punishment that can be inflicted on the largest firm is that it receives the stream of profits from its minimax strategy. In our setting, the per-period minimax profit of the largest firm is equivalent to its static Nash equilibrium profits. Thus, trigger strategies generate the harshest possible punishment for the largest firm, so strategies that punish firm n's rivals more severely than trigger strategies will not lower the critical discount factor below $\underline{\delta}^* (k_n)$.

its minimum, because the static Nash equilibrium profits are as low as possible, and the critical level of market demand, $\overline{m}^* (k_1)$, is at its maximum, because private monitoring is perfect for the widest range of demand fluctuations.

4.2 Imperfect private monitoring

We now wish to analyse collusion under private monitoring when market demand fluctuates to the extent that $\overline{m} \geq \overline{m}^* (k_1)$. In this case, firms can only imperfectly monitor the collusive agreement and a firm that receives low sales has to solve a non-trivial signal extraction problem: it does not know if its low sales are due to a low realisation of market demand or due to a deviation.

4.2.1 Trigger-sales strategies

We consider the following strategy profile in which there are "collusive phases" and "punishment phases". When period t is in a collusive phase, firm i sets the collusive price p^c and it privately observes its sales, which provide a signal of the prices of its rivals. The collusive phase continues into period $t+1$, if all firms' sales in period t are only consistent with all firms setting the same price. Otherwise, firms enter a punishment phase which lasts T periods. Firm i plays the static Nash equilibrium in each period of the punishment phase, regardless of its sales, and a new collusive phase follows the T periods of punishment.[10]

We refer to this strategy profile as trigger-sales strategies, because each firm effectively has a trigger level of sales, $\underline{s}_i$, where a punishment phase begins if at least one firm's sales in a collusive period fall below its trigger level, and where the collusive phase continues otherwise. More specifically, framing it in terms of collusive sales, firm i's trigger level is $\underline{s}_i \equiv s_i \left(p^c; m^*_{-1} (k_1, \overline{m})\right)$ for all i, where:

$$m^*_{-1} (k_1, \overline{m}) \equiv \frac{K}{K_{-1}} (\overline{m} - k_1),$$

and where $\overline{m} \geq \overline{m}^* (k_1)$ guarantees that $m^*_{-1} (k_1, \overline{m}) \geq \underline{m}$. This follows from (3) and it has two implications. First, if firm i's sales are greater than $\underline{s}_i$, then all firms' sales are only consistent with them all setting the same price. Second, if firm i's sales do not exceed $\underline{s}_i$, then at least one firm has received sales that are at least consistent with a deviation by the smallest firm. Again, the size of the smallest firm's capacity determines the critical threshold, because a deviation by

[10] As we discuss below, the event that triggers the punishment phase is always common knowledge. Furthermore, we also show in section 6 that this generates the same critical discount factor and collusive profits as optimal symmetric penal codes.

firm 1 provides the highest sales for the other (nondeviating) firms $i > 1$. Thus, a collusive phase will switch to a punishment phase on the equilibrium path, despite all firms setting p^c, whenever $m \in \left[\underline{m}, m^*_{-1}(k_1, \overline{m})\right]$; otherwise, the collusive phase will continue into the next period. Furthermore, any firm that deviates unilaterally will trigger a punishment phase with certainty.

The event that triggers the punishment phase is common knowledge for the following reasons. If firm i's sales do not exceed $\underline{s}_i$ in a collusive period due only to a low realisation of market demand, then its rivals' sales will also be below their trigger levels. Yet, if firm i's sales do not exceed $\underline{s}_i$ because a rival deviated, then other nondeviating rivals will also receive sales below their trigger levels, but the deviants will supply their capacities and they can infer from this that a punishment phase has been triggered if $\overline{m} < K$. Moreover, if $\overline{m} \geq K$, it is common knowledge that each firm's collusive sales will always be below their trigger levels for any level of demand. Notice that the event that triggers the punishment phase would not always be common knowledge, if the trigger levels were below $s_i\left(p^c; m^*_{-1}(k_1, \overline{m})\right)$ for all i. This is due to the fact that, if a firm unilaterally undercuts p^c, its sales provide it with no information about market demand (or the resultant sales of its rivals) that it does not already know, since it will supply its full capacity for any level of demand. Consequently, given such deviations would not be punished with certainty for any trigger level below $s_i\left(p^c; m^*_{-1}(k_1, \overline{m})\right)$, the deviant would have no way of determining whether it should be in a collusive phase or a punishment phase in the following period.[11]

In fact, we now show that the event that triggers a punishment phase is common knowledge, even when play is off the equilibrium path. To see this, consider the following history $h^t = (y_0, y_1, \ldots, y_{t-1})$ where:

$$y_\tau = \begin{cases} \overline{y} & \text{if } s_{i\tau}\left(p_{i\tau}, \mathbf{p}_{-i\tau}; m_\tau\right) > \underline{s}_i \ \forall\, i \\ \underline{y} & \text{otherwise} \end{cases} \tag{5}$$

for all $\tau = \{0, 1, \ldots, t-1\}$. It follows from (5) that, if period $t-1$ was a collusive period, then according to trigger-sales strategies period t should be a collusive phase if $y_{t-1} = \overline{y}$, but a punishment phase should begin if $y_{t-1} = \underline{y}$. Proposition 2 shows that each firm can indirectly observe h^t through its private information in period t.

[11] In contrast, trigger levels above $\underline{s}_i$ also ensure that the event that triggers the punishment phase is common knowledge. However, we show in section 6 that they raise the critical discount factor and lower collusive profits compared to the level analysed in the main analysis.

Proposition 2. *For any given $n \geq 2$, $\delta \in (0,1)$ and $\underline{m} \geq K_{-1}$, h^t is common knowledge at the beginning of period t, for all t.*

An implication of Proposition 2 is that, for each date t and for any history h^t, we can check whether trigger-sales strategies yield a Nash equilibrium from that date on. We say that collusion under imperfect private monitoring is sustainable when this is the case. Consequently, our equilibrium strategies satisfy one of the two necessary conditions for a perfect public equilibrium (PPE) (see Fudenberg and Tirole, 1994, p.188). They are not PPE strategies, however, because trigger-sales strategies are not public strategies since firms do not ignore their private information in choosing their prices. Instead, the information that our equilibrium strategies are conditioned on is common knowledge *because* firms use their private information to infer it.

There are two features of trigger-sales strategies that require further discussion, both of which are concerned with when some firms can have better information than their rivals but they do not act upon it. The first case (that does not occur on the equilibrium path) is when a firm knows for sure that there has been a deviation, because it receives sales that are not a member of its set of collusive sales, $s_i < s_i(p^c; \underline{m})$ for some i. In this case, a deviant would be unaware that its rivals know for sure that there was a deviation, because it sells its capacity regardless of the level of demand. Consequently, it is not possible to ensure coordination and punish such deviations differently, since this information is not common knowledge. The second case (that does occur on the equilibrium path if $k_2 > k_1$) is when firms $i > 1$ receive sales consistent only with a deviation by firm 1, but firm 1 knows that it has not deviated. Since firm 1's collusive sales are consistent with a rival undercutting p^c for any $m \leq \frac{K}{K_{-2}} (\overline{m} - k_2) \equiv m_1^* (k_2, \overline{m})$, this occurs for all $m \in \left(m_1^* (k_2, \overline{m}), m_{-1}^* (k_1, \overline{m})\right)$ where $m_{-1}^* (k_1, \overline{m}) > m_1^* (k_2, \overline{m})$ for any $k_2 > k_1$. Alternatively, we could assume that firm 1's trigger level is $\underline{s}_1 = s_1 (p^c; m_1^* (k_2, \overline{m}))$ and the event that triggers a punishment phase would still be common knowledge. The reason is that firm 1 would be able to infer that its rivals will enter the punishment phase for any $m \in \left(m_1^* (k_2, \overline{m}), m_{-1}^* (k_1, \overline{m})\right)$ because, given it knows for sure that a deviation has not taken place, it can infer that $s_{jt}(p^c; m_t) = \frac{k_j}{K} m_t$ for all j, so it can check whether firm $i > 1$ makes sales less than $\underline{s}_i = s_i \left(p^c; m_{-1}^* (k_1, \overline{m})\right)$.

4.2.2 Equilibria in trigger-sales strategies

We wish to find when collusion under imperfect private monitoring is sustainable. It is helpful to denote V_{iH}^c as the expected discounted profit in period t and thereafter, if period t belongs to a collusive phase. Similarly, denote V_{iL}^c as the expected discounted profit in period t and thereafter, if period t is the start of a punishment phase. Since no firms will deviate in equilibrium, it follows that $G\left(m_{-1}^*\left(k_1,\overline{m}\right)\right)$ is the probability that a collusive phase in period t will switch to a punishment phase in period $t+1$, where $G\left(m_{-1}^*\left(k_1,\overline{m}^*\left(k_1\right)\right)\right)=0$ and $G\left(K\right)=1$. Thus, it is possible to write V_{iH}^c and V_{iL}^c as:

$$\begin{aligned}
&V_{iH}^c=\pi_i^c+\delta\left[\left(1-G\left(m_{-1}^*\left(k_1,\overline{m}\right)\right)\right)V_{iH}^c+G\left(m_{-1}^*\left(k_1,\overline{m}\right)\right)V_{iL}\right]\\
&V_{iL}^c=\textstyle\sum_{\tau=0}^{T-1}\delta^\tau\pi_i^N+\delta^T V_{iH}^c.
\end{aligned}$$

The first term on the right-hand side of V_{iH}^c is the initial per-period collusive profit and the second term is the expected discounted profit conditional on whether a collusive phase continues into the next period or switches to a punishment phase. In contrast, the first term on the right-hand side of V_{iL}^c is the expected discounted profit during the punishment phase and the second term is the expected discounted profit when a new collusive phase begins.

Solving simultaneously yields:

$$V_{iH}^c=\frac{\pi_i^N}{1-\delta}+\frac{\pi_i^c-\pi_i^N}{1-\delta+G\left(m_{-1}^*\left(k_1,\overline{m}\right)\right)\delta\left(1-\delta^T\right)}$$

$$V_{iL}^c=\frac{\pi_i^N}{1-\delta}+\frac{\delta^T\left(\pi_i^c-\pi_i^N\right)}{1-\delta+G\left(m_{-1}^*\left(k_1,\overline{m}\right)\right)\delta\left(1-\delta^T\right)}.$$

It follows that $V_{iH}^c>V_{iL}^c$ for any $\pi_i^c>\pi_i^N$ and $T>0$, in which case it is more profitable to be in a collusive phase than at the start of a punishment phase.

