CMPO Working Paper Series No. 00/25

Promoting Competition in the Presence of Essential Facilities

Paul A Grout[1]
and
In-Uck Park[2]

[1]*University of Bristol and CMPO*
[2]*University of Pittsburgh and CMPO*

August 2000

Abstract

This paper addresses the issue of how regulators can use access pricing to promote entry by innovatory firms in the presence of essential facilities. The entrants have lower costs that spillover to firms in the market but the regulator is not able to distinguish which entrants have low costs and which do not. In a dynamic framework with entrants of differing quality technology spillovers have two effects. One is positive in that the incumbent can copy the cheaper technology of the entrant. This reduces cost in the industry and offsets the fixed entry cost associated with entry. The other is a negative effect in that the ability to use access pricing to deter entry of bad quality entrants is reduced. A low quality firm can free ride on the quality of a good entrant since it is protected from the consequences of its high costs and poor technology if a good firm has already entered or may be about to enter. The greater the spillover the greater the desire to attract good entrants but also the harder it is to penalise poor quality entrants.

This paper considers this dilemma and the consequences for public policy. The question we address is whether the lack of full information encourages the regulator to sustain entry enhancing policies for longer or whether the regulator makes entry harder. Generally, we show that the incentives are for the regulator to limit entry enhancement in the face of incomplete information rather than be more open in the face of the inability to determine the good firms from the bad ones. We also show that for certain configurations the good firm has an incentive to raise its costs, i.e., become a less good competitor.

JEL Classification: L51
Keywords: essential facilities; promoting competition; access pricing; technology spillovers

Acknowledgements

We are grateful to participants of the Institute of Economics and Statistics Seminar, University of Oxford and the AEA meetings at Chicago for helpful comments. This research has been funded by the Leverhulme Trust under the Regulation of Newly Privatised Entities project

Address for Correspondence
CMPO, Department of Economics
University of Bristol
8 Woodland Road
Bristol
BS8 1TN
p.a.grout@bristol.ac.uk

CMPO is funded by the Leverhulme Trust.

Non-Technical Summary

Essential facilities provide the incumbent firm with an advantage in the provision of associated downstream facilities. For example, there may be costs that are common between the downstream and essential facilities such that provision of the downstream facility by the provider of the essential facility is cheaper than having competitive provision. However, competitors may be more innovatory and this may reduce the cost of provision of the overall product even though there are more firms than are strictly necessary. The competing firms need access to the essential facility and in our framework the price of access to the essential facility and the price of the final product provided by the incumbent are regulated. The paper addresses the issue of how regulators can use access pricing to promote entry by innovatory firms in the presence of essential facilities. The entrants have lower costs that spillover to firms in the market but the regulator is not able to distinguish which entrants have low costs and which do not.

Where entrants are of differing quality technology spillovers have two effects. One is positive in that the incumbent can copy the cheaper technology of the entrant. This reduces cost in the industry and offsets the fixed entry cost associated with entry. The other is a negative effect in that the ability to use access pricing to deter entry of bad quality entrants is reduced. That is, it is protected from the consequences of its high costs and poor technology, if a good firm has already entered or may be about to enter, since it can copy the cheaper technology. The higher cost entrant is bad for efficiency since it causes additional fixed costs to be incurred and brings no benefit but the spillovers prevent it from being ‘hurt’ by its own relative inefficiency. The greater the spillover the greater the desire to attract good entrants but also the harder it is to penalise poor quality entrants.

This paper considers this dilemma and the consequences for public policy. The question we address is whether the lack of full information encourages the regulator to sustain entry enhancing policies for longer or whether the regulator makes entry harder. Generally, we show that the incentives are for the regulator to limit entry enhancement in the face of incomplete information rather than be more open in the face of the inability to determine the good firms from the bad ones. We also show that for certain configurations the good firm has an incentive to raise its costs, i.e., become a less good competitor.

1. Introduction

Optimal pricing for access to essential facilities has received considerable attention in recent years both from economists and policy makers throughout the world. This has focused mostly on network utilities but other issues such as access to ports have received regulatory attention. Recent interest has been driven in part by the wave of privatisations of network utilities around the world and international drive to open up network markets.

One of the most common access problems arises in networks where a service requires two legs, one a monopoly owned essential facility, and the other a potentially competitive segment. Suppliers other than the owner of the essential facility need to interconnect with the monopoly supplier and will generally be expected to contribute to the cost of the essential facility. The appropriate structure of this access charge has been the focus of significant debate within the economics profession. In basic models a Ramsey pricing rule, or sometimes a very simple version of this often referred to as the Baumol-Willig rule where the access charge is set at the marginal cost of provision plus the opportunity cost, is optimal (see, for example, Baumol and Sidak (1994) and Laffont and Tirole (1994, 1996)). Where there are issues such as network externalities or unregulated monopoly suppliers then there will be deviations from these rules (see, for example, the discussion in Economides (1996)).

A feature of conventional access pricing rules is that they make entry difficult. The potential entrant has to meet, in the form of an access charge, both the monopolist's marginal cost of the essential facility and the customer's contribution to the monopolist's common cost, and then cover the entrant's own cost before they can profitably enter the market. Once one includes the up-front cost of entry it is often difficult to compete in the presence of such an access pricing regime. At the same time is common for there to be a legal obligation on regulatory agencies to promote effective competition. This is the case in the European Union and within the framework of UK regulatory policy, where regulators have proved resistant to the implementation of conventional access pricing rules (see, for example, Grout (1996)). These two positions can be reconciled if there are external effects.[1] For example, new entrants may bring innovations which lower costs for all firms then positive entry assistance through lower access prices can be beneficial. This, however,

---

[1] See for example, De Fraja (1997, 1999).

raises questions such as how long should the entry assistance last. Furthermore, although a regulatory body may wish to reduce access charges to attract innovatory firms, in many cases it is difficult for the regulatory body to distinguish, at least in the medium term, between the entrants that will be most beneficial and those that are less beneficial. Indeed, there is an inherent dilemma when pursuing efficiency and wishing to promote competition.

In a dynamic framework with entrants of differing quality technology spillovers have two effects. One is positive in that the new technology of the entrant can be copied by the incumbent. This reduces cost in the industry and offsets the fixed entry cost associated with entry. The other is a negative effect in that the ability to use access pricing to deter entry of bad quality entrants is reduced. A low quality firm can free ride on the quality of a good entrant since it is protected from the consequences of its high costs and poor technology if a good firm has already entered or may be about to enter. The greater the spillover the greater the desire to attract good entrants but also the harder it is to penalise poor quality entrants. This paper considers this dilemma and the consequences for public policy. It complements the existing access pricing literature in that it focuses on issues that have not been addressed to date, in particular the limitations of simple access pricing in the presence of the spillovers in a game theoretic setting.

The layout of the paper is as follows. Section 2 outlines the model. There is a regulator, an incumbent that owns the essential facility and two potential entrants in the potentially competitive section of the network. The incumbent has a common cost between the two sections of the network. This favours monopoly provision but the two potential entrants have lower production costs and these spillover when they enter the market. The disadvantage of entry is that there is a fixed one-off entry cost per firm. The regulator sets a price cap which has to ensure that the incumbent can finance its activities (i.e. has non negative expected profit in equilibrium) and then sets an access pricing regime which may encourage or discourage entry. There are two time periods, one of the firms arrive in each period and each has equal probability of being first. We outline and discuss the subgame perfect equilibria of the model (technical proofs are omitted in this version)

Section 3 of the paper considers the position when the regulator can observe whether the first firm is the good one (i.e., lower production cost) or the bad one (i.e., higher production cost). In this case the regulator is able to implement the first-best solution and we characterise this. The equilibria 'subsidise entry' to accommodate the spillover effect. There are four profiles that are optimal. Either the prices encourage early entry by the good firm but discourage late entry, encourage entry by the good firm at any time, encourage entry by the first firm and discourage entry thereafter, or encourage entry by the good firm at all times and entry by the bad firm in the early period.

