Vertical Exclusion with Downstream Risk Aversion*

Stephen Hansen
Universitat Pompeu Fabra
Barcelona GSE
stephen.hansen@upf.edu

Massimo Motta
ICREA-UPF
Barcelona GSE
massimo.motta@upf.edu

December 18, 2011

PRELIMINARY DRAFT

Abstract

We examine a vertical market in which downstream firms have private information about their productivity and compete for consumers, and an upstream firm posts a public contract for each firm. When downstream firms are risk-neutral, the upstream firm offers all of them the input, but when they are sufficiently risk averse it sells to one. This eliminates the payment of a risk premium to compensate for the uncertainty created by the potential rival's hidden cost type. Thus exclusion arises when contracts are fully observable and downstream firms are ex ante symmetric. Fixed wage contracts provide insurance, but their welfare consequences are ambiguous.

Keywords: Exclusive Contracts, Adverse Selection, Risk

JEL Codes: D82, L22, L42

*We would like to thank Larbi Alaoui for extensive comments, as well as Jan Eeckhout, Patrick Rey and participants of the ANR-DFG workshop (Paris) for helpful advice.

1 Introduction

Very often, a manufacturer has to decide whether to sell its products through one or several retailers, a franchisor whether to have one or multiple franchisees, the owner of a patent whether to license its technology to one or more licensees. As a result, exclusive clauses may be signed whereby in a given geographical area only one agent will deal with the principal's product, brand, or technology. Such exclusivity clauses are a form of vertical restraint which has received the attention of antitrust agencies and courts,[1] as well as raised the interest of economic and legal scholars. The literature has so far identified pro-competitive and anti-competitive reasons for such exclusive clauses, which effectively foreclose all but one downstream agent from the market.

On the pro-competitive side, exclusive territorial protection clauses may have the effect of promoting investments by retailers. For instance, when local advertising and promotion activities matter, there may be under-provision of such activities because of free-riding among retailers of the same brand. By allocating exclusive selling rights in a given region to only one retailer, incentives to invest are restored.

Probably, the main reason why exclusive territories may have anticompetitive effects is due to Hart and Tirole (1990) (see also Rey and Tirole (2007)).[2] When contract offers are unobservable a monopolistic manufacturer may not be able to earn monopoly profits: anticipating its incentives to improve contract terms to any retailer at the expenses of the others, retailers' willingness to pay for the product lowers. The manufacturer can solve its commitment problem—conceptually similar to that of a durable good monopolist—by assigning exclusive rights to one dealer. As a result, it will be able to implement the monopolist prices and profits, but to the detriment of consumers.

We revisit the problem of an upstream firm to trade with one or several retailers without relying on any of the above-mentioned motives for exclusive clauses. In particular, while Hart and Tirole (1990) assume that contracts are private information, we show that exclusivity can arise when contracts are public information, but downstream firms have private information and are risk-averse.

More precisely, we analyze an adverse selection model where two downstream firms

[1]In the US, the Supreme Court has moved from a per se prohibition rule of exclusive territorial protection (*US v. Arnold Schwinn & Co.*, 1967) to a rule of reason (*Continental TV v. GTE Sylvania*, 1977), although it is probably fair to say that it is unlikely that such practices would be found today in violation of the Sherman Act. In the EU, such restraints would be allowed but subject to certain provisions; for instance, a manufacturer could not prohibit its retailers in one country from selling to unsolicited customers from another country.

[2]Exclusive territories may harm welfare also through strategic interbrand effects, as identified by Rey and Stiglitz (1995). By giving exclusivity to one retailer, a manufacturer will suppress intra-brand competition and induce a less aggressive price of its brand, in order to relax inter-brand competition. While industry profits will increase, consumers will end up paying higher prices, and total welfare will decrease.

(the "agents") competing for final consumers may differ in (privately-known) production costs. The upstream firm (the "principal") posts a public menu of contracts designed for each cost type. When downstream firms are risk-neutral, both of them will be offered the input at equilibrium, because resorting to multiple agents improves the manufacturer's profits.[3]

But when agents are sufficiently risk-averse, the upstream firm will optimally sell to only one firm, and exclude the other. This is because each firm's payoff is affected by the uncertainty about the cost of the rival, and it will want a rent which compensates for the risk of having to compete with a low-cost rival. By supplying only one downstream firm, the source of uncertainty is removed and the principal is better off, despite losing the benefit of having several agents.[4]

The most important assumption in the model is the ruling out of communication between the downstream firms and upstream firm prior to the posting of the contracts. By making this assumption, we depart from the well-established literature in which the principal selects one among the privately-informed agents by making them compete for the incentive contracts (see Laffont and Tirole (1987); McAfee and McMillan (1986); McAfee and McMillan (1987); Riordan and Sappington (1987)). Undoubtedly, there will be many real-world situations in which an auction can be easily and speedily designed and/or the value of the contract will justify the costs of organizing the auction (large procurement contracts being a case in point). But there will also be other situations in which a firm will not want to rely on an auction when choosing its retailers, licensees or franchisees. Our paper has something to say about the latter environment.[5]

Another crucial assumption in our model is that agents may be risk averse. To the extent that agents are distributors or retailers, or more generally small firms which are unable to diversify risk, it will probably not take much convincing that this assumption is realistic. But there are several reasons why even larger firms may be reluctant to take risk, starting from the fact that actual decisions are taken by managers who, being individuals, may well be risk averse. Nocke and Thanassoulis (2010) also endogenously explain risk aversion when firms are credit constrained. Asplund (2002) also mentions

[3] In our model with cost uncertainty, the principal benefits from having multiple agents because this helps smoothing output (intuitively, the principal reduces uncertainty by contracting with both firms. But one can think of other mechanisms, such as product differentiation, whereby contracting with both downstream firms raises profits.