Collusion under imperfect monitoring is sustainable if, for each date t and for any history h^t, trigger-sales strategies yield a Nash equilibrium from that date on. Clearly, equilibria in trigger-sales strategies do not exist if $\overline{m}\geq K$, because a punishment phase follows every collusive period, so it follows from Lemma 1 that there is always an incentive to deviate from any p^c. Consequently, we must consider the region of $\overline{m}^*\left(k_1\right)\leq\overline{m}<K$. Since firms play the static Nash equilibrium during each period of the punishment phase, it is clear that they have no incentive to deviate in such periods. Thus, we only need to consider deviations during a collusive phase.

A deviation in a collusive phase results in a one period gain in profits, followed by a definite T period punishment phase during which firms get π_i^N. After this, firms then return to a collusive phase. Therefore, firm i's expected discounted value of deviation profits is:

$$V_{iH}^d=\pi_i^d+\sum_{\tau=1}^{T}\delta^\tau\pi_i^N+\delta^{T+1}V_{iH}^c,$$

where $\pi_i^d = p^c k_i$ if firm i marginally undercuts the collusive price, p^c. Firm i will not deviate in a collusive phase if $V_{iH}^c \geq V_{iH}^d$. This can be rewritten as:

$$\left(1-\delta^{T+1}\right) V_{iH}^c \geq \pi_i^d + \sum_{\tau=1}^{T} \delta^\tau \pi_i^N. \tag{6}$$

Proposition 3 provides the lowest discount factor for which collusion under imperfect private monitoring is sustainable, which occurs when $p^c = p^m$.

Proposition 3. *For any given $n \geq 2$ and $\underline{m} \geq K_{-1}$, there exists a unique discount factor, $\delta^* (k_1, k_n) \geq \underline{\delta}^* (k_n)$, a unique length of the punishment phase, $T^* (k_1, k_n) > 0$, and a unique level of market demand, $\overline{m}' (k_1, k_n) \in [\overline{m}^* (k_1), K)$, such that collusion under imperfect private monitoring is sustainable for any $T \geq T^* (k_1, k_n)$ and $\overline{m} \in [\overline{m}^* (k_1), \overline{m}' (k_1, k_n))$, if $\delta \geq \delta^* (k_1, k_n)$.*

Similar to Tirole (1988), there are three necessary conditions that must be satisfied so that collusion under imperfect private monitoring is sustainable. The first two are standard for theories of collusion: firms must be sufficiently patient and the length of the punishment phase must last a sufficient number of periods. The critical length of the punishment phase is implicitly defined by the level of T where the incentive compatibility constraint (6) holds with equality. However, firms must also be sufficiently patient because otherwise even a punishment phase that lasts an infinite number of periods is insufficient to outweigh the short-term benefit from deviating. More specifically, firms are sufficiently patient if:

$$\delta \geq \delta^* (k_1, k_n) \equiv \frac{1}{1 - G\left(m_{-1}^* (k_1, \overline{m})\right)} \frac{k_n}{K}, \tag{7}$$

where $T^* (k_1, k_n) < \infty$ for any $\delta > \delta^* (k_1, k_n)$, but $T^* (k_1, k_n) \to \infty$ otherwise.

The final necessary condition is that the maximum market demand must be sufficiently low. The intuition is that an increase in the maximum market demand relative to the capacity of the smallest firm raises the probability that firms will enter a punishment phase following a collusive period, $G\left(m_{-1}^* (k_1, \overline{m})\right)$. This lowers the right-hand side of (6), which tightens the incentive compatibility constraint. When $G\left(m_{-1}^* (k_1, \overline{m})\right)$ is too high, the incentive compatibility constraint cannot hold for any δ and T, so the level of $\overline{m}$ that sets (7) equal to 1 implicitly defines the critical threshold, $\overline{m}' (k_1, k_n)$. Furthermore, note that the critical discount factor, $\delta^* (k_1, k_n)$, converges to the critical level under perfect private monitoring, $\underline{\delta}^* (k_n)$, when $\overline{m} = \overline{m}^* (k_1)$, but it is strictly above it for any $\overline{m} > \overline{m}^* (k_1)$, since $\delta^* (k_1, k_n)$ is strictly increasing in $\overline{m}$. These results, and the results of corrollary 1, are depicted in Figure 1.

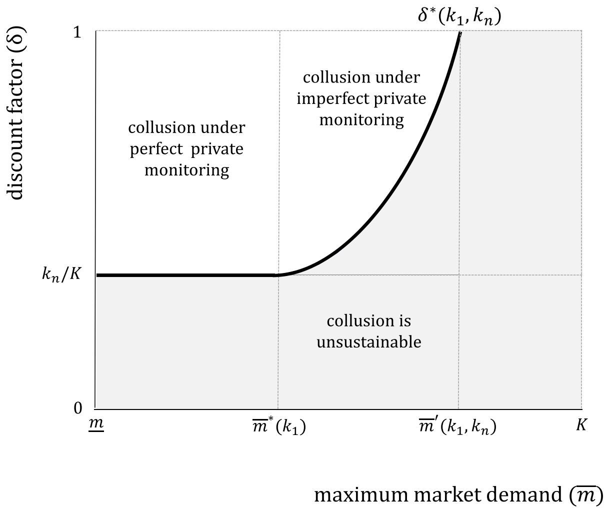

Figure 1: parameter space of collusion

4.2.3 Capacity allocation and the sustainability of collusion

We now analyse the effects of reallocating capacity among the firms on the critical discount factor and on the critical length of the punishment phase for given demand fluctuations. In particular, Proposition 4 considers an increase in the capacity of a given firm when total capacity is held constant, so such an increase may require capacity to be reallocated from a rival. For example, increasing the size of the smallest firm in a duopoly will mean that capacity of the largest firm decreases. In general, when the capacity of firm j changes by a small amount, the capacities of the other firms can change to the extent that $\frac{\partial k_i}{\partial k_j} \in [-1, 0]$ for all $i \neq j$, where $\sum_{i \neq j} \frac{\partial k_i}{\partial k_j} = -1$. Implications for mergers are drawn below.

Proposition 4. *For any given $n \geq 2$, $\underline{m} \geq K_{-1}$ and $\overline{m} \in [\overline{m}^* (k_1), \overline{m}' (k_1, k_n))$, raising the capacity of the smallest firm, k_1, strictly decreases $\delta^* (k_1, k_n)$ and $T^* (k_1, k_n)$, and strictly increases $\overline{m}' (k_1, k_n)$. In contrast, raising the capacity of the largest firm, k_n, strictly increases $\delta^* (k_1, k_n)$ and $T^* (k_1, k_n)$, and strictly decreases $\overline{m}' (k_1, k_n)$.*

As the size of the smallest firm's capacity increases, the incentive compatibility constraint slackens, because the left-hand side of (6) rises. This is due to the fact that sales of each of the smallest firm's rivals are less likely to be consistent with a deviation by the smallest firm, so firms expect to enter the punishment phase less often. This implies that collusion is easier to support for any T-period punishment, so both $\delta^*(k_1, k_n)$ and $T^*(k_1, k_n)$ fall and $\overline{m}'(k_1, k_n)$ rises. In contrast, increasing the capacity of the largest firm tightens the incentive compatibility constraint, because it weakens the punishment as the static Nash equilibrium profits increase for each firm. This tightens the incentive compatibility constraint because, although it increases both the left- and the right-hand side of (6), it increases the latter at a faster rater than the former, so both $\delta^*(k_1, k_n)$ and $T^*(k_1, k_n)$ rise and $\overline{m}'(k_1, k_n)$ falls. Both of these effects are present and work in the same direct when an increase in the capacity of the smallest firm implies a reallocation of capacity from the largest firm (and vice versa). Thus, similar to our results for perfect private monitoring, the parameter space of collusion under imperfect private monitoring is largest when firms' capacities are symmetric, $k_i = K/n$ for all i.

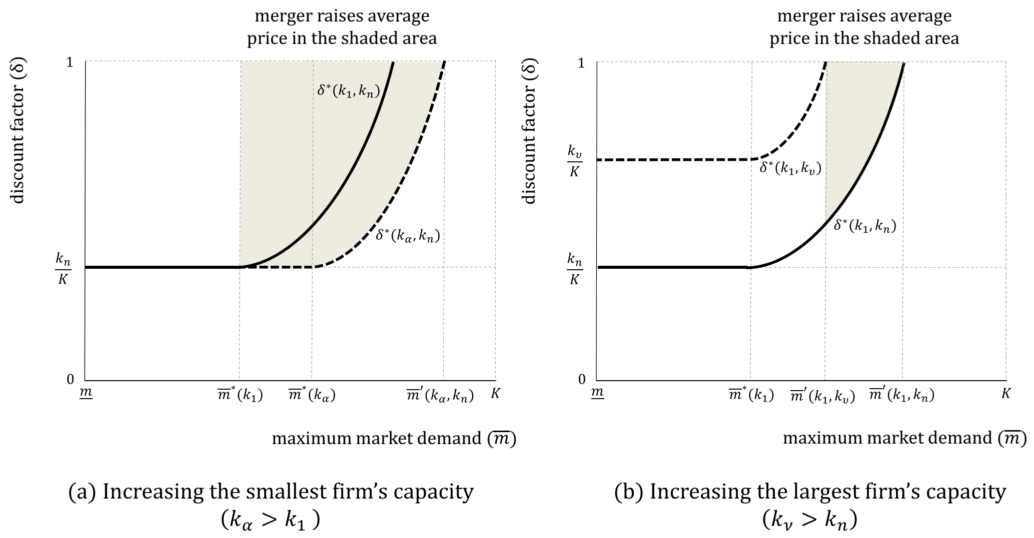

Figure 2: The effects of mergers on the parameter space of collusion

An implication of Proposition 4 is that a merger that only increases the size of the smallest firm will facilitate collusion by expanding the parameter space of collusion and a merger that only increases the size of the largest firm will destabilise collusion by reducing it. Both of these cases are illustrated in Figure 2. The result that collusion is more difficult to support as the capacity of the largest firm increases is consistent with Compte *et al.* (2002), but the effect on collusion due to the size of the capacity of the smallest firm is not. In contrast, both results are consistent with the findings of Vasconcelos (2005), but the underlying incentives for his results are very different to ours as they rely on capacities affecting marginal costs. However, as we stress in the next section, it is inappropriate to focus only on the parameter space of collusion in our setting, because the collusive profits under imperfect monitoring may also depend upon the capacity allocation.

4.3 Competitive effects of mergers

In this section, we analyse the effects of mergers on prices. Our focus is on prices as opposed to profits or welfare because they can be quickly translated into a subset of firms' profits and consumer surplus, where the latter is important because ensuring that mergers do not reduce consumer surplus is commonly perceived to be the main objective of merger control (see Lyons, 2002).[12] Following the terminology of Farrell and Shapiro (1990), we henceforth refer to the firms that merge as insiders and those not involved in the merger as outsiders. We say that a merger is privately optimal if the sum of insiders' profits post-merger is *strictly* greater than the sum of their profits pre-merger.