As indicated, the process of achieving first-best by subsidising entry assumes that it is possible to observe whether entrants are good or bad firms. In general, it is more plausible to assume that the regulator is very unsure when setting the policy. For example, the UK telecommunications regulatory regime only allowed for one new entrant, Mercury Communications plc, in the UK market for many years after the privatisation of British Telecommunications. One can think of this as a very extreme version of our model. It was far from clear at that time whether Mercury was a good quality competitor. Indeed, ex post there are mixed views as to the quality of Mercury as a competitor in this period and the policy was eventually abandoned in favour of a more open one. Section 4 considers the model when it is not possible to identify whether the entrant is the good or bad firm. The question we address is whether the lack of full information encourages the regulator to sustain entry enhancing policies for longer or whether the regulator makes entry harder. We have mentioned above that when it is not possible to observe the types then the spillover effect makes it harder to use the access pricing structure to deter poor firms since a regime that wishes to attract a low cost firm given a high cost one has entered cannot prevent a high cost firm entering in the wake of the low cost firm since the spillover protects the bad firm from the consequences of its own inefficiency. Similarly, a regime that wishes to attract a higher cost firm in the early stages cannot prevent a low cost firm earning a positive surplus should it be the first in the market. This prevents the implementation of the first-best. Generally, we show that the incentives are for the regulator to limit entry enhancement in the face of incomplete information rather than be more open in the face of the inability to determine the good firms from the bad ones.

Section 4 also addresses certain features of the equilibria. It shows that in the presence of imperfect information there are profiles which are superior to the implementible profiles but that they are not time consistent. More interestingly, it is shown that for certain configurations the good firm has an incentive to raise its costs, i.e., become a less good competitor. The intuition for this is that the regulator will not wish to encourage the high cost firm if its costs are significantly worse than the low cost firm. In this case the access pricing regime need provide no surplus to the good firm. In contrast, if the bad firm is not too inefficient in comparison to the good firm then the optimal access pricing regime encourages entry by the high cost firm which implies that the low cost firm earns a positive expected surplus. That is, the informational rent of the good firm can be increased by reducing the extent of its superiority over the high cost firm.

2. The Basic Model

The model consists of a regulated market with one incumbent and two potential entrants. The market demand for the final product is represented by a differentiable function $D(p)$ with derivatives $D'(p) < 0$ and $D''(p) > 0$. The incumbent has control of the upstream part of the network which is an essential facility for access to customers. The current state of technology available to the incumbent for provision of the downstream part of the network is a constant cost per unit. Each potential entrant to the downstream activity has a fixed cost of entry, F. The two potential entrants differ in their states of technology that they bring into the industry when they enter. Both costs are below the incumbent's cost per unit but one of the entrants, referred to as the good type, has the lowest cost technology and the other, referred to as the bad type, has a technology with costs between the incumbent and the good entrant. We use g and b as shorthand for good and bad types, respectively. For simplicity, we assume without loss of generality that the incumbent's cost per unit is 1, the good entrant's cost per unit is 0 and the bad entrant's cost is c, $0 < c < 1$. We assume that there is a complete spillover of technology. That is, the lowest cost technology in place in the market at any time can be copied costlessly by others in the industry.[2]

The extensive form game of the model consists of an initiation stage and two subsequent periods. Formally, we can think of the initiation stage as one where the regulator sets a price cap, p, which the incumbent either accepts or rejects. If it is rejected, there is no production and the payoffs to all involved parties are identically zero. This formalises

(2) The attraction of assuming a complete spillover is that it provides the richest set of outcomes and avoids having to decide on the form of market shares in a situation where there are differing costs and a binding price cap (see footnote 3). Subject to a resolution of this latter difficulty none of the main results in the paper appear to be dependent on the complete spillover assumption.

the idea that a regulator must allow the regulated firm to fund its activities, i.e., the regulated firm will only accept a price cap if the expected profit is non-negative. If it is accepted, the incumbent is locked in, that is, the incumbent must operate the network in the industry in both periods and provide access to new entrants if they wish. Finally, the regulated firm has a common and fixed cost of $L > 0$ which is necessary for it to operate in either the upstream or downstream market.

In each of the two periods, a sequence of moves take place: (i) one of the potential entrants arrives at the market, (ii) the regulator sets an access price a_i for the current period $i = 1, 2$, (iii) the arrived firm makes an entry decision, (iv) the market reaches the Cournot solution if the Cournot price is below the price cap, $\bar{p}$; the firms in the market share the market demand $D(\bar{p})$ evenly at the price $\bar{p}$ if the Cournot price is above $\bar{p}$, and (v) each firm in the market pays a_i to the incumbent. Access prices are allowed to be negative. We assume for the purposes of this paper that the market is sufficiently large to ensure that the optimal price cap imposed by the regulator is binding.[3]

One of the two potential entrants (i.e., g or b) arrives at the market in period 1 (equal probability of each event) and the other arrives in period 2. There are two arrival contingencies that describe candidates' types in the two periods: one arrival contingency is that the first candidate is g and the second candidate is b, and the other contingency is the reverse. Candidates are referred to as entrants when they actually enter the market. The type of each candidate is known to firms in the market when he enters but the type may not be known by the regulator. We consider two possibilities for the regulator's information on the candidates' type. As a benchmark, we consider the case where the regulator observes the firm's type on arrival. We then consider the case where the regulator observes the occurrence of entry but not the type of entrants. Access prices can be made contingent upon what the regulator has observed. In the former case, therefore, the regulator has more capacity to control entry by setting access prices contingent upon entrants' types.

In either case, the regulator sets the price cap and access prices to maximise the

(3) With a binding price cap the market shares of identically place firms in a Cournot equilibrium may not be exactly equal since they lie in a range around the equal market shares case. The closer the price cap is to the unconstrained price the smaller the range. The equal share is the only fixed sharing rule that is compatible with the Cournot assumption for all market sizes where the price cap is binding and makes this the natural case to employ. Note the complete spillover assumption implies all marginal costs are identical. As noted in footnote 2, in the absence of complete spillovers it would be far harder to determine a natural output assumption.

expected level of welfare (i.e. the consumers surplus and the producers surplus) over the two periods. The incumbent and each candidate select their strategy to maximise (expected) surplus over the two periods, which is total revenue in excess of total expense. There is no discounting. The description of the game is common knowledge and we focus on the subgame-perfect equilibria of this game.

Abusing notation slightly, it is convenient to use g and b to represent the entry of good and bad types, and we use $\emptyset$ to represent no entry. An entry sequenc is an ordered pair, r, in $\{(g,b,\emptyset) \times (g,b,\emptyset)\}$ and represents a sequence of entry decisions by the two potential entrants in periods 1 and 2. An entry profile is an ordered pair of entry sequences $A = \{r, r'\}$ where r is the entry sequence given that g arrives first and r' is the entry sequence given that b arrives first. For example, an entry profile $\{(g,b),(b,\emptyset)\}$ describes the following; If g arrives first both candidates enter in due course, while in the alternative case (i.e., if b arrives first) b enters in period 1 but the g does not enter in period 2. There are four possible entry sequences for each arrival contingency, so there are sixteen entry profiles. The regulator's objective is to implement the best possible entry profile in the most efficient way, by setting the price cap and access prices to provide the right incentives for the producers.

An access pricing strategy is a strategy of the regulator which consists of a price cap p and access prices contingent on the history observable by the regulator. Given an access pricing strategy, we apply a backward induction argument to determine each potential candidate's entry decision for each possible history. Recording the entry decisions that would be realised for each arrival contingency, we derive an entry profile that is a 'best-response' to the given access pricing strategy. Every entrant in this profile derives non-negative expected surplus. Note, however, that the fact that a profile is a best-response to an access pricing strategy does not by itself mean that the regulator can induce it using the associated access prices since the incumbent must also derive non-negative expected surplus from the access pricing strategy or else the incumbent would not accepted it in an equilibrium. We say that a profile is incentive compatible at a price cap p if i) it is a best-response to an access pricing strategy whose price cap component is p and ii) the incumbent derives non-negative expected surplus.