[4] We shall also show that for intermediate risk preferences partial exclusion arises, with one downstream firm being offered no trade if it is high cost, and lower trade than the other firm if it is low cost.

[5] Segal (1999) analyses different principal-agent models where there are externalities among agents. The simplest setting he studies (section III) is one where the principal commits to a set of publicly observable bilateral contract offers to agents. In section VI he studies the case of "general commitment mechanisms" where the principal is allowed to commit to mechanisms in which one agent's trade can be made contingent on other agents' messages. Our paper focuses on the former setting.

empirical research pointing to the risk aversion of firms.

The paper is organized thus. Section 2 describes the general model, which is then solved first for the extreme cases of risk-neutral and infinitely risk-averse agents in section 3. Section 4 then proposes a way of modelling intermediate risk aversion, characterizes optimal contracts under general demand, and solves for optimal contracts under linear demand. Section 5 then examines the case where the principal insures both downstream agents via fixed wage contracts (that we interpret as vertical integration), and shows that such contracts have ambiguous welfare consequences. Section 6 concludes the paper.

All proofs are relegated to the Appendix.

2 Model

We consider an industry in which a risk-neutral upstream firm M supplies an input that is transformed in a one-to-one relationship by two downstream firms $i = 1, 2$ whose product is homogenous. Aggregate demand for the product is $P(Q)$, where $Q \geq 0$ is aggregate quantity. We assume that $P'(Q) < 0$; that marginal revenue $P(Q) + QP'(Q) = \text{MR}(Q)$ is decreasing; and that there exists some finite quantity $\widetilde{Q}$ at which $\text{MR}(\widetilde{Q}) = 0$.

The downstream firms, which may be risk neutral or risk averse, are heterogenous in their productivity. Each has a constant marginal cost of production $c_i \in \{0, c\}$ where $c > 0$, $\Pr[c_i = 0] = r$, and c_1 and c_2 are independent. We also assume that c_i is private information for firm i.[6] The constant returns to scale embedded in the constant marginal cost assumption keeps aggregate production costs independent of the number of firms in the market, allowing one to focus on revenue volatility as the main driver of the exclusion results. To ensure high cost production is profitable we impose $P(0) > c$.

M offers the downstream firms contract menus $T_i(Q_i)$, where $Q_i \geq 0$ is the amount of input that firm i uses (and which it converts to output Q_i) and $T_i(Q_i) \in \mathbb{R}$ is the transfer that firm i pays to M for using Q_i. $(0, 0)$ is included as an element of $T_i(Q_i)$. The fact that in general $T_1 \neq T_2$ for the same input level means that M can price discriminate between the firms. Unlike in Hart and Tirole (1990), we assume that these contracts are publicly observable. We assume that downstream firms' cost types are realized *after* M posts the contract menus (see discussion in introduction about the ruling out of pre-contractual communication), after which they play a simultaneous game of incomplete information. When a pure strategy equilibrium exists, its outcome is $\{Q_i(c_i), T_i(c_i)\}_{i=1}^2$, or a quantity and transfer choice made by each cost type of each downstream firm. Using the standard revelation principle argument, one can without loss of generality focus on M offering

[6] An alternative interpretation of the model is that c_i represents a shock to a local demand intercept that reduces the price that firm i's customers are willing to pay for any given amount of aggregate output.

each downstream firm a two-point, incentive-compatible contract menu. To formalize this idea, one can denote the menu offered to firm i as $[Q_i(\widehat{c}_i), T_i(\widehat{c}_i)]$ for $\widehat{c}_i \in \{0, c\}$. Here $\widehat{c}_i$ corresponds to the cost type that firm i ''reports" at the stage it must choose an element from the menu.[7] Let

$$\pi_i(\widehat{c}_i, \widehat{c}_j, c_i) = Q_i(\widehat{c}_i)\left\{P\left[Q_i(\widehat{c}_i) + Q_j(\widehat{c}_j)\right] - c_i\right\} - T_i(\widehat{c}_i) \tag{1}$$

be firm i's profit from reporting cost type $\widehat{c}_i$ when firm $j \neq i$ reports cost type $\widehat{c}_j$ and firm i has marginal cost c_i. Here one can see the externality in the model: firm j's choice of $\widehat{c}_j$ affects π_i through the market price. Suppose it is common knowledge that firm j reports $\widehat{c}_j = 0$ with probability r and $\widehat{c}_j = c$ with probability $1 - r$. This induces the lottery

$$L_i(\widehat{c}_i \mid c_i) = \left\{\left[\pi_i(\widehat{c}_i, 0, c_i), \pi_i(\widehat{c}_i, c, c_i)\right]; (r, 1 - r)\right\} \tag{2}$$

for firm i. Let U be the common utility function over such lotteries. This function embeds downstream firms' risk preferences, for which we will provide specific functional forms below.