4.3.1 Coordinated effects of mergers

We want to understand which mergers can facilitate collusion by raising prices post-merger on average. To do so, we focus on the average price associated with the best collusive equilibrium, which we refer to as the best average price. We first investigate the effect of changing the capacity allocation on the best average price. Then we draw implications for mergers below. When private monitoring is perfect, the best average price is independent of the capacity allocation when firms are sufficiently patient, because firms set p^m in each period. Thus, we must consider when private monitoring is imperfect. The best average price for this case can be calculated by evaluating V^c_{iH} at $T^*(k_1, k_n)$, so that collusive profits are maximised subject to the constraint that collusion is sustainable, and noting that $V^c_{iH} = \frac{1}{1-\delta}\widehat{p}^c(k_1, \overline{m})\frac{k_i}{K}\widehat{m}$, where $\widehat{p}^c(k_1, \overline{m})$ denotes the best average price under imperfect private monitoring.[13] For comparison, it is helpful to note to that the average price of the static Nash equilibrium is $\widehat{p}^N(k_n, \widehat{m}) \equiv \frac{K}{k_n}\frac{(\widehat{m}-K_{-n})}{\widehat{m}}$ for all $\overline{m} < K$.

[12]Total welfare is independent of the capacity allocation, because market demand is perfectly inelastic.

[13]Although $T^*(k_1, k_n)$ may not be an integer, the expected length of the punishment phase could equal this length if there were some optimally set randomisation device that varied the length of punishment phases. Moreover, the same average price can be generated from optimal symmetric penal codes where the punishment may last for only one period, as we show in section 6.

Proposition 5. *For any given $n \geq 2$, $\underline{m} \geq K_{-1}$, $\overline{m} \in (\overline{m}^* (k_1), \overline{m}' (k_1, k_n))$ and $\delta \geq \delta^* (k_1, k_n)$, the best average price $\widehat{p}^c (k_1, \overline{m})$ satisfies $\widehat{p}^N (k_n, \widehat{m}) < \widehat{p}^c (k_1, \overline{m}) < p^m$. It is strictly increasing in the capacity of the smallest firm, k_1, and is independent of the capacities of all other firms, including the largest, k_n.*

The best average price is increasing in the size of the smallest firm's capacity for two reasons. First, as the capacity of the smallest firm increases, its rivals' collusive sales are less likely to be consistent with a deviation by the smallest firm, so profits rise on the equilibrium path because collusive phases are less likely to switch to punishment phases than before. Second, the increase in profits also raises the left-hand side of (6), which slackens the incentive compatibility constraint, and as a consequence the optimal punishment length shortens. Both effects imply that firms expect there to be more collusive periods on the equilibrium path than when the smallest firm has less capacity, so the best average price rises.

In contrast, the best average price is independent of the size of the largest firm's capacity, because there are two offsetting effects that perfectly cancel each other out. The first effect is that an increase in the capacity of the largest firm raises profits on the equilibrium path, because the static Nash equilibrium profits of each firm are greater than before. However, it also tightens the incentive compatibility constraint because, although it increases both the left- and the right-hand side of (6), it increases the latter at a faster rater than the former. Consequently, the second effect is that the length of the punishment phase has to increase to ensure that collusion is sustainable, and this decreases profits on the equilibrium path to the extent that the size of the largest firm has no effect on the best average price.

This implies that a merger that includes the smallest firm pre-merger as one of the insiders is the only merger that can facilitate collusion by raising the best average price. The complete parameter space for which such a merger raises the average price is illustrated in Figure 2(a). This follows since $\widehat{p}^c (k_1, \overline{m})$ is equal to p^m when $\overline{m} = \overline{m}^* (k_1)$, it converges on $\widehat{p}^N (k_n, \widehat{m})$ as $\overline{m} \to \overline{m}' (k_1, k_n)$, and it is also between these two levels over this range. Furthermore, a merger that raises the best average price is privately optimal and it also strictly increases the profits of the outsiders. This is because it is possible to write the present discounted value of profits for any set of firms μ as $\sum_{i \in \mu} \frac{k_i}{K} \widehat{m} \left(\frac{\widehat{p}^c (k_1, \overline{m})}{1-\delta} \right)$, which is only strictly higher post-merger if k_1 is increased. As a consequence, such a merger will also lower consumer surplus, since the present discounted value of consumer surplus is $\frac{\widehat{m}}{1-\delta} (1 - \widehat{p}^c (k_1, \overline{m}))$. Finally, the fact that $\widehat{p}^c (k_1, \overline{m})$ is never a function of k_n implies that it is higher, for a given $\overline{m}$, when firms are symmetric,

$k_i = K/n$ for all i, and when there are fewer firms in the market. For example, $\widehat{p}^c(k_1, \overline{m})$ is highest for a symmetric duopoly, but it is lower for a symmetric triopoly and it is even lower for an asymmetric triopoly, where $k_1 < K/3$.[14]

4.3.2 Comparing coordinated effects with unilateral effects

Given collusion may not enable firms to set p^m in every period, we want to understand when a merger that destabilises collusion will lead to higher prices on average than under collusion. So, first we compare the best average price with the static Nash equilibrium average price for two allocations, which we denote (k_1, k_n) and (k_α, k_ν) where k_α and k_ν are the capacities of the smallest and largest firms for an alternative allocation, respectively. Then we draw implications for mergers below. When the alternative allocation comes from a merger, it follows that there will be $\nu < n$ firms post-merger, where $k_\alpha \geq k_1$ and $k_\nu \geq k_n$.

Proposition 6. *For any given $n \geq 2$, $\underline{m} \geq K_{-1}$, $\overline{m} < \overline{m}'(k_1, k_n)$ and $\delta \geq \delta^*(k_1, k_n)$, the best average price of (k_1, k_n) is less than the static Nash equilibrium average price of (k_α, k_ν), $\widehat{p}^c(k_1, \overline{m}) < \widehat{p}^N(k_\nu, \widehat{m})$, if $k_\nu > k_n$ and $\overline{m} \geq \overline{m}'(k_1, k_\nu)$, where $\overline{m}'(k_1, k_\nu) > \overline{m}^*(k_1)$.*

The intuition is that punishment phases are expected to occur more often on the equilibrium path as the level of $\overline{m}$ increases, because firms are more likely to receive sales that are consistent with a deviation by the smallest firm. Thus, the best average price falls towards the average price of the static Nash equilibrium as $\overline{m}$ increases. Yet, the average price of the static Nash equilibrium is strictly increasing in the size of the largest firm's capacity, so an alternative allocation where the largest firm has more capacity will have a higher $\widehat{p}^N(k_\nu, \widehat{m})$ than $\widehat{p}^c(k_1, \overline{m})$ when $\overline{m}$ is sufficiently large. Notice that, as illustrated in Figure 2(b), the maximum market demand must be so large that collusion under imperfect monitoring will not be sustainable for the alternative allocation, when the size of the smallest firm is unchanged, $k_1 = k_\alpha$.

This has two implications for merger policy. First, a merger that destabilises collusion by increasing the size of the largest firm may not actually decrease prices post-merger. In fact, our model suggests that it is not in the insiders' interests to propose the merger if it destabilises

[14] This also implies that larger firms can increase their profits by divesting capacity to the smallest firm, so that collusion is facilitated. This is not unheard of in actual merger cases, because there are examples where insiders' capacity or market shares fall as a result of a divestment remedy (see Compte *et al.*, 2002; and Davies and Olczak, 2010).

collusion and decreases average prices. This follows since such a merger is privately optimal for any set of μ firms if $\sum_{i\in\mu}\frac{k_i}{K}\widehat{m}\left(\frac{\widehat{p}^N(k_\nu,\widehat{m})}{1-\delta}\right) > \sum_{i\in\mu}\frac{k_i}{K}\widehat{m}\left(\frac{\widehat{p}^c(k_1,\overline{m})}{1-\delta}\right)$, so the condition that ensures the insiders' profits increase post-merger also ensures that the average price rises post-merger. Second, when presented with the possibility of two merger outcomes, where one leads to a more asymmetric allocation than the other, it is not always the case that the most asymmetric allocation should be preferred, even if the other allocation facilitates collusion. Instead, it is important to consider the likelihood to which price wars will occur over time for the collusive allocation and this should be compared against the effect of stregthening one firm's market power unilaterally.

5 An Example

We complement our general results by analysing an example. We do this for two reasons. First, we wish to show that symmetric collusive capacity allocations can have substantially lower average prices than noncollusive capacity allocations. Second, we wish to illustrate how the unilateral and coordinated effects of mergers can be simulated in our framework. This is relevant for a recently emerging literature that aims to extend the merger simulation methodology to coordinated effects (Sabbatini, 2006; Hikisch, 2008; and Davis and Huse, 2010).

We consider an example in which total industry capacity is $K = 100$ and assume that this is divisible into 6 equal sized parts. We suppose there is an asymmetric triopoly pre-merger, in which firm 1 has 1/6 of this capacity, firm 2 has 2/6 and firm 3 has 3/6, and we denote this capacity allocation as (1/6, 2/6, 3/6). We then consider three alternative merger scenarios. The first is a merger between firms 1 and 2 that would create a symmetric duopoly, (3/6, 3/6). The second is a merger between firm 1 and 3 that would create an asymmetric duopoly, (2/6, 4/6). The final alternative is a merger between firms 2 and 3 that would create a very asymmetric duopoly, (1/6, 5/6).[15]

We want to analyse the effects of such mergers on consumer surplus. We focus on the expected consumer surplus per unit sold per-period of the most profitable equilibrium, $CS(\widehat{p}) \equiv 1-\widehat{p}$, which is simply the difference between the monopoly price and the average price, $\widehat{p}$.[16] The preceding analysis implies that $\widehat{p}$ is the best average price when collusion is sustainable,

[15] Alternatively, (3/6, 3/6) and (2/6, 4/6) could result from a remedy of the merger between 2 and 3, in which capacity of the merged entity is divested to firm 1.