The surplus of each producer is defined in the natural way. That is, the surplus of each entrant is the total revenue in excess of total cost including the entry cost F and the access price transfers. The (expected) surplus of the incumbent is total revenue (revenues and access price receipts) minus costs including the common cost L. The producers surplus is the sum of surpluses of all producers and total welfare is the consumers surplus and the producers surplus. Note, once a price cap, p, is accepted, a subgame starts in which the regulator can induce any profile as long as it is a best-response at p. In that subgame, therefore, the regulator will actually implement the profile that generates the highest welfare among all profiles that are a best-response at p. This profile is called subgame-optimal at p. Note that given a price cap, the welfare level generated from a profile is not affected by the access prices used to induce it, because they are only transfers between producers.

Finally, an entry profile is implementible at p if it is incentive compatible and subgame-optimal at p. The optimal profile that the regulator will implement is the one that generates the highest welfare among all profiles that are implementible (at some p).

3. Complete Information and the First-best

The key instruments that the regulator uses to induce the optimal entry profile are the access prices that transfer payoffs between the producers. In the benchmark case that the regulator observes the type of candidates, she can induce any transfers between producers by using appropriate access prices, as long as every producer has non-negative surplus. In particular, transfers can be made in such a way that every entrant has zero surplus and the incumbent reaps the entire producers surplus. Therefore, an entry profile is incentive compatible at a price cap p if and only if the producers surplus is non-negative at p. If the regulator can commit to an access pricing strategy, she will compare the welfare levels from all profiles at the price caps at which they are incentive compatible, and implement the one that generates the highest welfare. We refer to this profile as first-best (implicitly in association with the price cap that generates the highest welfare).

Note that we have not considered subgame-optimality in defining the first-best. When commitment is not possible as is the case in our model, therefore, the first-best is not necessarily implementible because it may not be subgame-optimal at the associated price cap. We show later that this problem does not arise in the benchmark case: the first-best is in fact subgame-optimal and therefore, the regulator will implement it. First, we identify the ¯rst-best.

The relative performance of entry pro¯les (hence, the ¯rst-best) varies depending on the parameter value of c and F. Given a `cost con¯guration' (c;F) in (0;1) £ R+ we say that an entry pro¯le A dominates another pro¯le A^0 at p, if the welfare from A exceeds the welfare from A^0 at the price cap p. The next results identify several entry pro¯les that are always dominated.

Lemma 3 1: Given any cost con¯guration (c;F) and price cap p,

(i) f(g; ;); r^{0}g dominates f(g; b); r^{0}g at p;

(ii) f(g; ;); r^{0}g dominates f(; ; b); r^{0}g at p;

(iii) f(g; ;); (; ; ;)g dominates f(; ; ;); r^{0}g at p if r^0 is not (; ; ;).

Sketch proof: Obviously (g; ;) generates a larger producers surplus than (g; b) because the second period entry b incurs the entry cost F without lowering cost of production. Since the consumers surplus depends only on the price cap (not the pro¯le), part (i) follows. Analogous arguments establish parts (ii) and (iii). Q.E.D.

An entry pro¯le is not ¯rst-best for any (c;F) if it is always dominated by another pro¯le. Lemma 1 shows the pro¯les that may survive this dominance test are the four pro¯les of the form f(g; ;); r^{0}g and the null pro¯le f(; ; ;); (; ; ;)g, which we denote as:

$$A^0 = f(; ; ;); (; ; ;)g$$
$$A^1 = f(g; ;); (; ; ;)g$$
$$A^2 = f(g; ;); (; ; g)g$$
$$A^3 = f(g; ;); (b; ;)g$$
$$A^4 = f(g; ;); (b; g)g$$

By Lemma 1 we need only consider these ¯ve pro¯les and it turns out that the ranking of each pro¯le is determined by its consumers surplus since the optimal solution will provide zero pro¯t to entrants and zero expected pro¯t to the incumbent. That is, each pro¯le is implemented most e±ciently with the lowest price cap at which the pro¯le is incentive compatible, because the producers surplus from any higher price cap would be more than o®set by the reduction in consumers surplus due to a `deadweight loss'. So, the optimal price cap for profile A^i, denoted by p^i, is the smallest solution to the following equations:

$$\begin{aligned}
A^0 &: \quad (2p^0 - 2)D(p^0) = L \\
A^1 &: \quad (2p^1 - 1)D(p^1) = F/2 + L \\
A^2 &: \quad (2p^2 - 1/2)D(p^2) = F + L \\
A^3 &: \quad (2p^3 - c)D(p^3) = F + L \\
A^4 &: \quad (2p^4 - c/2)D(p^4) = 3F/2 + L
\end{aligned} \tag{3.1}$$

In each case the left hand side of the equation is the expected aggregate sales profit and the right hand side is the common cost plus expected entry cost. The level of p^i depends on the parameter values, F and possibly c, except for A^0 where p^0 is independent of both c and F.

Since a lower price cap means higher consumers surplus, the first-best for a given $(c; F)$ is the entry profile A^i whose optimal price p^i is the lowest among all five viable profiles. Figure 1 illustrates typical regions of cost configurations $(c; F)$ in which the four profiles with possible entry ($A^1 \sim A^4$) are first-best. If $F > F^* = 2D(p^0)$ then the null profile, A^0, is first-best for all c. Verification of Figure 1 is rather lengthy and is provided in the Appendix. The intuition for the structure of Figure 1 is relatively clear. In all four cases the good firm g is always made to enter if it arrives first. Clearly, if the regulator does not wish the good firm to enter if it arrives first then the regulator must not want any entry (as indeed occurs if $F > F^*$). Where entry costs are high and there is a significant difference in quality between the good and bad firm the optimal strategy is to allow nothing other than entry by the good firm in the first period. If the good firm does not arrive until later then the costs of entry make it inefficient for it to enter since there is only one period of benefit from the entry of the good firm. As the fixed entry cost falls then it becomes sensible to allow more entry. If the production cost of the bad firm is close to the good firm then the optimal strategy is to make the first firm enter whether good or bad. Conversely, if the production cost of the bad firm, c, is closer to the incumbent than the good firm then the optimal strategy is to make the good firm enter whether it arrives first or second and to prevent the bad firm in all situations. Finally, if the costs of entry are low then it becomes sensible to force entry of either firm in the first period and to force entry of the good firm if it arrives later.

The purpose of this section is to outline the case where the regulator has complete information. She would implement the first-best if it is subgame-optimal but this is not guaranteed: once the incumbent accepts the optimal price cap for the first-best profile and lock himself in, the regulator may induce another profile that generates a higher welfare but incurs a loss to the incumbent. In this case, the incumbent would anticipate this and reject the price cap and the first-best profile would not be implementible. It turns out that this problem does not arise in the complete information regime: since any profile can be induced in such a way that the incumbent reaps the whole producers surplus, a profile that dominates the first-best in the subgame would be incentive compatible at the same price from the beginning and hence, would be preferred by the regulator (see Section A2 in the Appendix for details). Therefore, the first-best will be indeed implemented by the regulator.

Theorem 1: If the regulator observes the types of entrants, she will implement the first-best profile for each $(c; F)$ at the minimum price cap that satisfies incentive compatibility.

Proof: See Appendix, in particular, Section A2.

4. Incomplete Information.

We now consider the case where the regulator is unable to identify which firm has arrived in each period and so cannot make the access pricing rule a function of the type of entry. It can, of course, be a function of entry which is observable by the regulator. If the regulator could still implement all five of the viable profiles as considered in the previous section the optimal access pricing and its welfare implication would remain the same. Although there are some profiles that can be implemented in the same way, the inability to observe the types of entrants generally restricts the regulator's capacity to control entry using access pricing. This implies that some entry profiles are implemented less efficiently because producers surplus cannot be extracted fully while some other profiles cannot be implemented at all because the right incentives cannot be provided.

The complication for the regulator is that either of the two first candidates, once entered, will face the same revenues and access prices. There are two consequences:

(a) One is that if the regulator wishes to accommodate b in the first period and sets access prices accordingly, then g must also be accommodated in the same way if it arrives ¯rst even though g will enter at a higher access price. In these cases, the g entrant necessarily enjoys a positive pro¯t, i.e., \informational rent", which arises because of the incomplete information. This implies that such pro¯les cannot display zero expected surplus for g. Pro¯les such as A^3 are now far more expensive to implement.