M's problem can be expressed as

$$\max_{\{Q_i(\widehat{c}_i), T_i(\widehat{c}_i)\}_{i=1}^2} \sum_{i=1}^{2} rT_i(0) + (1 - r)T_i(c) \text{ such that} \tag{3}$$

$$U\left[L_i(0 \mid 0)\right] \geq U[0] \tag{PC_i^L}$$

$$U\left[L_i(c \mid c)\right] \geq U[0] \tag{PC_i^H}$$

$$U\left[L_i(0 \mid 0)\right] \geq U\left[L_i(c \mid 0)\right] \tag{IC_i^L}$$

$$U\left[L_i(c \mid c)\right] \geq U\left[L_i(0 \mid c)\right]. \tag{IC_i^H}$$

Here there are eight constraints corresponding to participation (denoted PC) and incentive compatibility (denoted IC) constraints for each cost type of each downstream firm. $U[0]$ corresponds to the utility from receiving the certain wealth level 0. We can focus without loss of generality on contracts that induce participation since $(0, 0)$ is an element of each contract. Since optimal contracts are incentive compatible, one can interpret M as posting contract menus but not actually observing which element is chosen until after each downstream firm has undertaken production. For simplicity we sometimes refer to a contract menu simply as a contract. We want to establish the conditions under which

[7] Here, the use of the word report should be treated with care. Since we are ruling out communication prior to the posting of contracts, this is not a literal report in the sense of the mechanism design literature, but simply a convenient device for representing the two-point contract menus that M posts. In general, the principal will propose firm i a menu which will contain a contract intended for low-cost firm i and a contract intended for high-cost firm i. We say that firm i "report" a low cost (respectively, high cost) when it chooses the former contract over the latter (respectively, vice versa).

M chooses to deal with one or both of the downstream firms. Let $\{Q_i^*(\widehat{c}_i), T_i^*(\widehat{c}_i)\}_{i=1}^2$ be a solution to (3). We say that firm i is *excluded* if $Q_i^*(0) = Q_i^*(c) = 0$. The rest of the paper studies how exclusion depends on U.

To summarize, the timing of the game is the following:

1. The upstream firm M posts a contract for each downstream firm i
2. Nature draws c_i, which i learns
3. Downstream firms order input, commit to pay the corresponding transfer, and produce output
4. The output market clears, downstream firms' profits are realized, and they pay the transfer to the upstream firm

The assumption on the timing of the transfer payments is not essential, but we make it to ensure that neither downstream firm observes the action of the competitor. Observing the competitor's action would in fact not violate incentive compatibility in our model, but would violate the participation constraints. Here we have also assumed that downstream firms are not credit-constrained and can pay the agreed transfer even if doing so means negative profits ex-post. With the basic structure of the model in place, we now solve it for various specifications of the utility function U.

3 Exclusion and Attitudes Towards Risk

To begin our analysis of the relationship between risk aversion and exclusion, we consider the extreme cases of risk neutrality and infinite risk aversion. Here we show that under risk neutrality, the upstream firm gains from dealing with both firms, while under infinite risk aversion it prefers to exclude one of the two firms.

3.1 Risk Neutrality

When downstream firms are risk neutral, their utility from facing the lottery induced by the competitor selling in the same market is simply the expected profit, so that

$$U\left[L_i(\widehat{c}_i \mid c_i)\right] = r\pi_i(\widehat{c}_i, 0, c_i) + (1-r)\pi_i(\widehat{c}_i, c, c_i). \tag{4}$$

From here, one can simplify M's optimization problem in (3) using standard arguments (for example, those in Chapter Two of Laffont and Martimort (2002)) to show that, in the optimal contract menu, the participation (incentive compatibility) constraint of each high (low) cost type binds; the participation (incentive compatibility) constraint of each low (high) cost type is slack; and the only incentive compatible contracts are those in which low cost types sell more than high cost types. These conditions pin down the resulting transfers that M can ask of downstream firms.

Lemma 1 *The optimal contract under risk neutrality satisfies* $Q_i^*(0) \geq Q_i^*(c)$ *and*

$$T_i^*(0) = Q_i^*(0)\left\{rP\left[Q_i^*(0)+Q_j^*(0)\right]+(1-r)P\left[Q_i^*(0)+Q_j^*(c)\right]\right\}-cQ_i^*(c) \quad (5)$$

$$T_i^*(c) = Q_i^*(c)\left\{rP\left[Q_i^*(c)+Q_j^*(0)\right]+(1-r)P\left[Q_i^*(c)+Q_j^*(c)\right]-c\right\}. \quad (6)$$

The transfer asked of each high cost firm—$T_i^*(c)$—is simply its expected profit; on the other hand, the transfer asked of each low cost firm—$T_i^*(0)$—is its expected profit minus an information rent equal to the total cost of the high cost type's output.