[16] Multiplying $CS(\widehat{p})$ by $\widehat{m}$ gives the expected consumer surplus per-period, then dividing this by $1-\delta$ gives the present discounted value of consumer surplus, assuming buyers have the same discount factor as firms.

otherwise it is the static Nash equilibrium average price. Figure 3 plots $CS(\widehat{p})$ as a function of $\Delta m \equiv \frac{\overline{m}-\underline{m}}{\widehat{m}}$ for the various scenarios under the assumption that demand is drawn from a uniform distribution, where $G(m) = \frac{m-\underline{m}}{\overline{m}-\underline{m}}$. Parameter values are chosen such that $\widehat{m} = 92$ for all Δm and that $K_{-1} \leq \frac{5}{6}(100) \leq \underline{m} \leq \overline{m} \leq K = 100$. We let $\delta \rightarrow 1$ so collusion is sustainable for all $\overline{m} \in [\underline{m}, \overline{m}'(k_1, k_n))$.[17] Finally, the analysis above implies that each merger is privately optimal whenever $CS(\widehat{p})$ is strictly lower post-merger than pre-merger.

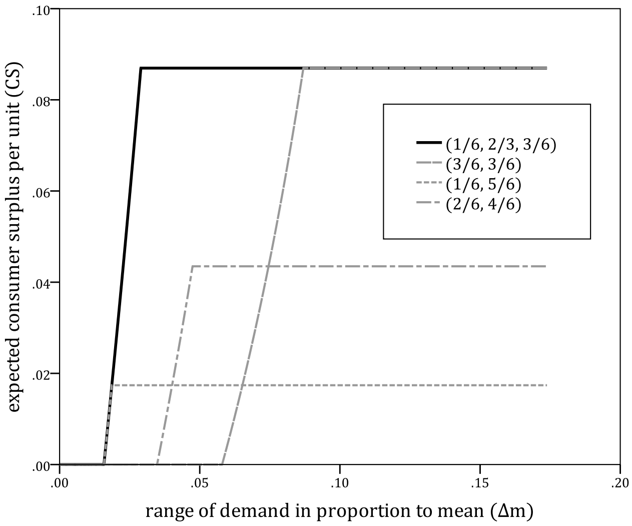

Figure 3: $G(m) = \frac{m-\underline{m}}{\overline{m}-\underline{m}}$, $\widehat{m} = 92$, $K = 100 > \frac{5}{6}(100) \geq K_{-1}$, and $\delta \rightarrow 1$

Each of the plotted lines in Figure 3 has a similar shape. For low levels of Δm, where $\overline{m} \leq \overline{m}^*(k_1)$, private monitoring is perfect so $CS(\widehat{p}) = 1 - p^m = 0$. When Δm is in an intermediate range, such that $\overline{m}^*(k_1) < \overline{m} \leq \overline{m}'(k_1, k_n)$, private monitoring is imperfect and $CS(\widehat{p}) = 1 - \widehat{p}^c(k_1, \overline{m})$ is upward-sloping because $\widehat{p}^c(k_1, \overline{m})$ is strictly decreasing in $\overline{m}$. For high levels of Δm, where $\overline{m} > \overline{m}'(k_1, k_n)$, $CS(\widehat{p}) = 1 - \widehat{p}^N(k_n, \widehat{m})$ is constant since $\widehat{m}$ is held fixed. Note that comparing (1/6, 2/6, 3/6) and (3/6, 3/6) in Figure 3 is consistent with moving horizontally from left to right on Figure 2(a) for $\delta \rightarrow 1$, because only the smallest firm's capacity is increased post-merger. Hence, collusion under perfect and imperfect private monitoring would be sustainable for a wider range of Δm post-merger. Likewise, comparing (1/6, 2/6, 3/6) and (1/6, 5/6) can be thought of as moving horizontally from left to right on Figure

[17]For $\delta < 1$, the only difference is that there is a discontinuity in $CS(\widehat{p})$ at the point where collusion under imperfect monitoring becomes unsustainable, so the line jumps to the $CS(\widehat{p})$ associated with the static Nash equilibrium for a lower level of Δm.

2(b) for $\delta \to 1$, because only the largest firm's capacity is increased post-merger. Consequently, private monitoring becomes imperfect at the same level of Δm but collusion under imperfect monitoring is sustainable for a (slightly) wider range pre-merger. For (2/6, 4/6), both the capacities of the smallest and the largest firms have increased compared to (1/6, 2/6, 3/6).

Figure 3 shows that all merger scenarios lower $CS(\widehat{p})$ for some levels of Δm. Of particular interest is that it shows that a merger that creates a more asymmetric allocation (1/6, 5/6) than pre-merger (1/6, 2/6, 3/6) can reduce consumer surplus, even if it destabilises collusion. This occurs at approximately $\Delta m = 0.025$ and the effect can be considerable: $CS(\widehat{p})$ under (1/6, 2/6, 3/6) is over four times higher than under (1/6, 5/6) for values of Δm close to 0.025, even though collusion is still sustainable for the former but not the latter. Furthermore, Figure 3 also shows that, when demand fluctuations are sufficiently large, a merger that facilitates collusion by creating a symmetric duopoly reduces consumer surplus less than a merger that creates a more asymmetric duopoly. This occurs at around $\Delta m = 0.075$ when collusion is sustainable for (3/6, 3/6) but it is not for the asymmetric duopolies. Again, the difference in $CS(\widehat{p})$ can be substantial: when Δm is approximately 0.09, it is over four times higher for (3/6, 3/6) than for (1/6, 5/6) and two times higher for (3/6, 3/6) than for (2/4, 4/6).

6 Robustness

In this section, we explore the robustness of our results. First, we show that optimal symmetric penal codes generate the same critical discount factor and best average prices as in our main analysis. Second, we show that higher trigger levels raise the critical discount factor and lower the best average price. Third, we consider the implications of downward sloping demand. Finally, we discuss and relax the assumption that each firm supplies all units at the same price.

6.1 Penal codes

In the main analysis, we assumed that firms play the static Nash equilibrium for a fixed number of periods during any punishment phase. In many models of collusion, such strategies are inferior to penal codes where the punishment can be harsher than the static Nash equilibrium (see Abreu *et al.*, 1986). However, similar to under perfect private monitoring, here we show that the critical discount factor and the best average price under optimal symmetric penal codes are the same as in our main analysis for imperfect private monitoring. The strategies considered in this section are equivalent to the strategies that generate the lowest possible discount factor in Compte *et* *al.* (2002).

Consider the following strongly symmetric strategy profile, in which firms behave identically after all histories. When firms are in a collusive phase in period t, firm i sets the collusive price p^c and it privately observes its sales. The collusive phase continues into period $t+1$, if all firms' sales in period t are only consistent with them all setting the same price. Otherwise, the firms enter a punishment phase in which firm i sets p^p in period $t+1$. If each firm's sales in period $t+1$ are only consistent with them all setting the same price, then a new collusive phase begins in period $t+2$. Otherwise, the punishment phase continues into period $t+2$, and the process is repeated until a period in which all firms' sales are only consistent with all firms having set the same price.[18]

In this case, V_{iH}^c and V_{iL}^c are as follows:

$$\begin{aligned} V_{iH}^c &= \pi_i^c + \delta\left[\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right) V_{iH}^c + G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right) V_{iL}^c\right] \\ V_{iL}^c &= \pi_i^p + \delta\left[\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right) V_{iH}^c + G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right) V_{iL}^c\right], \end{aligned}$$

where the first terms on the right-hand side are the expected per-period profit of the initial period, and the second terms are the expected discounted profit conditional on whether the next period is in a collusive or punishment phase. Solving simultaneously yields:

$$V_{iH}^c = \frac{\pi_i^c - \delta G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\left(\pi_i^c - \pi_i^p\right)}{1-\delta} \quad \text{and} \quad V_{iL}^c = \pi_i^p + \frac{\delta\left(\pi_i^c - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\left(\pi_i^c - \pi_i^p\right)\right)}{1-\delta}.$$

For this strategy profile, we must consider each firm's incentive to deviate in a collusive phase and in a punishment phase. Firm i will not deviate in a collusive phase if $V_{iH}^c \geq \pi_i^d + \delta V_{iL}^c$, and it will not deviate in a punishment phase if $V_{iL}^c \geq \pi_i^{dp} + \delta V_{iL}^c$, where π_i^{dp} denotes the per-period profit firm i receives from its optimal deviation in a punishment phase. Substituting V_{iH}^c and V_{iL}^c into these two incentive compatibility constraints and rearranging shows that firm i will not deviate in a collusive phase or in a punishment phase if:

$$\pi_i^c - \frac{\left(\pi_i^d - \pi_i^c\right)}{\delta\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right)} \geq \pi_i^p \geq \frac{\pi_i^{dp} - \delta\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right)\pi_i^c}{1 - \delta\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right)}. \tag{8}$$

This implies that a necessary condition for collusion to be sustainable is:

$$\delta \geq \frac{1}{\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right)}\left(\frac{\pi_i^d - \pi_i^c}{\pi_i^d - \pi_i^{dp}}\right) \forall\, i. \tag{9}$$

Otherwise, both incentive compatibility constraints cannot hold.

[18] Proposition 2 ensures that the event that triggers a punishment phase is still common knowledge for any h^t.

Since both V_{iH}^c and V_{iL}^c are increasing in π_i^p, the firms have an incentive to set π_i^p as high as possible, subject to collusion being sustainable. Thus, it is optimal for π_i^p to equal to the left-hand side of (8), which is the case for $p^c = p^m$ if:

$$p^p = 1 - \frac{1}{\delta\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right)}\left(\frac{K - \widehat{m}}{\widehat{m}}\right). \tag{10}$$

Substituting (10) into V_{iH}^c gives:

$$V_{iH}^c = \frac{1}{(1-\delta)}\widehat{p}^c\left(k_1, \overline{m}\right)\frac{k_i}{K}\widehat{m} \qquad \text{where} \qquad \widehat{p}^c\left(k_1, \overline{m}\right) = \frac{\widehat{m} - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)K}{\widehat{m}\left(1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\right)},$$

which is the same as in our main analysis (see Proposition 5). Finally, to solve for the critical discount factor, we must find the expression for π_i^{dp}. The optimal deviation for firm i in a punishment phase will either be to set the monopoly price and supply the residual demand, $\widehat{m} - K_{-i}$, or to undercut p^p marginally to supply its capacity. Comparing the profits for these two options yields:

$$\pi_i^{dp} = \begin{cases} k_i\left[1 - \frac{1}{\delta\left(1-G\left(m_{-1}^*\right)\right)}\left(\frac{K-\widehat{m}}{\widehat{m}}\right)\right] > 0 & \text{if } \delta \geq \frac{1}{\left(1-G\left(m_{-1}^*\left(k_1,\overline{m}\right)\right)\right)}\frac{k_i}{\widehat{m}} \\ \widehat{m} - K_{-i} & \text{otherwise.} \end{cases}$$

Substituting π_i^{dp} into (9) shows that collusion is sustainable if $\delta \geq \frac{1}{\left(1-G\left(m_{-1}^*\left(k_1,\overline{m}\right)\right)\right)}\frac{k_i}{K}$ for all i. This implies that, compared with trigger-sales strategies, optimal symmetric penal codes make collusion easier to sustain for all firms, except the largest. Moreover, the largest firm's incentives to collude are the same under these two strategies, so the critical discount factor is same as for our main analysis (see Proposition 3).[19]

6.2 Trigger-sales levels

Imperfect private monitoring in the main analysis is based on the assumption that firms enter a punishment phase if at least one firm's sales in a collusive period fall below its trigger level $\underline{s}_i \equiv s_i\left(p^c; m_{-1}^*\left(k_1, \overline{m}\right)\right)$ for all i. While this is the lowest trigger level that ensures that the event that begins a punishment phase is common knowledge, it is not unique because higher trigger levels would also achieve the same result. We show below that higher trigger levels raise the critical discount factor and lower collusive profits, so firms would not have incentive to

[19] A secondary implication of this analysis is that our results are robust if only $K_{-1} < \underline{m}$ buyers are informed of prices in each period, provided buyers are still allocated according to the proportional rationing rule. The reason is that each firm's incentive to deviate in a collusive phase and a punishment phase will be unaffected by the proportion of buyers who are unaware of the secret price cut.

implement higher than necessary trigger levels. Before doing so, we note that higher trigger sales are also less natural candidates in the real world, because they imply that firms should enter a punishment phase even for sales that are not consistent with a deviation.