(b) If the regulator wishes to encourage g to enter in the second period then it cannot stop b from also entering (in the second period) once g has entered. This follows from the complete spillover assumption: although b's entry does nothing to reduce production costs, b's production costs on entry will immediately fall to zero since g is already in the market. This means that the pro¯le A^4 is no longer implementible. Either A^4 must be replaced with the existing pro¯les (such as A^2 or A^3) or an alternative pro¯le

$$A^5 = \{(g;b);(b;g)\}$$

which recognises that $r^0 = (b;g)$ implies $r = (g;b)$. It is straightforward to verify that among the pro¯les that always dominated under complete information as per Lemma 3.1, A^5 is the only that ceases to be so under incomplete information (because A^4 is no longer implementible and hence, is eliminated from consideration).

It is obvious that the null pro¯le A^0 can be implemented in the ¯rst-best way without observing the ¯rms' type. The same is true for the pro¯les A^1 and A^2: since entry by b need be discouraged, the desired entries can be implemented via self-selection without incurring informational rent. So, the welfare level in the ¯rst-best regions for these three pro¯les is una®ected by the inability to observe the entrant's type. Given that the pro¯le A^3 is more costly to implement and A^4 cannot be implemented then this provides the intuition for the following result.

Lemma 4.1: Under incomplete information, the regulator will implement A^0; A^1 and A^2 in the ¯rst-best way in the regions in which they are ¯rst-best. In addition, A^1 and A^2 are optimal pro¯les to implement for some $(c;F)$ outside their ¯rst-best regions.

The ¯rst-best cannot be achieved for con¯gurations $(c;F)$ for which A^3 or A^4 is ¯rst-best: the optimal pro¯le to implement is the one that generates the highest welfare subject to incentive compatibility and subgame-optimality under incomplete information. Figure 2 illustrates a typical pattern of the optimal entry pro¯les under incomplete information.

The broken lines indicate the regions from Figure 1. In the light of Lemma 4.1, to justify Figure 2 we need to examine the optimal profiles for for configurations $(c; F)$ for which A^3 or A^4 is first-best, referred to as the "inefficient area."

A^3 is implementible but at a price cap above the optimal one, p^3, due to informational rent; A^4 is no longer implementible at all and A^5 need be considered instead.

For $A^3 = \langle(g; ;); (b; ;)\rangle$, the maximum possible transfer from each entrant to the incumbent (as access price payments over the two periods) is $(\hat{p} - c)D(\hat{p}) - F$, the revenue of b in excess of F, where $\hat{p}$ is the effective price cap. Then, the incumbent's incentive constraint is

$$(2\hat{p} - \frac{3}{2}c)D(\hat{p}) \geq F + L \qquad (4.1)$$

which obviously implies that the price cap must be higher than $3c/4$. Recall from (A.1) that $W_3(p)$ achieves its unconstrained maximum at $p = c/2$ and decreases monotonically for $p > c/2$. So, the optimal price cap that maximizes W_3 subject to (4.1) is the lowest one that satisfies (4.1), which we denote by p^3_*.

For $A^5 = \langle(g; b); (b; g)\rangle$, the maximum possible transfer from the second period entrant in either contingency is $\hat{p}D(\hat{p})/3 - F$. The maximum possible transfer from the first period entrant in either contingency is $(\hat{p} - c)D(\hat{p})/2 + \hat{p}D(\hat{p})/3 - F$, and g earns a positive rent if enters in the first period. So, the incumbent's incentive constraint is

$$(2\hat{p} - \frac{3}{4}c)D(\hat{p}) \geq 2F + L \qquad (4.2)$$

and, as before, the optimal price cap that maximizes W_5 subject to (4.2) is the lowest one that satisfies (4.2), which we denote by p^5_*.

Because A^3 and A^5 perform worse than the first-best in the inefficient area, we need to consider other profiles as well to determine the optimal one. Profiles $A^0; A^1$ and A^2 may not be implementible at their optimal price cap p^i in this area, because A^i may not be subgame-optimal at p^i. For each $(c; F)$, the range of price caps for which the profiles are subgame-optimal are calculated in the same way as in Section A.2 of the Appendix (because it is determined by the total surplus, not by its distribution, from the profiles implementible in the subgame) except that we now compare $A^0 \sim A^3$ and A^5. At $(c; F)$, the subgame-optimal profile for $\hat{p}$ is

$$\begin{cases} A^0 & \text{if } \hat{p} \ge D^{-1}(\frac{F}{2}) \\ A^1 & \text{if } c \ge \frac{1}{2} \text{ and } D^{-1}(\frac{F}{2}) \ge \hat{p} \ge D^{-1}(F), \text{ or } c \le \frac{1}{2} \text{ and } D^{-1}(\frac{F}{2}) \ge \hat{p} \ge D^{-1}(\frac{F}{2-2c}) \\ A^2 & \text{if } c \ge \frac{1}{2} \text{ and } D^{-1}(F) \ge \hat{p} \ge D^{-1}(\frac{2F}{1-c}) \\ A^3 & \text{if } c \le \frac{1}{2} \text{ and } D^{-1}(\frac{F}{2-2c}) \ge \hat{p} \ge D^{-1}(\frac{2F}{c}) \\ A^5 & \text{if } c \ge \frac{1}{2} \text{ and } D^{-1}(\frac{2F}{1-c}) \ge \hat{p}, \text{ or } c \le \frac{1}{2} \text{ and } D^{-1}(\frac{2F}{c}) \ge \hat{p} \end{cases} \tag{4.3}$$

To find the optimal profile in the inefficient area, we need to consider subgame-optimality (4.3) simultaneously with incentive compatibility: (4.1) for A^3, (4.2) for A^5, and (3.1) for $A^0 \sim A^2$ with the equality replaced with $\ge$, which we denoted by (3.1'). For given (c, F), each profile A^i is in one of three kinds of status:

α. The optimal price cap (p^0, p^1, p^2, p^3_* or p^5_*) exceeds the upper bound of the subgame-optimal range of price caps for A^i as specified in (4.3), so that A^i is not implementible.

β. The optmal price cap belongs to the subgame-optimal range of price caps for A^i, so that A^i is optimally implementible at the optimal price cap.

γ. The optimal price cap is below the lower bound of the subgame-optimal range of price caps for A^i, so that A^i is sub-optimally implementible at the lower bound of the range.[4]

Now we are ready to determine the optimal profile in the inefficient area: for each (c, F), we find the profiles that are implementible (optimally or sub-optimally) and compare the welfare levels from them. According to (4.3), A^2 is eligible for an optimal profile only if $c \ge \frac{1}{2}$ and A^3 is eligible only if $c \le \frac{1}{2}$. Due to the positive rent of the g entrant, A^3 is dominated by A^2 at the critical value $c = \frac{1}{2}$. So, we consider A^3 only for $c < \frac{1}{2}$.

First, we examine the inefficient area for $c \ge \frac{1}{2}$. It is easily verified from (4.3) and Lemma A.1 that A^0 and A^1 perform worse than A^2 in this area. So, we only need to compare A^2 and A^5. To do this we fix $c \ge \frac{1}{2}$ and divide the horizontal segment $0 \le F \le F_{24}(c)$ according to the status (α, β or γ) of A^2 and A^5.

Note that, being first-best at $F = F_{24}(c)$, both A^2 and A^4 are subgame-optimal at the optimal price cap $p^2 = p^4 = D^{-1}(\frac{F_{24}(c)}{1-c})$, the second equality of which follows from (A.5).

---

[4] For this to be the case the incentive constraint for A^i also need be satisfied at the lower bound. (Recall that the LHS of the incentive constraint may eventually decrease.) This condition is satisfied for subsequent analysis as is verified, for example, in the proof of Lemma 4.2.