The relationship between M's profits and the distribution of output across the two downstream firms is more transparent if one rewrites the contract variables. Let $Q^H = \sum_{i=1}^{2} Q_i(c)$ be the total production of high cost firms; let $\Delta_i = Q_i(0) - Q_i(c)$ be the difference between the quantity produced by the low and high cost types of downstream firm i, which we will call the *low cost output gap* of firm i; and let $\Delta = \Delta_1 + \Delta_2$ be the aggregate low cost output gap. When these variables carry asterisk superscripts, they should be understood to represent optimal values. One can then write the profit of M as:[8]

$$\begin{aligned}&r^2\left(Q^H+\Delta\right)P\left(Q^H+\Delta\right)+(1-r)^2Q^HP\left(Q^H\right)-cQ^H+\\&r(1-r)\left[\left(Q^H+\Delta_1\right)P\left(Q^H+\Delta_1\right)+\left(Q^H+\Delta-\Delta_1\right)P\left(Q^H+\Delta-\Delta_1\right)\right].\end{aligned} \quad (7)$$

The terms $\left(Q^H+\Delta\right)P\left(Q^H+\Delta\right)$ and $Q^HP\left(Q^H\right)$ are the total revenue that M gets conditional on there being two low cost or high cost firms, respectively. Neither term depends on the distribution of output between the downstream firms since only aggregate production quantities influence the quantity and price. Similarly, the term cQ^H reflects aggregate production costs and aggregate information rents, which again do not depend on the distribution of high cost output between the firms. In fact, the distribution of output between the two firms only depends on the term

$$\left(Q^H+\Delta_1\right)P\left(Q^H+\Delta_1\right)+\left(Q^H+\Delta-\Delta_1\right)P\left(Q^H+\Delta-\Delta_1\right), \quad (8)$$

which is the total revenue conditional on there being one high cost and one low cost firm downstream. As such, we will refer to (8) as *revenue with heterogeneity.*

The important distributional variables affecting the level of revenue with heterogeneity are the low cost output gaps. For a fixed level of the aggregate low cost output gap Δ,

[8]This results from computing the expression $r\left[T_1(0)+T_2(0)\right]+(1-r)\left[T_1(c)+T_2(c)\right]$ using the values of the transfers given by (5) and (6).

the optimal gap of firm 1 is defined by the condition

$$\mathrm{MR}\left(Q^H + \Delta_1^*\right) = \mathrm{MR}\left(Q^H + \Delta - \Delta_1^*\right). \tag{9}$$

In other words, the optimal allocation of Δ across firms 1 and 2 equates the marginal revenue from increasing the low cost output gap for each firm. To see why, suppose we begin with a situation in which $\Delta_1 = 0$. For simplicity, assume as well that $\mathrm{MR}\left(Q^H\right)$ and $\mathrm{MR}\left(Q^H + \Delta\right)$ are both positive. Now, increasing Δ_1 has two offsetting effects. First, it increases total revenue when firm 1 is low cost and firm 2 is high cost. At the same time, it decreases total revenue when the opposite is true. But, because marginal revenue is decreasing in aggregate quantity, the first effect dominates the second. One can use similar arguments for any allocation of the aggregate low cost output gap for which $\Delta_1 \neq \Delta_2$: there is always a possibility to increase revenue with heterogeneity by distributing Δ more evenly. Of course, one needs to make sure that the upstream firm actually wants to contract low cost firms to produce more than high cost firms. The following result establishes this.

Proposition 1 *In the optimal contracts,* $\Delta^* > 0$ *and* $\Delta_1^* = \Delta_2^* = \frac{\Delta^*}{2}$.

The optimal value of Δ is positive because, starting from $\Delta = 0$, increasing Δ a small amount has a positive effect on revenue without generating any additional costs (both the value of information rents and the total production cost depend only on the high cost output, not the low cost).

Proposition 1 has an important implication for exclusion, the focus of this paper. Since the optimal Δ is positive and the optimal distribution of Δ between the firms provides each an equal share, the upstream firm is strictly better off offering each firm a part of the market and never chooses to exclude either.

Corollary 1 *In the optimal contracts, neither firm is excluded.*

Notice that this finding breaks the indeterminacy of the optimal number of firms in complete information models of vertical markets in which the upstream firm can fully commit to contracts. In such models there is one known monopoly quantity that maximizes the upstream firm's profits, and it can be distributed arbitrarily among any number of downstream firms. Here the upstream firm is uncertain about the optimal quantity: it can either be high or low depending on the realizations of the downstream firms' cost types. Having two firms in the market helps it "hedge its bets" by making sure that when one of the two firms is the low cost type it gets a piece of the market.