Suppose that each firm now has a trigger level of $\underline{s}_i \equiv s_i(p^c; m^*)$ for all i, for some $m^* > m^*_{-1}(k_1, \overline{m})$, which implies that firms will enter the punishment phase more often than in the main analysis. Since sales above $s_i\left(p^c; m^*_{-1}(k_1, \overline{m})\right)$ are only consistent with collusive play by all firms, it follows that the event that triggers the punishment phase is still common knowledge: if a firm's sales are below its trigger level, then so are its rivals. Following the steps in the main analysis (or in the subsection above), it is possible to find that collusion under imperfect monitoring in this case is sustainable if:

$$\delta \geq \frac{1}{1 - G(m^*)} \frac{k_n}{K} \qquad \text{in which case} \qquad \widehat{p}^c(k_1, m^*) = \frac{\widehat{m} - G(m^*) K}{\widehat{m}(1 - G(m^*))}.$$

Notice that, in contrast to the main analysis, both are independent of the smallest firm's capacity. However, differentiating both with respect to m^* shows that the critical discount factor is strictly increasing in m^* and that the best average price is strictly decreasing in m^*. Thus, compared with the trigger level analysed in the main analysis, firms need to be more patient to sustain collusion under imperfect private monitoring and profits are lower. This implies that firms have do not have an incentive to set their trigger level above that in the main analysis.

6.3 Downward-sloping demand

Throughout the paper, we have assumed, consistent with Compte *et al.* (2002), that market demand is always perfectly inelastic. While it is often considered that collusion is more likely to occur in industries where demand is inelastic, the main reason for this assumption is that it ensures that the model is tractable. To address this limitation, we sketch some results that can be generated from our model when demand is downward sloping for a restricted parameter space where the model is not completely intractable. For simplicity, we also restrict the analysis to the equilibrium path.[20] Contrary to the main analysis, we show that monitoring is easier for a higher collusive price when demand is downward sloping, but firms do not have an incentive to raise the price to the point where private monitoring is perfect, when this requires setting a price above the monopoly level. The latter implies that imperfect private monitoring in the best collusive equilibrium is not restricted to the special case of perfectly inelastic demand.

[20] A complete proof is available from the authors upon request.

Let market demand in period t be denoted by $D(p_t; m_t)$, where demand is downward sloping, $\frac{\partial D}{\partial p} < 0$. Similar to the main analysis, let market demand shift with m_t where $\frac{\partial D}{\partial m} > 0$ and where, as before, m_t is independently drawn from a distribution $G(m)$, with mean $\widehat{m}$ and density $g(m) > 0$ on the interval $[\underline{m}, \overline{m}]$. For simplicity, assume that $\frac{\partial D}{\partial p}$ is independent of m_t. Suppose there exists a choke price, $P(m_t)$, for all $m_t \in [\underline{m}, \overline{m}]$, such that $D(p_t; m_t) = 0$ for any $p_t \geq P(m_t)$ and $D(p_t; m_t) > 0$ for any $p_t \in [0, P(m_t))$. Finally, assume that the expected industry profit function, $pD(p; \widehat{m})$, is strictly concave, so there exists a unique monopoly price, p^m, that satisfies $\arg\max_p \{pD(p; \widehat{m})\}$. To keep the analysis as simple as possible, we make the following assumptions on the parameter space:

Assumption 2. $K_{-1} \leq D(\overline{p}; \underline{m})$ for some $\overline{p} \in [p^m, P(\underline{m}))$

Assumption 3. $D(0; \overline{m}) < K$

Assumption 2 is the downward-sloping demand equivalent to Assumption 1, as it ensures that even the highest-priced firm will always receive positive sales, provided its price is below $\overline{p}$. Assumption 3 ensures that the market demand is always supplied, and this is equivalent to the necessary condition for collusion to be sustainable under inelastic demand, $\overline{m} < K$.

Similar to (3), each firm's collusive sales under the proportional allocation rule and the set of sales it can receive when at least one rival deviates are mutually exclusive if:

$$\frac{k_i}{K_{-1}}\left(D(p^c; \overline{m}) - k_1\right) < \frac{k_i}{K} D(p^c; \underline{m}) \leq \frac{k_i}{K} D(p^c; \overline{m}) < k_i.$$

Consequently, there is perfect private monitoring if and only if $m_t \in [\underline{m}, \overline{m}^*(k_1, p^c))$, where $\overline{m}^*(k_1, p^c)$ is the level of $\overline{m}$ that satisfies:

$$\frac{1}{K_{-1}}\left(D(p^c; \overline{m}) - k_1\right) - \frac{1}{K} D(p^c; \underline{m}) = 0. \tag{11}$$

Notice that $\frac{\partial D}{\partial \overline{m}} > 0$ guarantees that $\overline{m}^*(k_1, p^c)$ is unique, and that Assumption 3 ensures that $\overline{m}^*(k_1, p^c) \in (\underline{m}, K)$. Using the implicit function theorem, it is possible to check that $\overline{m}^*(k_1, p^c)$ is strictly decreasing in the capacity of the smallest firm, k_1, and strictly increasing in the collusive price, p^c. The former is consistent with the main analysis, but the latter is not. The intuition of the latter is that raising the collusive price reduces the market demand, which makes it easier for firms to monitor the smallest firm. This is because a deviation by the smallest firm would now supply relatively more of the market demand, so the resultant sales of each of its rivals will be further from their set of collusive sales.

An implication of this is that firms may want to set the collusive price above the monopoly level to make monitoring easier, when colluding on the monopoly price requires imperfect private monitoring. However, this involves a tradeoff as raising the collusive price above the monopoly level lowers the per-period collusive profits. We wish to demonstrate that firms do not have an incentive to set their price at a level that ensures that private monitoring is perfect. Thus, let p^* denote the level of p^c that satisfies $\overline{m} = \overline{m}^*(k_1, p^c)$ and suppose that $p^* \in (p^m, \overline{p}]$. This implies that there is perfect private monitoring if firms collude on p^*, but there will be imperfect monitoring for any collusive price below p^*, including p^m. Following the steps of section 6.1, it is possible to show that firm i's discounted collusive profits when demand is downward sloping are:

$$V_{iH}^c = \frac{\pi_i^c - G\left(m_{-1}^*(k_1, \overline{m}, p^c)\right)\pi_i^d}{(1-\delta)\left(1 - G\left(m_{-1}^*(k_1, \overline{m}, p^c)\right)\right)} \;\forall\, i$$

where $m_{-1}^*(k_1, \overline{m}, p^c)$ is the level of m_t that satisfies:

$$\frac{1}{K_{-1}}\left(D(p^c; \overline{m}) - k_1\right) - \frac{1}{K} D(p^c; m_t) = 0. \tag{12}$$

Similar to the main analysis, there is perfect private monitoring when $m_{-1}^*(k_1, \overline{m}, p^c) = \underline{m}$, since this implies that $\overline{m}^*(k_1, p^c) = \overline{m}$ from (11). By definition, this occurs when the collusive price is p^*, so $G\left(m_{-1}^*(k_1, \overline{m}, p^*)\right) = 0$.

Consider the effect of a small change in p^c on firm i's discounted collusive profits for some $p^c \in [p^m, p^*]$. Differentiating V_{iH}^c with respect to p^c (and surpressing notation slightly) yields:

$$\frac{\partial V_{iH}^c}{\partial p^c} = \frac{1}{(1-\delta)\left(1 - G\left(m_{-1}^*\right)\right)}\left[\frac{\partial \pi_i^c}{\partial p^c} - G\left(m_{-1}^*\right)\frac{\partial \pi_i^d}{\partial p^c} - g\left(m_{-1}^*\right)\frac{\partial m_{-1}^*}{\partial p^c}\left(\pi_i^c - (1-\delta)V_{iH}^c\right)\right].$$

This highlights the tradeoff that firms face when demand is downward sloping: the first two terms in square brackets have a nonpositive effect on V_{iH}^c for any $p^c \geq p^m$, implying that raising the collusive price above the monopoly level decreases the discounted collusive profits because the per-period collusive profits are lower and the per-period deviation profits are higher; whereas the third term has a nonnegative effect due to the fact that raising the collusive price improves monitoring, so the discounted collusive profits increase because $G\left(m_{-1}^*(k_1, \overline{m}, p^c)\right)$ falls.[21] Nevertheless, $\frac{\partial V_{iH}^c}{\partial p^c}$ is negative when p^c is evaluated at $p^* > p^m$, because the first term is negative for any $p^c > p^m$, and the second and third terms are zero at p^*, since $G\left(m_{-1}^*(k_1, \overline{m}, p^*)\right) = 0$. Thus, the above implies that reducing the collusive price from p^* raises each firm's discounted collusive profits, so firms do not have an incentive to eliminate imperfect monitoring completely by setting $p^* > p^m$, when demand is downward sloping.

[21] Using the implicit function theorem on (12) yields $\frac{\partial m_{-1}^*}{\partial p^c} = \frac{k_1}{K_{-1}}\left(\frac{\partial D}{\partial p^c} / \frac{\partial D}{\partial m}\right) < 0$.