This price cap is in the interior of the price cap range for which A^2 is subgame-optimal under incomplete information as derived in (4.3), namely, $[D^{-1}(\frac{2F_{24}(c)}{1-c}); D^{-1}(F_{24}(c))]$. As F falls from $F_{24}(c)$ to 0, the lower bound $D^{-1}(\frac{2F}{1-c})$ increases but p^2 decreases. Therefore, A^2 switches its status from β (optimally implementible) to γ (sub-optimally implementible) at F_h where

$$0 \le F_h \le F_{24}(c) \quad \text{and} \quad D^{-1}(\frac{2F_h}{1-c}) = p^2(c; F_h)$$

The range of price caps for which A^5 is subgame-optimal is $0 \le \bar{p} \le D^{-1}(\frac{2F}{1-c})$. Because incentive compatibility is harder to satisfy for A^5 than for A^4, we have $p^5_* > p^4$. At $F = F_{24}(c)$ in particular, $p^5_* > p^4 = p^2 > D^{-1}(\frac{2F_{24}(c)}{1-c})$ and so p^5 is not implementible (status α). As F falls from $F_{24}(c)$ to 0, p^5_* decreases and, therefore, A^5 becomes optimally implementible at $F = F_\ell$ and stays so for $F < F_\ell$ where

$$0 \le F_\ell \le F_{24}(c) \quad \text{and} \quad D^{-1}(\frac{2F_\ell}{1-c}) = \bar{p}^5_*(c; F_\ell)$$

Lemma 4.2: $0 < F_\ell < F_h < F_{24}(c)$ for $\frac{1}{2} \le c < 1$.

Proof: To show $F_\ell < F_h$, we take $\frac{F_{24}(c)}{2}$ as a reference point. Note that

$$D^{-1}(\frac{F_{24}(c)}{1-c}) = p^2(c; F_{24}(c)) > p^2(c; \frac{F_{24}(c)}{2})$$

In addition, since A^2 is incentive compatible at the price cap $D^{-1}(\frac{F_{24}(c)}{1-c})$ when $F = F_{24}(c)$, so it is at the same price cap when F is lower, in particular, when $F = \frac{F_{24}(c)}{2}$. Hence, A^2 is sub-optimally implementible at $F = \frac{F_{24}(c)}{2}$ and, therefore, $\frac{F_{24}(c)}{2} < F_h$. Since $p^2(c; F) > D^{-1}(\frac{2F}{1-c})$ for all $F_h < F < F_{24}(c)$, we would have proved $F_\ell < F_h$ provided that $p^5_* > p_2$ for all $\frac{F_{24}(c)}{2} < F < F_{24}(c)$. We show that this provision indeed holds.

Pick any F^0 strictly between $\frac{F_{24}(c)}{2}$ and $F_{24}(c)$. If $p^5_*(c; F^0) \ge p^2(c; F_{24}(c))$, then $p^5_*(c; F^0) > p^2(c; F^0)$ is trivial because $F^0 < F_{24}(c)$. So, suppose $p^5_*(c; F^0) < p^2(c; F_{24}(c))$. Since $p^2(c; F_{24}(c)) = p^4(c; F_{24}(c))$ and the LHS of (3.1) is quasiconcave, we have

$$(2p^5_*(c; F^0) - \frac{1}{2}c)D(p^5_*(c; F^0)) < (2p^2(c; F_{24}(c)) - \frac{1}{2}c)D(p^2(c; F_{24}(c))) = \frac{3}{2}F_{24}(c) + L \quad (4.4)$$

Subtracting (4.2) evaluated at $p^5_*(c; F^0)$ from (4.4) side by side, we get

$$\frac{1}{4}cD(p^5_*(c; F^0)) < \frac{3}{2}F_{24}(c) - 2F^0 \le F_{24}(c) - F^0 \quad (4.5)$$

where the second inequality holds because $\frac{F_{24}(c)}{2} < F^0$. On the other hand, because $c \geq \frac{1}{2}$,

$$
\begin{aligned}
(2\hat{p}^5_*(c;F^0) - \tfrac{1}{2})D(\hat{p}^5_*(c;F^0)) &\geq (2\hat{p}^5_*(c;F^0) - c)D(\hat{p}^5_*(c;F^0)) \\
&= (2\hat{p}^5_*(c;F^0) - \tfrac{3}{4}c)D(\hat{p}^5_*(c;F^0)) - \tfrac{1}{4}cD(\hat{p}^5_*(c;F^0)) \\
&> 2F^0 + L - F_{24}(c) + F^0 > F^0 + L
\end{aligned}
$$

where the last two inequalities can be verified, respectively, from (42) and (45) and from $2F^0 > F_{24}(c)$. This means that incentive compatibility for p^2 (the third equation of (31) with the equality replaced by $\geq$) is satisfied as a strict inequality at $p^5_*(c;F^0)$. Since p^2 is the smallest solution, we conclude $p^5_*(c;F^0) > p^2(c;F^0)$ as desired. Q.E.D.

For $F_\ell < F \leq F_{24}(c)$, A^2 is obviously optimal because A^5 is not implementible. For $F \leq F_\ell$, A^5 is now implementible at $\hat{p}^5_*$; A^2 is also implementible but at the price cap $D^{-1}(\frac{2F}{1-c})$. Since A^2 and A^5 are equivalent as subgame-optimal profiles at $D^{-1}(\frac{2F}{1-c})$, it follows that A^5 generates higher welfare at its optimal price cap $\hat{p}^5_*$ and hence, is optimal. So, the boundary between the optimal regions for A^2 and A^5 is $\{(c;F_\ell) : \frac{1}{2} \leq c < 1\}$. As c increases, $\hat{p}^5_*(c;F)$ rises while $D^{-1}(\frac{2F}{1-c})$ falls and hence, F_ℓ falls. As c tends to 1, $D^{-1}(\frac{2F}{1-c})$ tends to 0 and hence, F_ℓ tends to 0. This justifies the partition structure of Figure 2 for $c \geq \frac{1}{2}$.

The partition structure for $c < \frac{1}{2}$ can be justified by an analogous analysis and we sketch it here. As before, we fix $c < \frac{1}{2}$ and divide the horizontal segment $0 \leq F \leq F_{13}(c)$ according to the status of A^1, A^3 and A^5 (A^0 performs worse than A^1 in this area).

At $F = F_{13}(c)$, the optimal price cap p^1 coincides with the lower bound of the price caps for which A^1 is subgame-optimal. Therefore, A^1 is optimally implementible at $F = F_{13}(c)$, but for $F < F_{13}(c)$ it is sub-optimally implementible at the price cap $D^{-1}(\frac{F}{2(1-c)})$.

Profile A^3 is not implementible for $F = F_{13}(c)$. As F falls it switches the status to being optimally implementible, at $\tilde{F}_h$ where the optimal price cap p^3 coincides with the upper bound of the price cap range for which A^3 is subgame-optimal:

$$0 \leq \tilde{F}_h \leq F_{13}(c) \quad \text{and} \quad D^{-1}\left(\frac{\tilde{F}_h}{2(1-c)}\right) = p^3_*(c;\tilde{F}_h)$$

As F falls still further, A^3 switches the status again to being sub-optimally implementible, at $\tilde{F}_\ell$ where the optimal price cap coincides with the lower bound of the range:

$$0 \leq \tilde{F}_\ell \leq \tilde{F}_h \quad \text{and} \quad D^{-1}\left(\frac{2\tilde{F}_\ell}{c}\right) = p^3_*(c;\tilde{F}_\ell)$$

As for A^5, as F falls from $F_{13}(c)$, it switches the status from being not implementible to being optimally implementible, at $\tilde{F}_m$ where the optimal price cap p^5_* coincides with the upper bound of price caps for which A^5 is subgame-optimal:

$$0 \le \tilde{F}_m \le F_{13}(c) \quad \text{and} \quad D^{-1}\left(\frac{2\tilde{F}_m}{c}\right) = p^5_*(c;\tilde{F}_m)$$

The following relationship can be verified:

$$0 < \tilde{F}_\ell < \tilde{F}_m < \tilde{F}_h < F_{13}(c) \qquad \text{for} \qquad 0 < c \le \frac{1}{2}:$$

For $\tilde{F}_h < F \le F_{13}(c)$, A^1 is obviously optimal because A^3 and A^5 are not implementible. For $\tilde{F}_m < F \le \tilde{F}_h$, A^1 and A^3 are implementible at price caps $D^{-1}(\frac{F}{2(1-c)})$ and p^3_*, respectively, but A^5 is not. Since A^1 and A^3 generate the same level of welfare as subgame-optimal profiles at these price caps and $p^3_* \le D^{-1}(\frac{F}{2(1-c)})$, it follows that A^3 is the unique optimal profile except at $F = \tilde{F}_h$ where both A^1 and A^3 are optimal profiles at the same price cap $p^3_* = D^{-1}(\frac{F}{2(1-c)})$.