To see the connection between two firms and revenue uncertainty more clearly, the linear demand case is helpful. In this situation, aggregate revenue is given by $\mathbb{E}[\,Q(1-Q)\,] =$

$\mathbb{E}[\,Q\,] - \mathbb{E}[\,Q\,]^2 - V[\,Q\,]$. So distributing output between the two firms should be done to decrease the variance in aggregate output. This is not because M is risk averse (in fact, it is risk-neutral), but because aggregate revenue is concave in aggregate output. Now, firm i's output in an incentive compatible contract is the random variable $Q_i = Q_i(c) + \widetilde{x}_i \Delta_i$ where $\widetilde{x}_i$ is a Bernoulli random variable with mean r and variance $r(1-r)$. So $V[\,Q\,] = r(1-r)\sum_i \Delta_i^2$, which is clearly minimized by equating Δ_i across firms. In other words, equal low cost output gaps maximally reduce aggregate output volatility, which makes M better off.

Proposition 1 pins down the distribution of Δ across downstream firms, but not that of Q^H. As long as $\Delta_1 = \Delta_2$ holds, any split of Q^H between the firms is optimal. In particular, asymmetric contracts are optimal. The important point is that both firms have a positive probability of producing under the optimal contract menus.

3.2 Infinite Risk Aversion

If serving two firms is useful for the upstream firm because of a reduction in the uncertainty about the aggregate output level, the opposite is true for the downstream firms. If a downstream firm knows that it alone produces, it knows for certain what will be its profits conditional on producing. On the other hand, when it knows that the other firm is offered a menu with different output levels for high and low cost type realizations, its profit conditional on producing is uncertain. In the case of risk neutrality, this has no effect since downstream firms are happy to pay a transfer equal to expected profit in order to enter the market. In reality, however, one might imagine that downstream firms have some aversion to the uncertainty that competition creates. To begin the analysis of how this affects the optimal menus, we make the extreme assumption—which we later relax—that downstream firms are *infinitely* risk averse in the sense that their utility from a lottery is its worst realization:

$$U\left[L_i(\widehat{c}_i \mid c_i)\right] = \min\{\pi_i(\widehat{c}_i, 0, c_i), \pi_i(\widehat{c}_i, c, c_i)\}. \tag{10}$$

One can follow similar steps as in the risk neutral case to show that the optimal contract makes the participation constraints of high cost types bind and the incentive compatibility constraints of low cost types bind, leading to the following characterization of the optimal menus.

Lemma 2 *The optimal contract under infinite risk aversion satisfies* $Q_i^*(0) \geq Q_i^*(c)$ *and*

$$\begin{aligned} T_i^*(0) &= Q_i^*(0) P\left[Q_i^*(0) + Q_j^*(0)\right] - c Q_i^*(c) \\ T_i^*(c) &= Q_i^*(c)\left\{P\left[Q_i^*(c) + Q_j^*(0)\right] - c\right\}. \end{aligned}$$

The expressions in lemma 2 differ from those in lemma 1 in that downstream firms' certainty equivalent income corresponds to the profit level realized when facing a low cost competitor. Because the upstream firm contracts low cost firms to produce more output than high cost firms, the profit for a downstream firm, no matter its cost type, is lower when it meets a low cost competitor since the market price falls with increased output. At the same time, low cost types continue to receive an information rent of $cQ_i^*(c)$, just as in the risk neutral case. Replacing the transfers from lemma 2 into the upstream firm's profit gives

$$r\left(Q^H+\Delta\right)P\left(Q^H+\Delta\right)+(1-r)Q_1(c)P\left(Q^H+\Delta_2\right)+(1-r)Q_2(c)P\left(Q^H+\Delta_1\right). \quad (11)$$

The total transfers that M collects from low cost firms, given by

$$\left(Q^H+\Delta\right)P\left(Q^H+\Delta\right), \quad (12)$$

does not depend on the distribution of output between them: each low cost firm is only willing to pay the profit realized when the highest possible aggregated quantity is realized, and it does not matter whether itself or the competitor contributes to this output level. On the other hand, the total transfers collected from high cost firms, given by

$$Q_1(c)P\left(Q^H+\Delta_2\right)+Q_2(c)P\left(Q^H+\Delta_1\right), \quad (13)$$

does depend on the distribution of output. We will refer to (13) as *worst case high cost revenue*. For a fixed level of Q^H and Δ, M's optimal distribution of output between downstream firms will maximize worst case high cost revenue.[9] From observing (13), one can fairly quickly see that its upper bound is $Q^H P(Q^H)$.[10] This has a natural interpretation: maximizing worst case high cost revenue means eliminating the worst case, i.e. giving the whole downstream market to just one firm. When just one firm produces, (13) achieves its upper bound of $Q^H P(Q^H)$ because the high cost firm knows that its output alone will determine the output price. In this case, an exclusive contract is valuable to the upstream firm because it keeps it from having to pay *any* risk premium to high cost producers. Of course, an implicit assumption in this discussion is that M actually wants to contract high cost firms to produce positive output levels, but one can show that when r is small this is indeed the case.

Proposition 2 *There exists an $r^* \in (0,1)$ such that for $i,j=1,2$, $i \neq j$:*

[9] Note that aggregate production costs and information rents equal cQ^H, which also does not depend on the distribution of output between firms.

[10] To see this, note that $P\left(Q^H\right) \geq \max\{P\left(Q^H+\Delta_2\right), P\left(Q^H+\Delta_1\right)\}$, so that (13) is less than $Q_1(c)P\left(Q^H\right)+Q_2(c)P\left(Q^H\right)=Q^H P(Q^H)$.