6.4 Customer-specific prices

Up to this point, we have assumed, consistent with Tirole (1988) and Campbell *et al.* (2005), that each firm supplies all units at the same price in any given period. In contrast to this context, it is possible in some settings that a firm could make a limited number of units available at a low price or offer specific terms to each individual buyer. Consequently, when a firm considers deviating from the collusive agreement, it may want to try to hide its deviation by limiting how much is sold at the deviation price. In terms of our model, this would mean that the set of sales that are consistent with a deviation are larger than what has been considered in the main analysis, and that punishment phases would need to occur for a wider range of collusive sales. In fact, if a firm can sell any amount they wish at the deviation price, then all collusive sales will be consistent with even the smallest of deviations, and this would make collusion unsustainable.

Despite this, our results are of interest for two reasons. First, it is reasonable to expect that there are a number of contexts in which firms are likely to sell all units at the same price in any given period. For example, collusion is likely to be difficult to sustain if a firm's prices are set by a number of different sales representatives, so the pricing strategy may need to be coordinated and organised by management at a higher level. This is especially true if the firm is looking deviate from a collusive agreement without triggering a punishment phase. Thus, as argued by Harrington and Skrzypacz (2007, p.318), setting different prices to different buyers will require the managers to predict which buyers make up the market demand. This is unlikely to be feasible if the buyers are small compared to the market demand and market demand varies over time, as in our model. Furthermore, firms may avoid charging consumers different prices if they are concerned that it will create incentives for arbitrage, which could ultimately undermine a firm's attempts to limit the units sold at the low price.[22]

Second, our results are robust under some conditions when firms can offer customer-specific prices, so that a firm can restrict how much is sold at the deviation price by offering such a price to a limited number of buyers. In this context, collusion may still be sustainable if firms have some information on each buyer's demand. For example, suppose that a finite number of symmetric buyers are expected to account for an equal proportion of the market demand (and that our other assumptions on demand in the main analysis are still satisfied). When firms can charge customer-specific prices, they may be able to sustain the monopoly price to each consumer, if

[22]In a setting where arbitrage within any given period is possible, the firms could ensure that no arbitrage occurs in equilibrium by following strategies in which they all set the same price in punishment phases as well as in collusive phases, as described in section 6.1.

they enter the punishment phase whenever at least one firm's collusive sales are consistent with a rival offering a lower price to at least one buyer. Under this strategy, the results will be the same as in the main analysis, if the capacity of the smallest firm is less than or equal to the lowest possible demand of each buyer. The reason is that, even though the largest firms $i > 1$ may only supply one buyer when they deviate, they will still supply at least as much as when the smallest firm deviates. As a result, it remains true that if it is possible for a firm to infer that the smallest firm has not deviated, then it follows that larger firms have also not deviated. Thus, when each buyer accounts for a sufficiently large proportion of the market demand, the size of the smallest firm's capacity still determines when a punishment phase is entered, so the critical discount factor and the best collusive profits will be the same, because each firms' incentives are unaffected.

In contrast, our results do not hold if the smallest firm can supply the lowest possible demand of a buyer, but following the logic of our model implies that collusion can still be sustainable. For instance, in this case, it will be the size of the buyers, not the size of the smallest seller, that determines the range of collusive sales for which firms enter a punishment phase. In particular, as the size of the smallest buyer decreases, collusion will less successful, until it is unsustainable when the buyers are infinitesimally small.

7 Concluding Remarks

We have explored the effects of asymmetries in capacity constraints on collusion in a setting where there is demand uncertainty and where firms never directly observe their rivals' prices and sales. Despite the fact that each firm must monitor the collusive agreement using their private information, we have showed that such private monitoring perfectly detects deviations when demand fluctuations are sufficiently small, and that the critical level is determined by the size of the smallest firm's capacity. Otherwise, private monitoring is imperfect and punishment phases must occur on the equilibrium path. In either case, collusion is hindered when largest firm has more capacity and when the smallest firm has less capacity. We also analysed both the unilateral and coordinated effects of mergers in a unified framework. We showed, in contrast to collusion under perfect observability, that a merger that creates a near monopoly would lead to lower consumer surplus than one that facilitates collusion by creating a more symmetric capacity distribution, when demand fluctuations are sufficiently large. Using an example, we showed that a collusive symmetric duopoly could have substantially higher consumer surplus than more asymmetric noncollusive duopolies.

Our results have two implications for merger policy. First, although market transparency is rightly an important criteria in assessing coordinated effects, our model reemphasises the fact that a lack of transparency about rivals' prices and sales is not a sufficient condition to rule out such effects: it is also necessary to check that firms are unable to detect deviations using only their own prices and sales. While this possibility is explicitly mentioned in the most recent US and European horizontal merger guidelines, our model suggests that this possiblity is less likely when the market structure is more asymmetric. Second, mergers that create symmetric market structures and raise concerns of coordinated effects should not be presumed to be more harmful than asymmetric market structures where collusion is not considered a problem. A collusive agreement may require sufficiently frequent and long price wars that actually increase consumer surplus compared to an alternative outcome where collusion will not arise. This outcome is more likely according to our model, when market demand fluctuates to a large extent over time.

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

[3] Campbell, C., Ray, G., and Muhanna, W. (2005) "Search and Collusion in Electronic Markets." *Management Science*, 51, 497–507

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

[5] Dasgupta, P. and Maskin, E. (1986) "The Existence of Equilibrium in Discontinuous Economic Games, II: Applications" *Review of Economic Studies*, 53(1), 27-41

[6] Davies, S. W. and Olczak, M. (2010) "Assessing the Efficacy of Structural Merger Remedies: Choosing Between Theories of Harm?" *Review of Industrial Organization*, 37(2), 83-99

[7] Davis, P. and Huse, C. (2010) "Estimating the Coordinated Effects of Mergers," Competition Commission Working Paper

[8] Farrell, J. and Shapiro, C. (1990) "Horizontal Mergers: An Equilibrium Analysis," *American Economic Review*, 80(1), 107-126

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

[10] Fudenberg, D. and Levine D. "An Approximate Folk Theorem with Imperfect Private Monitoring," *Journal of Economic Theory*, 54(1), 26-47

[11] Fudenberg, D. and Tirole, J. (1994), *Game Theory*, MIT Press

[12] Gal-Or, E. (1984) "Price Dispersion with Uncertain Demand," *International Economic Review*, 25(2), 441-457

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

[14] Harrington, J. and Skrzypacz, A. (2007) "Collusion under Monitoring of Sales," *RAND Journal of Economics*, 38(2), 314-331

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

[16] Hikisch, S. (2008) "Unilateral and Coordinated Effects Merger Simulation with Non-Linear Demands," Working Paper

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

[18] Ivaldi, M., Jullien, B., Rey, P., Seabright, P., and Tirole, J. (2003b) "The Economics of Unilateral Effects," Report for DG Comp, European Commission

[19] Kühn, K.-U., (2001) "An Economist's Guide through the Joint Dominance Jungle," John M. Olin Centre for Law and Economics Working Paper 02-014

[20] Kühn, K.-U., (2004) "The Coordinated Effects of Mergers in Differentiated Product Markets," John M. Olin Centre for Law and Economics Working Paper 04-020

[21] Lambson, E. (1994) "Some results on optimal penal codes in asymmetric Bertrand supergames," *Journal of Economic Theory*, 62(2), 444-468

[22] Lyons, B. (2002) "Could Politicians be More Right than Economists? A Theory of Merger Standards," Centre for Competition and Regulation Working Paper 02-1

[23] Mailath, G. and Morris, S. "Repeated Games with Almost-public Monitoring," *Journal of Economic Theory*, 102(1), 189-228

[24] Marshall, R. and L. Marx (2008) "Explicit Collusion and Marsket Share Allocation," Duke University, mimeo

[25] Motta, M., Polo, M., and Vasconcelos, H. (2003). "Merger Remedies in the European Union: An Overview," in Lévêque, F. and Shelanski, H., (eds.), *Merger Remedies in American and European Union Competition Law*, Edward Elgar

[26] Roller, L.-H. and Mano, M. (2006) "The Impact of the New Substantive Test in European Merger Control," *European Competition Journal*, 2(1), 9-28

[27] Skrzypacz, A. and Hopenhayn, H. (2004) "Tacit Collusion in Repeated Auctions." *Journal of Economic Theory*, 114, 153–169

[28] Sabbatini, P. (2006) "How to Simulate the Coordinated Effect of a Merger," Working Paper

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

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

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

Appendix

Proof of Lemma 1. There exists a pure strategy Nash equilibrium if $\underline{m} \geq K$, because, from (1), $s_i(p_i, \mathbf{p}_{-i}; m) = k_i \; \forall \; p_i \leq 1$ and $m \in [\underline{m}, \overline{m}]$. Consequently, the best reply of firm i is to set $p_i = 1$ for any $\mathbf{p}_{-i}$, so there is a unique pure strategy Nash equilibrium in which $p_i = 1$ and $\pi_i^N = k_i \; \forall \; i$. In contrast, there is no pure strategy Nash equilibrium if $\underline{m} < K$. To see this, note that any pure strategy Nash equilibrium that can exist requires firms to set $p_j = p_t^{\max} \; \forall \; j$, where $p_t^{\max} \equiv \max\{\mathbf{p}_t\}$. Otherwise, firm $j \in \Omega\left(p_t^{\max}\right)$ has an incentive to increase its price towards $p_t^{\max} \; \forall \; p_j < p_t^{\max}$, since $s_j\left(p_j, \mathbf{p}_{-j}; m\right) = k_j \; \forall \; j \in \Omega\left(p_t^{\max}\right)$. However, for any candidate equilibrium in which $p_j = p \in (0, 1] \; \forall \; j$, firm i has an incentive to lower its price, because

$s_i(p-\epsilon, p; m) > s_i(p,p;m)$ if $\underline{m} < K$, where $\epsilon > 0$ but small. Moreover, for $p_j = p = 0 \; \forall \; j$, firm i has an incentive to raise its price, since Assumption 1 ensures that $s_i(\epsilon, 0; m) > 0 \; \forall \; i$.

Nevertheless, if $K > \underline{m} \geq K_{-1}$, the existence of a mixed strategy is guaranteed by Thereom 1 of Dasgupta and Maksin (1986). To characterise the mixed strategy Nash equilibrium, let $H_i(p)$ denote the probability that firm i charges a price less than or equal to p. Below we demonstrate that the mixed strategy Nash equilibrium profits are given by (2) for all i and that:

$$H_i(p) = \frac{1}{k_i}\left[\frac{(\overline{s}_n - pk_n)}{pk_n(\overline{s}_i - k_i)}\prod_{j=1}^{n} k_j\right]^{1/(n-1)}, \tag{13}$$

where, if firm i is strictly the highest-priced firm, its sales are:

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

This converges to the mixed strategy Nash equilibrium in Fonseca and Normann (2008) as $\overline{m} \to \underline{m}$.