For $F \le \tilde{F}_m$, all three profiles are implementible: A^1 at $D^{-1}(\frac{F}{2(1-c)})$; A^5 at p^5_*; A^3 at $D^{-1}(\frac{2F}{c})$ if $F \le \tilde{F}_\ell$ and at $p^3_* > D^{-1}(\frac{2F}{c})$ if $\tilde{F}_\ell < F \le \tilde{F}_m$. It is intuitive that A^3 performs better than A^1. Since both A^3 and A^5 are subgame-optimal at the price cap $D^{-1}(\frac{2F}{c})$, the welfare level from A^5 at p^5_* exceeds that of A^3 at $D^{-1}(\frac{2F}{c})$, which in turn exceeds that of A^3 at p^3_* for $\tilde{F}_\ell < F \le \tilde{F}_m$. Therefore, A^5 is optimal for $F \le \tilde{F}_m$.

From the above, the boundary between the optimal regions for A^1 and A^3 is $\{(c;\tilde{F}_h) : 0 < c < \frac{1}{2}\}$, and the boundary between A^3 and A^5 is $\{(c;\tilde{F}_m) : 0 < c < \frac{1}{2}\}$. As c decreases, $p^3_*(c;F)$ falls while $D^{-1}(\frac{F}{2(1-c)})$ rises and hence, $\tilde{F}_h$ rises and converges to $2D(\hat{p}_0)$ as c tends to 0. It is intuitive that $\tilde{F}_m$ increases in c: A^3 becomes less attractive than A^5 for a larger c (because c is effective for longer in A^3), which need be cancelled off by a heavier damage to A^5 from higher F.

Finally, $F_\ell = \tilde{F}_m$ at $c = \frac{1}{2}$ is straightforward from the definitions of F_ℓ and $\tilde{F}_m$. This completes verification of Figure 2.

The following follows immediately from Figure 2:

Theorem 2: The set of cost configurations $(c;F)$ that support the possibility of both entrants joining the market when the regulator has incomplete information, A^5, is a proper subset of the set of cost configurations that support the possibility of both entrants joining the market when the regulator has complete information, A^4.

This is the sense in which we suggest that the inability to identify good and bad potential entrants will encourage a regulator to be less supportive of potential entry.

Once the regulator has incomplete information then the incentives become complex. Here we present intuitive arguments for two of these. One that better solutions exist but they are time inconsistent.[5] Second, that the good firm will often have incentives to raise their cost, i.e., become a less good competitor. We will cover these in turn.

Proposition 1 : For some configurations $(c;F)$ there exists an entry profile that exhibits higher welfare than the optimal one, but is not time consistent, i.e., not subgame-optimal.

The intuition for this result is as follows. As first-best profiles A^2 and A^3 are equivalent at the boundary between the corresponding regions. With incomplete information the welfare value of A^2 is unaffected (since b does not appear). However, in the presence of incomplete information there is a positive rent enjoyed by the good firm and so welfare is lower in A^3 compared to the complete information regime. It follows that there must be an abrupt drop at the boundary between the optimal regions for A^2 and A^3 when there is incomplete information. For values of c lower than but 'close' to 1/2, profile A^2 will generate a higher level of welfare than profile A^3. However, A^2 cannot be implemented for such a value of c because it is not subgame-optimal: That is, whatever the level of the price cap, once it is set the regulator finds A^3 a better profile to induce than A^2. Implementing A^2 rather than A^3 at this price cap would generate an insufficient return to the incumbent. Realising this, the incumbent will only accept price caps that are sufficiently high to generate enough return to the incumbent when A^3 is induced.

Proposition 2 : For some configurations $(c;F)$ the good firm has an incentive to raise its production cost, i.e., to become a less good (but still better than the other) competitor.

The intuition for this result follows from the discussion of Proposition 1. The good firm earns an information rent in A^3 that it does not earn in A^2. At the border of A^2 and

---

(5) The effect here is akin to the conventional incomplete contracts problem, e.g., Grout (1984), Hart (1995) and Hart and Holmstrom (1987).

A^3 the good firm has a discrete benefit from being in A^3 rather than A^2. The boundary between these two regions is determined by c, the relative position of the production cost of the bad firm to those of the good firm and the incumbent. The profit of the good firm increases as the bad firm improves its position relative to the good firm. At the boundary of A^3 and A^2 the good firm can change the relative position of itself compared to the bad firm by raising its own production cost. Thus the good firm has apparently perverse incentives when the regulator has incomplete information. Analogous incentives exist at the boundary between A^5 and A^2, and between A^3 and A^1.

References

Baumol, W J. and J G. Sidak (1994) The Pricing of Inputs Sold to Competitors, Yale Journal of Regulation, Volume 11.

De Fraja, G. (1997), Pricing and entry in regulated industries: The role of regulatory design', Journal of Public Economics

De Fraja, G. (1999), Regulation and access pricing with asymmetric information', European Economic Review.

Economides, N. (1996) The Economics of Networks, International Journal of Industrial Organisation, Volume 2.

Grout, P A. (1984), 'Investment and wages in the absence of binding contracts: A Nash bargaining approach', Econometrica

Grout, P A. (1996), Promoting the Superhighway: Telecommunications Regulation in Europe', Economic Policy, Volume 22.

Hart, O D. (1995) Firms Contracts and Financial Structure (Oxford University Press).

Hart, O D. and B. Holmstrom (1987) 'The theory of contracts' in T. F. Bewley (ed.) Advances in Economic Theory (Cambridge University Press).

La®ont, J J. and J. Tirole (1994), "Access Pricing and Competition", European Economic Review, 38.

La®ont, J J. and J. Tirole (1996), "Creating Competition through Interconnection: Theory and Practice", Journal of Regulatory Economics.

Appendix

Given a profile A^i, let $\bar{c}_i$ and $\bar{e}_i$ be, respectively, the mean production cost per unit and the expected number of entry. For example, $\bar{c}_1 = \bar{e}_1 = \frac{1}{2}$ and $\bar{c}_4 = c/4$ and $\bar{e}_4 = 1.5$. Then, the expected level of welfare $W_i(\hat{p})$ from A^i at a price cap $\hat{p}$ (which is binding) is:

$$W_i(\hat{p}) \equiv 2\int_{\hat{p}}^{\infty} D(p)dp + 2\hat{p}D(\hat{p}) - 2\bar{c}_i D(\hat{p}) - \bar{e}_i F - L \qquad \text{(A.1)}$$

where the first term is consumers surplus over the two periods and the rest calculates producers surplus: the total industry revenue over the two periods, $2\hat{p}D(\hat{p})$, minus the total expected industry expense that consists of expected industry production cost $2\bar{c}_i D(\hat{p})$, expected entry cost $\bar{e}_i F$, and the network operation cost L.

From the first derivative $W_i'(\hat{p}) = 2(\hat{p} - \bar{c}_i)D(\hat{p})$, it follows that the welfare from A^i monotonically decreases in $\hat{p}$ for $\hat{p} > \bar{c}_i$. Since incentive compatibility implies $\hat{p} > \bar{c}_i$ (otherwise the industry revenue does not cover total production cost, let alone the operation cost L and the entry cost), the optimal price cap subject to incentive compatibility is the lowest one that satisfies it.

Since each profile can be induced with zero surplus for all producers in the complete information regime, the first-best profile is the one with the smallest optimal price cap, i.e., the smallest solution to the corresponding equation in (3.1). In Section A.1 we derive the first-best profiles and justify Figure 1. In Section A.2 we show that the first-best profiles are subgame-optimal at their optimal prices, thereby proving Theorem 1.