1. *Whenever* $r < r^*$ *the optimal contracts are such that* $Q_i^*(0) > 0$, $Q_i^*(c) > 0$, *and* $Q_j^*(0) = Q_j^*(c) = 0$.

2. *Whenever* $r \geq r^*$ *the optimal contracts are such that* $Q_i^*(c) = Q_j^*(c) = 0$ *and* $Q_i^*(0) + Q_j^*(0) = \Delta^* > 0$.

An immediate implication is that

Corollary 2 *Under infinite downstream risk aversion, exclusion of one downstream firm always solves M's profit maximization problem, and whenever* $r < r^*$, *exclusion is the only solution to its problem.*

This result contains the basic message of the paper. Even if the upstream firm can fully commit to contracts, it may simply be too costly to include both firms in the downstream market if they are sufficiently risk averse. In the infinite risk aversion case, the upstream firm does not gain by smoothing aggregate output across different downstream cost configurations, as in the risk neutral case, because the transfers it can extract from each downstream firm no longer vary with the cost realization of the competitor. Rather than include both firms in the market and pay each one a risk premium, the upstream firm simply chooses to avoid paying a risk premium to downstream firms.

At this point, one might ask why the upstream firm chooses to contract just one firm to produce output when there is another instrument available to decrease downstream firms' uncertainty: reducing the low cost output gaps Δ_i. This would lower the risk premium of downstream firms by reducing the difference between the possible profit realizations. In the limit as $\Delta_i \to 0$, the upstream firm would also pay no risk premia to downstream firms, just like with exclusive contracts. This is a poor alternative because when Δ_i is small, M does not take advantage of the greater productivity of low cost firms. It keeps both downstream firms in the market and minimizes risk premia, but sacrifices production efficiency. By contrast, when contracting with just one firm, it minimizes risk premia but also takes advantage of the productivity gains that arise from contracting the low cost firm to produce more than the high cost firm. There is also another possibility available to cope with M's problem, namely to provide a contract that insures downstream firms from risk. In section 5 we study these contracts.

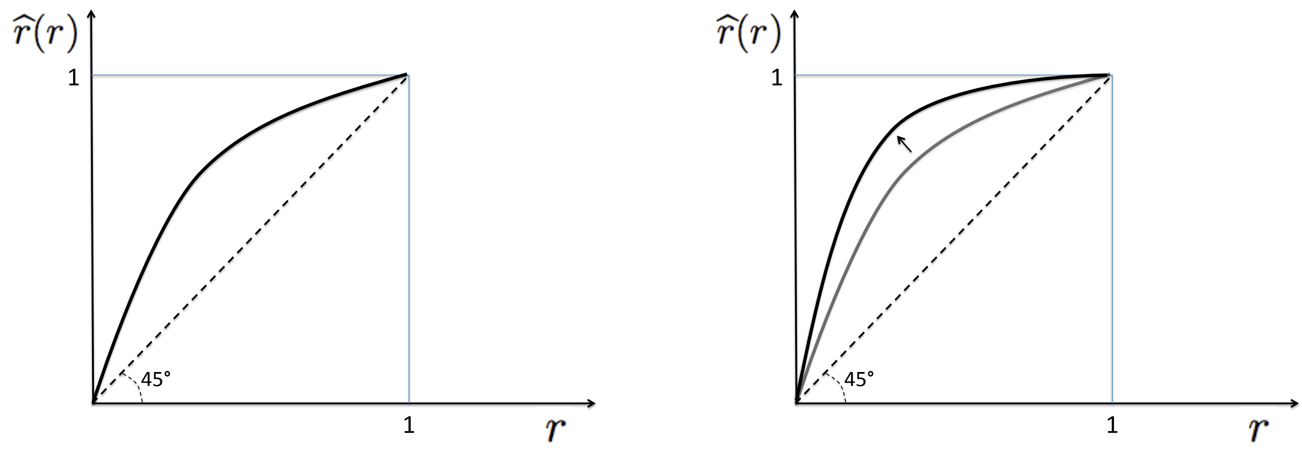
While the case of infinite risk aversion provides a clear illustration of the main idea of the paper, it is admittedly an extreme case. The next section explores the optimal contracts with intermediate risk aversion.

4 Intermediate Risk Preferences

A basic challenge for introducing intermediate levels of risk aversion in our model is that curvature in firms' utility of wealth functions would make characterizing optimal contracts much more difficult than in the baseline cases, in which firms' payoffs were linear in profits and we obtained clean results. Therefore we adopt an alternative approach taken from the well-established rank dependent utility literature that represents preferences over lotteries both in terms of the probability weights attached to wealth outcomes and the utility of wealth function.[11] When utility is linear in wealth, the probability weights alone embody an individual's attitudes towards risk. We use this framework to model downstream firms' utility functions, as suggested by Yaari (1987).[12] More specifically we assume that

$$U\left[L_i(\widehat{c}_i \mid c_i)\right] = \widehat{r}(r)\pi_i(\widehat{c}_i, 0, c_i) + \left[1 - \widehat{r}(r)\right]\pi_i(\widehat{c}_i, c, c_i) \tag{14}$$

where $\widehat{r}(r)$ is a *weighting function* defined on $r \in (0,1)$ with the following properties: (1) $\widehat{r}(r) > r$; (2) $\widehat{r}'(r) > 0$; (3) $\lim_{r\to 0}\widehat{r}(r) = 0$; and (4) $\lim_{r\to 1}\widehat{r}(r) = 1$.[13] Figure 1a plots an example of a weighting function.