In equilibrium, firm i must receive the following expected profit from charging p:

$$p\left(\prod_{j\neq i} H_j(p)\overline{s}_i + \left(1 - \prod_{j\neq i} H_j(p)\right)k_i\right) = \frac{k_i}{k_n}\overline{s}_n, \; \forall \; i \tag{14}$$

where $\prod_{j\neq i} H_j(p)$ is the probability that firm i is the highest-priced firm. The right-hand side of (14) comes from the fact that firm i expects to receive profit of $\overline{s}_i$ if $p_i = 1$, which implies that firm i will have no incentive to price below $\overline{s}_i/k_i \equiv \underline{p}_i$, where $\underline{p}_n \geq \underline{p}_{n-1} \geq \ldots \geq \underline{p}_1$. Moreover, any firm $j < n$ can guarantee profits of $\frac{k_j}{k_n}\overline{s}_n \geq \overline{s}_j$ by charging a price marginally below $\underline{p}_n$, so all firms have no incentive to price below $\underline{p}_n$. Finally, the fact that all firms $j < n$ place positive probability on charging $\underline{p}_n$ is necessary and sufficient to ensure that the lowest price that firm n will charge is also $\underline{p}_n$. This implies that the lower bound of $H_i(p)$ is $\underline{p} = \underline{p}_n = \overline{s}_n/k_n \; \forall \; i$. Manipulating (14) yields:

$$\prod_j H_j(p) = \frac{(\overline{s}_n - pk_n)}{pk_n(\overline{s}_i - k_i)}k_i H_i(p). \tag{15}$$

It follows from (13) that:

$$\begin{aligned}\prod_j H_j(p) &= \prod_{l=1}^{n}\left\{\left[\frac{(\overline{s}_n - pk_n)}{pk_n(\overline{s}_i - k_i)}\prod_{j=1}^{n} k_j\right]^{1/(n-1)} k_l^{-1}\right\} \\ &= \left[\frac{(\overline{s}_n - pk_n)}{pk_n(\overline{s}_i - k_i)}\prod_{j=1}^{n} k_j\right]^{n/(n-1)} \prod_{l=1}^{n} k_l^{-1}\end{aligned}$$

$$= \frac{(\overline{s}_n - pk_n)}{pk_n (\overline{s}_i - k_i)} \left[\frac{(\overline{s}_n - pk_n)}{pk_n (\overline{s}_i - k_i)} \prod_{j=1}^{n} k_j \right]^{1/(n-1)} \tag{16}$$

Substituting (16) into (15) shows that $H_i(p)$ is as claimed in (13).

It follows from (13) that $H_i(1) \lesseqgtr 1$ if $\frac{k_i^{n-1}}{\prod_{j \neq n} k_j} \gtreqless 1$. This has two implications. First, if $\frac{k_i^{n-1}}{\prod_{j \neq n} k_j} \geq 1$, then firm i randomises over $[\underline{p}, 1]$ and puts mass of $1 - H_i(1)$ on a price of 1 when the inequality is strict. Note that $\frac{k_i^{n-1}}{\prod_{j \neq n} k_j} > 1$ never holds if $k_i = k \ \forall \ i$ but always holds for firm n if $k_n > k_1$. Second, if $\frac{k_i^{n-1}}{\prod_{j \neq n} k_j} < 1$ for some $i < n$, then firm i randomises over $[\underline{p}, \overline{p}_i]$ where $\overline{p}_i < 1$ solves $H_i(\overline{p}_i) = 1$. Consequently, the probability distributions of the larger firms with higher upper bounds must be adjusted accordingly. For example, if $\overline{p}_i < 1$ only for firm 1 (which is the case for any triopoly with $k_1 < k_2$), then the largest $n-1$ firms play with the $H_i(p)$ adjusted so that $n-1$ replaces n over $[\overline{p}_1, 1]$. Note that $\frac{k_i^{n-1}}{\prod_{j \neq n} k_j} < 1$ never holds if $n = 2$ or if $k_i = k \ \forall \ i$ for any $n \geq 2$. ■

Proof of Proposition 1. There is perfect private monitoring if firm i, $\forall \ i$, can detect with certainty whether $p_{jt-1} = p_{t-1} \ \forall \ j$ from its private information, z_{it}. Since only p_{it-1} and s_{it-1} provide information about p_{jt-1}, $j \neq i$, we must find the levels of $\overline{m}$ for which firm i's set of sales when $p_{jt-1} = p \ \forall \ j$ and the sales it would receive if $p_{jt-1} \neq p$ for some j are mutually exclusive, $\forall \ i$.

To begin, suppose for the moment that firm i set $p \leq 1$ and that $s_i < k_i$. Thus, it follows from (1) that it knows that no firm priced higher than it, $p = p^{\max}$. With this information, from (3) it can infer with certainty whether $p_j = p \ \forall \ j$ if:

$$\frac{k_i}{K}\underline{m} > \frac{k_i}{K - \sum_{j \in \Omega(p)} k_j} \left(\overline{m} - \textstyle\sum_{j \in \Omega(p)} k_j \right), \ \forall \ \ j \neq i \tag{17}$$

where $\Omega(p)$ is nonempty in (17). Since the right-hand side of (17) is decreasing in $\sum_{j \in \Omega(p)} k_j$, it follows that firm i, $\forall \ i$, can infer with certainty, $\forall \ m \in [\underline{m}, \overline{m}]$, that no rival (or any set of rivals) charged $p_j < p$, for some j, if $\frac{k_i}{K}\underline{m} > \frac{k_i}{K_{-1}}(\overline{m} - k_1)$, where rearranging yields:

$$\overline{m} < \frac{k_1}{K}(K - \underline{m}) + \underline{m} \equiv \overline{m}^* (k_1).$$

It follows from the above that $\overline{m}^* (k_1)$ is uniquely defined by its parameters and that $\overline{m}^* (k_1) < K$. An implication of the latter is that $\overline{m} < \overline{m}^* (k_1)$ is a necessary and sufficient condition for firm i, $\forall \ i$, to detect with certainty whether $p_j = p \ \forall \ j$, because it can also infer with certainty that at least one rival has not set a higher price for all $\overline{m} < K$. This is because $s_{it-1}(p; m_{t-1}) < k_i$ $\forall \ m_{t-1} < K$, yet it receives sales of $s_{it-1} = k_i$ if $p < p_{t-1}^{\max}$. Thus, it follows from the above that there is perfect private monitoring, if and only if $\overline{m} \in [\underline{m}, \overline{m}^* (k_1))$. ■

Proof of Proposition 2. To prove that h^t is common knowledge despite it not being directly observable, we show that: (a) firm i, $\forall\ i$, can infer h^t in period t given its private information, z_i^t ; and that (b) firm i, $\forall\ i$, can infer that firm j, $\forall\ j \neq i$, can infer h^t in period t given firm j's private information, z_j^t. It then easily follows that firm i can infer that firm j can infer that firm i can infer h^t, and so on. Since only $p_{i\tau}$ and $s_{i\tau}$ provide information about y_τ and denoting $y_{j\tau}$ as firm j's private signal of y_τ, it suffices to check that firm i, $\forall\ i$, can infer that $y_{j\tau} = y_\tau\ \forall\ j$, for any given $p_{i\tau}$ and $s_{i\tau}$. It follows from (1) that there are two possibilities that we must check: (i) $s_{i\tau} = k_i$, and (ii) $s_{i\tau} < k_i$.

First, suppose that firm i sets $p_{i\tau}$ and that $s_{i\tau} = k_i$. If $k_i \leq \underline{s}_i$, then it can infer from its sales that $s_{j\tau} = k_j \leq \underline{s}_j\ \forall\ j$. Otherwise, if $k_i > \underline{s}_i$, then from (1) it can infer from its sales that firm $j \neq i$'s sales are:

$$s_{j\tau}\left(p_{i\tau}, \mathbf{p}_{-i\tau}; m_\tau\right) \begin{cases} = k_j & \text{if } p_{j\tau} < p_\tau^{\max} \\ \leq \underline{s}_j & \text{if } p_{j\tau} = p_\tau^{\max} \end{cases} \tag{18}$$

for all j and m_τ. Thus, $s_{i\tau} = k_i > \underline{s}_i$ implies that firm i can infer that any rival j with $p_{j\tau} = p_\tau^{\max}$ receives sales such that $s_{j\tau} \leq \underline{s}_j$. Similarly, it can infer that any rival j with $p_j < p_\tau^{\max}$, that receives sales of $s_{j\tau} = k_j$, can also infer the same as it.

Next, suppose that firm i sets $p_{i\tau}$ and that $s_{i\tau} < k_i$. In this case, firm i can infer that $p_{i\tau} = p_\tau^{\max}$, because firm i would supply its full capacity for any m_τ otherwise. If $s_{i\tau} \leq \underline{s}_i$, then it can infer from its sales that firm $j \neq i$'s sales are given by (18). Thus, $s_{i\tau} \leq \underline{s}_i$ implies that firm i knows that a rival j with $p_{j\tau} = p_\tau^{\max}$ also receives sales of $s_{j\tau} \leq \underline{s}_j$, and it follows from above that it knows that any rival with a price below $p_\tau^{\max}$ can infer this. Finally, if $s_{i\tau} \in (\underline{s}_i, k_i)$, then it can infer from its sales that $s_{j\tau} \in (\underline{s}_j, k_i)\ \forall\ j$.