In principle, the left hand side (LHS) of (3.1) may attain multiple local maxima (as functions of p), in which case the analysis is cumbersome: the optimal price cap for each profile cannot be identified as a solution to the corresponding equation at which the LHS is increasing, because there may be more than one such solutions. In this Appendix we circumvent this complication by focusing on demand functions D such that the LHS of the equations in (3.1) are quasiconcave(6) as functions of $\hat{p}$. Then, they are either single-peaked (possibly with a plateau) or monotonically non-decreasing and, therefore, a solution to each equation at which the LHS is increasing (more precisely, at immediate left of which the first derivative of the LHS is strictly positive) must be the optimal price cap. This condition is satisfied by various (demand) functions such as $D(p) = a + b/p$ and $D(p) = a + e^{-bp}$. In

(6) A function f is quasiconcave if the upper contour set $\{x : f(x) \geq r\}$ is convex for every $r \in \mathbb{R}$.

addition, we note here that the main results of this paper (such as Theorems 2 – 4) hold for a large class of demand functions that do not meet this condition, although Figure 1 may not be an exact description of the partition of the first-best regions (see Section A.3 for details). Finally, we assume that the first equation of (3.1) has a solution: otherwise some profiles are not implementible for any $(c;F)$ and as a result, no optimal profile exists for some $(c;F)$.

A.1. Derivation of first-best

We derive a series of lemmas that justify Figure 1 as the first-best regions. The optimal price caps p^1 and p^2 are functions of F and p^3 and p^4 are functions of c and F. When we need to emphasize such dependence, we specifically write as $p^1(F); p^2(F); p^3(c;F)$ and $p^4(c;F)$, or more generally, $p^i(c;F)$. These price caps may not exist for higher levels of F. Whenever applicable, subsequent statements need be understood with the qualification "if such price caps exist."

For $i; j = 0; \cdots; 4$, we say that p^i undercuts p^j at a given $(c;F)$ if the following holds: if p^j exists, so does p^i and $p^i < p^j$. Given c, let $F_{ij}(c)$ be the value of F such that $p^i = p^j$ at $(c;F_{ij}(c))$. The next lemma shows inter alia that $F_{ij}(c)$ is unique if exists, except for $F_{23}(\frac{1}{2})$ which is shown in Lemma A.2 to be any F at which $p^2(F)$ exists, because $p^2(F) = p^3(\frac{1}{2};F)$.

Lemma A.1: Consider $i < j$ such that $\{i; j\} \neq \{2; 3\}$. If $F_{ij}(c)$ exists (for a given c), then

(a) p^j exists and undercuts p^i at all $(c;F)$ with $F < F_{ij}(c)$, and

(b) p^i undercuts p^j at all $(c;F)$ with $F > F_{ij}(c)$.

Proof: By supposition, $p^i(c;F_{ij}(c)) = p^j(c;F_{ij}(c))$ which we denote by $\hat{p}$ for convenience. Also, $\hat{p}$ solves two equations of (3.1):

$$(2p - 2\acute{c}_k)D(p) = \acute{e}_k F_{ij}(c) + L \qquad (A.2)$$

for $k = i; j$, where $\acute{c}_i \geq \acute{c}_j$ and $\acute{e}_i < \acute{e}_j$. Taking the differences between these two equations (valued at $\hat{p}$) side by side, we get

$$2(\acute{c}_i - \acute{c}_j)D(\hat{p}) = (\acute{e}_j - \acute{e}_i)F_{ij}(c) \qquad (A.3)$$

The LHS of (A.2) has a negative value at $p = 0$ and increases as p increases at least up to $p^k(c;F_{ij}(c))$. For $F < F_{ij}(c)$, therefore, both $p^i(c;F)$ and $p^j(c;F)$ exist and are smaller than $\hat{p}$. Moreover, $p^j(c;F) < p^i(c;F)$ if

$$(2p^i(c;F) - 2\hat{c}_j)D(p^i(c;F)) > \hat{c}_j F + L \qquad (A.4)$$

We now prove part (a) by verifying (A.4). Note that i) $(\hat{c}_j - \hat{c}_i)F < (\hat{c}_j - \hat{c}_i)F_{ij}(c)$, and ii) $(\hat{c}_i - \hat{c}_j)D(p^i(c;F)) \geq (\hat{c}_i - \hat{c}_j)D(\hat{p})$ because $\hat{p} > p^i(c;F)$ and $D(\cdot)$ is decreasing. In conjuction with (A.3), we get

$$2p^i(c;F)D(p^i(c;F)) + 2(\hat{c}_i - \hat{c}_j)D(p^i(c;F)) > 2p^i(c;F)D(p^i(c;F)) + (\hat{c}_j - \hat{c}_i)F$$

from which we derive (A.4) because $(2p^i(c;F) - 2\hat{c}_i)D(p^i(c;F)) = \hat{c}_i F + L$.

For $F > F_{ij}(c)$, an analogous argument verifies part (b), that is, if $p^j(c;F)$ exists, so does $p^i(c;F)$ and $p^i(c;F) < p^j(c;F)$. Q.E.D.

Lemma A.2: p^2 undercuts p^3 at all $(c;F)$ with $c > \frac{1}{2}$; $p^2(F) = p^3(\frac{1}{2};F)$ for all F; p^3 undercuts p^2 at all $(c;F)$ with $c < \frac{1}{2}$.

Proof: The second assertion is immediate because the equations for A^2 and A^3 are identical if $c = \frac{1}{2}$. The first (last) assertion is easily verified because the LHS of the equation for A^2 in (3.1) exceeds (falls short of) that for A^3 while the right hand sides are the same. Q.E.D.

By Lemma A.2, we need to compare $p^0 \equiv p^2$ and p^4 for $c \geq \frac{1}{2}$ and p^0, p^1, p^3 and p^4 for $c \leq \frac{1}{2}$. We do this for the case that F_{01}, F_{12}, F_{13}, F_{24} and F_{34} exist for all relevant c. This case generates the prototype partition structure of the first-best regions. In Section A.3 we discuss possible variations in the partition when some of F_{ij} above do not exist.

Lemma A.3: (a) $F_{01}(c) = 2D(p^0)$ for all $c \in [0,1]$.

(b) If $c \geq \frac{1}{2}$, then $0 \leq F_{24}(c) < F_{12}(c) < F_{01}$; $F_{24}(c)$ is continuous and decreases in c; $F_{12}(c)$ is a constant.

(c) If $c \leq \frac{1}{2}$, then $0 \leq F_{34}(c) < F_{13}(c) \leq F_{01}$; $F_{34}(c)$ is continuous and increases in c; $F_{13}(c)$ is continuous and decreases in c.

Proof: (a) The optimal price cap p^0 for A^0 is a constant. Since p^0 also solves the second equation of (3.1) at $(c;F_{01}(c))$, part (a) follows easily.

(b) Let $\hat{p}$ be the common solution to the second and third equations of (31) at $F_{12}(c)$. Multiply the former equation by 2 and subtract the latter side by side to verify that $\hat{p}$ is a solution to $(2p - 1.5)D(p) = L$. In fact, $\hat{p}$ is the smallest solution to this equation: otherwise $\hat{p}$ would not be the smallest solutions to the second and third equations of (31), either. Obviously, $\hat{p}$ is independent of c, and $\hat{p} < p^0$ follows from $(2p - 2)D(p) < (2p - 1.5)D(p)$.

Now, by taking differences between the second and third equations of (31), note $F_{12}(c) = D(\hat{p})$. So, $F_{12}(c)$ is a constant. Furthermore, $F_{12} < F_{01}$: the value of the LHS of the second equation of (31) at $\hat{p}$ is $F_{12}/2 + L$ which is lower than that at p^0, namely $F_{01}/2 + L$, because $\hat{p} < p^0$ and the LHS is increasing up to p^0.

By an analogous argument, $F_{24}(c) = (1 - c)D(\tilde{p})$ where $\tilde{p}$ is the smallest solution to $(2p + c - 1.5)D(p) = L$. So, $\tilde{p} < \hat{p}$. That $F_{24}(c)$ is continuous is trivial. That it decreases in c can be shown by calculus, but an intuition suffices: A^4 becomes less attractive than A^2 for a larger c, which need be compensated by lower F so as to make A^4 stay as attractive as A^2. Finally, $F_{24}(c) < F_{12}$: the value of the LHS of the third equation of (31) at $\tilde{p}$ is $F_{24}(c) + L$ which is lower than that at $\hat{p}$, namely $F_{12} + L$, because $\tilde{p} < \hat{p}$ and the LHS is increasing up to $\hat{p}$.