(a) Weighting Function

(b) Increase in Risk Aversion

Figure 1: Examples of Weighting Functions

This formulation of risk aversion nests the two baseline cases. Risk neutrality is

[11] See Quiggin (1982) for a seminal reference.

[12] "In studying the behavior of firms, linearity in payments may in fact be an appealing feature. Under the dual theory, maximization of a linear function of profits can be entertained simultaneously with risk aversion. How often has the desire to retain profit maximization led to contrived arguments about firms' risk neutrality?" (Yaari (1987), page 96).

[13] The general formulation of rank-dependent utility is the following. Let n wealth outcomes be ordered so that $w_1 \geq w_2 \geq \ldots \geq w_n$ where p_i is the probability of w_i being realized. Then utility is

$$\sum_{i=1}^{n}\left[\alpha\left(\sum_{j=1}^{i} p_j\right) - \alpha\left(\sum_{j=1}^{i-1} p_j\right)\right] u(w_i)$$

where α is a weighting function such that $\alpha(0) = 0$ and $\alpha(1) = 1$. In our formulation $u(w_i) = w_i$, $w_1 = \pi_i(\widehat{c}_i, c, c_i)$, $w_2 = \pi_i(\widehat{c}_i, 0, c_i)$, and $p_1 = 1 - r$. So $\alpha(p_1) = \alpha(1 - r)$ describes preferences, which we have represented by the mirror function $\widehat{r}(r) = 1 - \alpha(1 - r)$.

obtained as $\widehat{r}(r)$ approaches r and infinite risk aversion is obtained as $\widehat{r}(r)$ approaches 1. Essentially the weighting function captures pessimism since the weight attached to the worse profit realization—the one corresponding to meeting a low cost competitor—is higher than the probability of meeting such a firm.[14] To see that this notion is closely related to risk aversion, one can rewrite $U[L_i(\widehat{c}_i \mid c_i)]$ in (14) as

$$r\pi_i(\widehat{c}_i, 0, c_i) + [1-r]\,\pi_i(\widehat{c}_i, c, c_i) - [\widehat{r}(r) - r]\,[\pi_i(\widehat{c}_i, c, c_i) - \pi_i(\widehat{c}_i, 0, c_i)]\,, \qquad (15)$$

or the expected value of the lottery minus $\widehat{r}(r) - r$ times the distance between the two wealth outcomes. This is similar to the standard approach of endowing firms with mean-variance preferences, but replaces variance with an alternative measure of outcome dispersion.[15] This alternative representation also makes clear that $\widehat{r}(r) - r$ measures risk aversion. Suppose some firm x has a weighting function $\widehat{r}_x(r)$ such that $\widehat{r}_x(r) - r = \chi_x(r)$ and some firm y has a weighting function $\widehat{r}_y(r)$ such that $\widehat{r}_y(r) - r = \chi_y(r) < \chi_x(r)\ \forall r \in (0,1)$. Then whenever firm x accepts an uncertain bet, firm y does too. Thus one can model an increase in risk aversion as an increase in $\widehat{r}(r)$ for all values of r, as shown in figure 1b. We now use this framework to study the implications of intermediate risk aversion for optimal contracts.

4.1 Optimal contracts with general demand

By now the arguments for obtaining the transfers of low and high cost firms should be familiar, so we directly state them.

Lemma 3 *The optimal contract under risk neutrality satisfies* $Q_i^*(0) \geq Q_i^*(c)$ *and*

$$T_i^*(0) = Q_i^*(0)\left\{\widehat{r}P\left[Q_i^*(0) + Q_j^*(0)\right] + (1-\widehat{r})P\left[Q_i^*(0) + Q_j^*(c)\right]\right\} - cQ_i^*(c) \qquad (16)$$

$$T_i^*(c) = Q_i^*(c)\left\{\widehat{r}P\left[Q_i^*(c) + Q_j^*(0)\right] + (1-\widehat{r})P\left[Q_i^*(c) + Q_j^*(c)\right] - c\right\}. \qquad (17)$$

In incentive compatible contracts, low cost firms must produce at least weakly more than high cost firms, meaning that profits upon meeting a low cost competitor are indeed lower than upon meeting a high cost competitor, as claimed above. For a fixed value of r, we are interested in the influence that the value of $\widehat{r}$ has on the optimal contracts, and in particular exclusion. As a starting point, we provide a useful expression for M's profits.

[14] Below we prove that $\pi_i(\widehat{c}_i, c, c_i) \geq \pi_i(\widehat{c}_i, 0, c_i)$ is indeed satisfied in the optimal contract.