The above implies that firm i, for all i, can infer that $y_{j\tau} = \overline{y}\ \forall\ j$, when it receives sales of $s_{i\tau} \in (\underline{s}_i, k_i)$. Likewise, it can infer that $y_{j\tau} = \underline{y}\ \forall\ j$, when it receives sales of $s_{i\tau} \notin (\underline{s}_i, k_i)$. Thus, given each firm can always observe its sales in each period, it follows that h^t is common knowledge at the beginning of period t, for all t. ■

Proof of Proposition 3. Substituting V_{iH}^c into (6) and rearranging yields:

$$\delta^T \leq \frac{\pi_i^d - \pi_i^c - \delta[1 - G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)]\left(\pi_i^d - \pi_i^N\right)}{\delta\left[-\left(\pi_i^c - \pi_i^N\right) + G\left(m_{-1}^*\left(k_1, \overline{m}\right)\right)\left(\pi_i^d - \pi_i^N\right)\right]} \tag{19}$$

For any $\overline{m} < K$, it follows from the proportional rationing rule that $\pi_i^c = p^c \frac{k_i}{K} \widehat{m}$ and $\pi_i^d = p^c k_i$, and all other terms are independent of p^c. Thus, the right-hand side of (19) is increasing in

p^c, which implies that collusion is easier to sustain for higher p^c. Furthermore, V^c_{iH} is strictly increasing in p^c. Consequently, we focus on $p^c = 1$, where (19) simplifies to:

$$\delta^T \leq \frac{k_n - K\delta\left(1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)\right)}{\delta\left[k_n - K\left(1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)\right)\right]} \equiv X \tag{20}$$

Collusion under imperfect private monitoring is sustainable for a sufficiently large $T > 0$ if both the numerator and the denominator of X are negative. This is true if $G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right) < 1 - \frac{k_n}{K}$ and if:

$$\delta \geq \frac{1}{\left(1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)\right)} \frac{\pi^d_i - \pi^c_i}{\pi^d_i - \pi^N_i} = \frac{1}{\left(1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)\right)} \frac{k_n}{K} \equiv \delta^*\left(k_1, k_n\right)$$

Notice that $G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right) < 1 - \frac{k_n}{K}$ ensures that $\delta^*\left(k_1, k_n\right) < 1$. Since $\frac{\partial G\left(m^*_{-1}\right)}{\partial \overline{m}} = g\left(m\right) \frac{\partial m^*_{-1}}{\partial \overline{m}} > 0$, it follows that there is a unique level of $\overline{m}$, denoted $\overline{m}'\left(k_1, k_n\right)$, that sets $G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right) = 1 - \frac{k_n}{K} < 1$, where $G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right) \in \left[0, 1 - \frac{k_n}{K}\right)$ for all $\overline{m} \in \left[\overline{m}^*\left(k_1\right), \overline{m}'\left(k_1, k_n\right)\right)$ and $\overline{m}'\left(k_1, k_n\right) < K$ since $G\left(K\right) = 1$.

Finally, given V^c_{iH} is strictly decreasing in T, the optimal length of the punishment phase, denoted $T^*\left(k_1, k_n\right)$, solves $\max_T V^c_{iH}$ subject to $\delta^T \leq X$, which implies $T^*\left(k_1, k_n\right) \equiv \frac{\ln X}{\ln \delta} > 0$. Thus, collusion under imperfect private monitoring is sustainable for any $T \geq T^*\left(k_1, k_n\right)$, if $\delta \geq \delta^*\left(k_1, k_n\right)$ and $\overline{m} \in \left[\overline{m}^*\left(k_1\right), \overline{m}'\left(k_1, k_n\right)\right)$. ■

Proof of Proposition 4. Differentiating $\delta^*\left(k_1, k_n\right) = \frac{1}{\left(1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)\right)} \frac{k_n}{K}$ with respect to k_1 yields:

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

Thus, it follows from $\frac{\partial k_n}{\partial k_1} \in [-1, 0]$ and $\frac{\partial m^*_{-1}}{\partial k_1} = -\frac{K(K - \overline{m})}{(K - k_1)^2} < 0$ that $\frac{\partial \delta^*}{\partial k_1} < 0$. The fact that δ^* falls as k_n decreases implies that $\frac{\partial \delta^*}{\partial k_n} > 0$.

Differentiating $T^*\left(k_1, k_n\right) = \frac{\ln X}{\ln \delta}$ with respect to k_1 yields $\frac{\partial T^*}{\partial k_1} = \frac{1}{\ln \delta} \frac{1}{X} \frac{\partial X}{\partial k_1}$, where $X > 0$ is defined in (20) and:

$$\frac{\partial X}{\partial k_1} = -\frac{(1-\delta)K\left(1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)\right)}{\delta\left[k_n - K\left(1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)\right)\right]^2} \left[\frac{\partial k_n}{\partial k_1} + k_n \frac{g\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)}{1 - G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right)} \frac{\partial m^*_{-1}}{\partial k_1}\right].$$

It follows from $\delta \in (0, 1)$, $G\left(m^*_{-1}\left(k_1, \overline{m}\right)\right) \in \left[0, 1 - \frac{k_n}{K}\right)$, $\frac{\partial k_n}{\partial k_1} \in [-1, 0]$, $g(m) > 0$ and $\frac{\partial m^*_{-1}}{\partial k_1} < 0$ that $\frac{\partial X}{\partial k_1} > 0$. Together with $\ln \delta < 0$ and $X > 0$, this implies that $\frac{\partial T^*}{\partial k_1} < 0$. The fact that X rises as k_n decreases implies that a fall in k_n decreases T^*, so $\frac{\partial T^*}{\partial k_n} > 0$.

Finally, recall that $\overline{m}'(k_1, k_n)$ is the level of $\overline{m}$ that sets $G\left(m_{-1}^*(k_1, \overline{m})\right) = 1 - \frac{k_n}{K}$, so let $Z \equiv 1 - \frac{k_n}{K} - G\left(m_{-1}^*(k_1, \overline{m})\right) = 0$. Using the implicit function theorem it follows that:

$$\frac{\partial \overline{m}'}{\partial k_1} = -\frac{\frac{\partial Z}{\partial k_1}}{\frac{\partial Z}{\partial \overline{m}}} > 0 \qquad \text{and} \qquad \frac{\partial \overline{m}'}{\partial k_n} = -\frac{\frac{\partial Z}{\partial k_n}}{\frac{\partial Z}{\partial \overline{m}}} < 0$$

since $\frac{\partial Z}{\partial \overline{m}} = -g\left(m_{-1}^*(k_1, \overline{m})\right)\frac{\partial m_{-1}^*}{\partial \overline{m}} < 0$, $\frac{\partial Z}{\partial k_1} = -\frac{1}{K}\frac{\partial k_n}{\partial k_1} - g\left(m_{-1}^*(k_1, \overline{m})\right)\frac{\partial m_{-1}^*}{\partial k_1} > 0$, and $\frac{\partial Z}{\partial k_n} = -\frac{1}{K} - g\left(m_{-1}^*(k_1, \overline{m})\right)\frac{\partial m_{-1}^*}{\partial k_1}\frac{\partial k_n}{\partial k_1} > 0$. ■

Proof of Proposition 5. It follows from (19) that:

$$1 - \delta^{T^*} = \frac{-(1-\delta)\left(\pi_i^d - \pi_i^c\right)}{\delta\left[-\left(\pi_i^c - \pi_i^N\right) + G\left(m_{-1}^*(k_1, \overline{m})\right)\left(\pi_i^d - \pi_i^N\right)\right]}.$$

Substituting the above into V_{iH}^c gives:

$$V_{iH}^c = \frac{\left(\pi_i^c - G\left(m_{-1}^*(k_1, \overline{m})\right)\pi_i^d\right)}{(1-\delta)\left(1 - G\left(m_{-1}^*(k_1, \overline{m})\right)\right)} = \frac{1}{(1-\delta)}\frac{k_i}{K}\widehat{m}\left(\frac{\widehat{m} - G\left(m_{-1}^*(k_1, \overline{m})\right)K}{\widehat{m}\left(1 - G\left(m_{-1}^*(k_1, \overline{m})\right)\right)}\right),$$

where:

$$\widehat{p}^c(k_1, \overline{m}) = \frac{\widehat{m} - G\left(m_{-1}^*(k_1, \overline{m})\right)K}{\widehat{m}\left(1 - G\left(m_{-1}^*(k_1, \overline{m})\right)\right)}.$$

This implies that $\widehat{p}^c(k_1, \overline{m}^*(k_1)) = 1$, since $G\left(m_{-1}^*(k_1, \overline{m}^*(k_1))\right) = 0$, $\widehat{p}^c(k_1, \overline{m}'(k_1, k_n)) = \widehat{p}^N(k_n, \widehat{m})$, since $G\left(m_{-1}^*(k_1, \overline{m}'(k_1, k_n))\right) = 1 - \frac{k_n}{K}$. Furthermore, $\widehat{p}^N(k_n, \widehat{m}) < \widehat{p}^c(k_1, \overline{m}) < p^m$ for all $\overline{m} \in (\overline{m}^*(k_1), \overline{m}'(k_1, k_n))$.

Note that $\widehat{p}^c(k_1, \overline{m})$ is a function of k_1 and it is independent of the size of the capacity of all other firms, including k_n. Differentiating $\widehat{p}^c(k_1, \overline{m})$ with respect to k_1 yields:

$$\frac{\partial \widehat{p}^c}{\partial k_1} = -\frac{(K - \widehat{m})\, g(m_{-1}^*)}{\widehat{m}\left(1 - G(m_{-1}^*)\right)^2}\frac{\partial m_{-1}^*}{\partial k_1} > 0,$$

since $g(m) > 0$, $0 < \widehat{m} < \overline{m} < K$ and $\frac{\partial m_{-1}^*}{\partial k_1} < 0$. ■

Proof of Proposition 6. Substituting in for $\widehat{p}^N(k_\nu, \widehat{m}) > \widehat{p}^c(k_1, \overline{m})$ and rearranging yields:

$$G\left(m_{-1}^*(k_1, \overline{m})\right) > 1 - \frac{k_\nu}{K}.$$

Recall that $\overline{m}'(k_1, k_n)$ is the level of $\overline{m}$ that sets $G\left(m_{-1}^*(k_1, \overline{m})\right) = 1 - \frac{k_n}{K}$ and that a necessary condition for collusion under imperfect private monitoring to be sustainable is $G\left(m_{-1}^*(k_1, \overline{m})\right) \in \left[0, 1 - \frac{k_n}{K}\right)$. Thus, $\widehat{p}^c(k_1, \overline{m}) < \widehat{p}^N(k_\nu, \widehat{m})$ if $G\left(m_{-1}^*(k_1, \overline{m})\right) \in \left[1 - \frac{k_\nu}{K}, 1 - \frac{k_n}{K}\right)$, which can only be true if $k_\nu > k_n$. It follows from $\frac{\partial G(m_{-1}^*)}{\partial \overline{m}} = g(m)\frac{\partial m_{-1}^*}{\partial \overline{m}} > 0$ that $G\left(m_{-1}^*(k_1, \overline{m})\right) = 1 - \frac{k_\nu}{K} < 1 - \frac{k_n}{K}$ for some $\overline{m} < \overline{m}'(k_1, k_n)$. Furthermore, by definition $\overline{m}'(k_1, k_\nu)$ is the level of $\overline{m}$ that sets $G\left(m_{-1}^*(k_1, \overline{m})\right) = 1 - \frac{k_\nu}{K} > 0$, where $\overline{m}'(k_1, k_\nu) > \overline{m}^*(k_1)$ since $G\left(m_{-1}^*(k_1, \overline{m}^*(k_1))\right) = 0$. Thus, $\widehat{p}^c(k_1, \overline{m}) < \widehat{p}^N(k_\nu, \widehat{m})$ if $k_\nu > k_n$ and $\overline{m} \in [\overline{m}'(k_1, k_\upsilon), \overline{m}'(k_1, k_n))$. ■