(c) The proof for part (c) consists of essentially the same (albeit slightly more involved) arguments as those used to prove part (b), which we omit here. Q.E.D.

We use these lemmas to verify Figure 1. For $c \geq \frac{1}{2}$, we compare $A^0; A^1; A^2$ and A^4. From Lemmas A1 and A3 and transitivity of "undercutting" relationship, we deduce that the first-best profiles are: A^4 with $p^4(c; F)$ for $F \leq F_{24}(c)$; A^2 with $p^2(F)$ for $F_{24}(c) \leq F \leq F_{12}$; A^1 with $p^1(F)$ for $F_{12} \leq F \leq F_{01}$. Moreover, the boundary between A^4 and A^2, the graph of $F_{24}(c)$, is a downward sloping curve joining conguraion points $(0; 1)$ and $(F_{24}(\frac{1}{2}); \frac{1}{2})$; the boundary between A^2 and A^1 is the vertical line at $F = F_{12}$; the boundary between A^1 and A^0 is the vertical line at $F = F_{10}$. The first-best profiles for $c \leq \frac{1}{2}$ are analogously deduced and they conform to Figure 1. The first-best regions for $c \geq \frac{1}{2}$ and $c \leq \frac{1}{2}$ are joined without slip (i.e., $F_{24}(\frac{1}{2}) = F_{34}(\frac{1}{2})$ and $F_{12} = F_{13}(\frac{1}{2})$) because A^2 and A^3 are equivalent with the same optimal price cap at $c = \frac{1}{2}$.

A2. First-best profiles are implementible

Since $D(\hat{p})$ is higher for a lower $\hat{p}$, we reason from (A1) that the lower the price cap, the more important $\hat{e}_i$ is relative to $\acute{e}_i$ in improving the welfare, and vice versa. That is, as

$\hat p$ increases the subgame-optimal profile switches from the ones with low $\hat e_1$ such as A^4, to the ones with low $\hat e_1$ such as A^0. By comparing $W_0 \sim W_4$ [7] for various $\hat p$, we verify that the ranges of price caps $\hat p$ for which each profile is subgame-optimal are as follows:

At $(c;F)$, the subgame-optimal profile for $\hat p$ is

$$\begin{cases} A^0 & \text{if } \hat p \geq D^{-1}(\frac{F}{2}) \\ A^1 & \text{if } c \geq \frac{1}{2} \text{ and } D^{-1}(\frac{F}{2}) \geq \hat p \geq D^{-1}(F), \text{ or } c \leq \frac{1}{2} \text{ and } D^{-1}(\frac{F}{2}) \geq \hat p \geq D^{-1}(\frac{F}{2-2c}) \\ A^2 & \text{if } c \geq \frac{1}{2} \text{ and } D^{-1}(F) \geq \hat p \geq D^{-1}(\frac{F}{1-c}) \\ A^3 & \text{if } c \leq \frac{1}{2} \text{ and } D^{-1}(\frac{F}{2-2c}) \geq \hat p \geq D^{-1}(\frac{F}{c}) \\ A^4 & \text{if } c \geq \frac{1}{2} \text{ and } D^{-1}(\frac{F}{1-c}) \geq \hat p, \text{ or } c \leq \frac{1}{2} \text{ and } D^{-1}(\frac{F}{c}) \geq \hat p \end{cases} \tag{A.5}$$

It is straightforward to verify from (A.5) that for each $(c;F)$ the first-best profile is indeed subgame-optimal at the optimal price. We show this for A^2 here. Exactly analogous arguments work for other profiles.

Consider $(c;F)$ for which A^2 is first-best, that is, $c \geq \frac{1}{2}$ and $F_{24}(c) \leq F \leq F_{12}(c)$. Then,

$$p^2(c;F_{24}(c)) \leq p^2(c;F) \leq p^2(c;F_{12}). \tag{A.6}$$

Since $p^2(c;F_{12})$ solves the second and third equations of (3.1) at $(c;F_{12})$, we have $D(p^2(c;F_{12})) = F_{12}$ and so $p^2(c;F_{12}) = D^{-1}(F_{12}) < D^{-1}(F)$. Similarly, $p^2(c;F_{24}(c))$ solves the third and fifth equations of (3.1) at $(c;F_{24}(c))$, from which we deduce $p^2(c;F_{24}(c)) = D^{-1}(\frac{F_{24}(c)}{1-c}) > D^{-1}(\frac{F}{1-c})$. In conjuction with (A.6), $p^2(c;F)$ is in the range specified in (A.5) for which A^2 is subgame-optimal.

A.3. Variations of the partition structure

We justified the first-best partition in Figure 1 for the case that all relevant $F_{ij}(c)$ exist. In this case, the graph of $F_{ij}(c)$ forms the boundary between A^i and A^j. If this is not the case, the parition structure is altered but the main features are retained.

Consider $F_{24}(c)$ for example. $F_{24}(c)$ necessarily exists for sufficiently high $c < 1$, converging to $F_{24}(1) = 0$. For lower c, however, it may happen that $p^4(c;F)$ exists precisely when F is not too large, say $F \leq G_4(c)$, and whenever it exists it undercuts $p^2(c;F)$, in which case $F_{24}(c)$ does not exist. Suppose this is the case for $c < \hat c$ where $\hat c > \frac{1}{2}$. Then,

(7) The profiles shown to be dominated in Lemma 3.1 are also dominated in the subgame after $\hat p$ is accepted and so, need not be considered.

the boundary of A^4 starts from the configuration point $(1;0)$ downward along $F_{24}(c)$ until $c = \hat{c}$; then it continues along the graph of $G_4(c)$ which is also downward sloping until it meets the graph of $F_{34}(c)$; at this point it turns around and continues along $F_{34}(c)$ to the point $(0;0)$. The area to the right of this boundary and to the left of F_{12} and $F_{13}(c)$, is divided between A^2 and A^3 with a border along $c = \frac{1}{2}$. (For some $(c;F)$ in this area sufficiently close to $G_4(c)$, the optimal price cap p^2 or p^3 may be below the subgame-optimal range specified in (A5), in which case A^2 or A^3 need be implemented at a price cap above the optimal one, generating a positive producers surplus.)

For another example, suppose that $p^2(F)$ exists precisely when F is not too large, say $F \leq G_2$, and whenever it exists it undercuts $p^1(F)$. In this case, F_{12} does not exist and the boundary between A^1 and A^2 is a vertical line at $F = G_2$ for $c \geq \frac{1}{2}$; the boundary between A^1 and A^3 starts from $(\frac{1}{2};G_2)$ and continues along the graph of $G_3(c)$ which is downward sloping and penetrates the lower-left corner of the first-best region of A^0 ($G_3(c)$ is the maximum F at which p^3 exists). For a simpler example, F_{01} does not exist if $p^1(F)$ exists precisely when $F \leq G_1$, and is lower than p^0: in this case the boundary between A^0 and A^1 is a vertical line at $F = G_1$.

It is easy to see that A^0 is first-best for sufficiently large F; A^4 is first-best for sufficiently low c; A^2 is first-best for sufficiently high c and low F; A^3 is first-best for sufficiently low c and F. But A^1 may not be first-best for any $(c;F)$: this is so if p^1 is undercut by either p^2 or p^3 whenever p^1 undercuts p^0. In this case the partition looks like Figure 1 with the change that A^0 replaces A^1. The boundary between A^0 and A^2 (A^3) is either F_{02} or G_2 ($F_{03}(c)$ or $G_3(c)$).

Finally, further variation of the partition structure results when quasiconcavity of the LHS of (31) is relaxed. However, the relative position of the first-best regions remain the same. One notable variation is that the boundary between A^4 and A^2 and that between A^4 and A^3 may extend to meet the first-best region for A^1, in which case the first-best regions for A^2 and A^3 are not adjacent.