[15] One might wonder why we do not simply combine CARA preferences with normally distributed payoffs to arrive at mean-variance preferences as is typical. In our model introducing normally distributed payoffs would be difficult even if we allowed for a continuum of cost types, as the noise in firms' payoffs emerges endogenously through the upstream firm's choice of contracts.

Lemma 4 *M's profits from contracts that satisfy the conditions in lemma 3 are given by*

$$r\widehat{r}(Q^H+\Delta)P(Q^H+\Delta)+(1-r)(1-\widehat{r})Q^H P(Q^H)-cQ^H+$$
$$r(1-\widehat{r})\left[(Q^H+\Delta_1)P(Q^H+\Delta_1)+(Q^H+\Delta_2)P(Q^H+\Delta_2)\right]+$$
$$(\widehat{r}-r)\left[Q_1(c)P(Q^H+\Delta_2)+Q_2(c)P(Q^H+\Delta_1)\right] \quad (18)$$

The terms in (18) in fact appeared in the two extreme cases of section 3. $(Q^H+\Delta)P(Q^H+\Delta)$ and $Q^H P(Q^H)$ correspond to aggregate revenue when both downstream firms are low cost and high cost, respectively, while cQ^H corresponds to aggregate production costs and information rents. None of these terms change with the distribution of aggregate output across downstream firms. The bracketed term in the second line of (18) corresponds to revenue with heterogeneity, which guided the optimal allocation between firms in the risk neutral case, while the bracketed term in the third line corresponds to worst case high cost revenue, which guided the optimal allocation in the infinite risk aversion case. Thus the optimal allocation in the intermediate case is influenced both by the desire to smooth aggregate output and to reduce the risk premium paid out to high cost firms. Moreover, the relative influence of each of these factors depends on downstream firms' risk aversion via the parameter $\widehat{r}$. When firms are more risk averse and $\widehat{r}$ increases, the weight attached to revenue with heterogeneity in profits decreases, while the weight attached to worst case high cost revenue increases. Indeed, as $\widehat{r}\to r$, revenue with heterogeneity alone determines profits and, as $\widehat{r}\to 1$, worst case high cost revenue alone does.

Another interesting message from the profit representation in lemma 4 is that the upstream firm appears to only care about the *high* cost firms' risk premia in deciding downstream allocations. One might then fairly wonder what role low cost firms' risk premia play in the model. The answer is that rather than show up as an independent term in overall profit, they appear implicitly by reducing the importance of revenue with heterogeneity. Recall from section 3.1 that the motivation for increasing Δ_1 when it was low relative to Δ_2 was that it allowed the upstream firm to extract additional revenue when firm 1 was low cost and firm 2 was high cost (albeit at the expense of reducing revenue in the opposite cost downstream cost configuration). When downstream firms are risk averse, the transfers that M collects from low cost firms are relatively insensitive to the profit difference across states of the world—in the case of infinite risk aversion they are in fact completely insensitive to the difference across states. So the risk aversion of low cost firms reduces the gains from shifting aggregate revenue across different states.

We now turn to discussing the optimal contract. Without loss of generality, let firm 1 be the firm for which $\Delta_1\geq\Delta_2$. The following gives the important properties of the optimal contract.

Proposition 3 *There exist values c ,r^*, and $\widehat{r}^*$ such that, whenever $c<c^*$ and $r<r^*$:*

1. $(Q_1^{H*}, \Delta_1^*) = (Q^{H*}, \Delta^*)$ *uniquely optimal whenever* $\widehat{r} > \widehat{r}^*$

2. $Q_1^{H*} = Q^{H*}$ *and* $\Delta_2^* < \frac{\Delta^*}{2}$ *is uniquely optimal whenever* $\widehat{r} \leq \widehat{r}^*$.

The conditions that r and c take low values are imposed in order to ensure that the upstream firm wishes to contract high cost firms to produce positive output, which will be the case when the probability of meeting high cost firms is high (so that r is low) and when high cost firms are relatively efficient. When no high cost output is contracted, the optimal contract splits Δ evenly across firms. This result splits the optimal contracts into two regions (within a restricted parameter space) that depend on downstream firms' risk aversion. When risk aversion crosses a threshold $rhat^*$, exclusive contracts are optimal, while when risk aversion is lower a type of "partial" exclusion emerges in which firm 1 produces all the high cost output as well as a higher proportion of low cost output.

To build intuition for this result, one can observe that $Q_1(c) = Q^H$ and $Q_2(c) = 0$ is always an optimal allocation since it minimizes worst case high cost revenue. Given this distribution of high cost output, the optimal value of Δ_2 for a fixed Δ is pinned down by the equation

$$r(1-\widehat{r})\left[MR(Q^H + \Delta - \Delta_2^*) - MR(Q^H + \Delta_2^*)\right] = (r - \widehat{r})Q^H P'(Q^H + \Delta_2^*). \qquad (19)$$

The choice of Δ_2^* must resolve a tradeoff for the upstream firm. First, increasing Δ_2 when $\Delta_1 > \Delta_2$ helps smooth aggregate revenue: the left hand side of (19) gives the increase in revenue from increasing Δ_2 in the cost configuration $(c_1, c_2) = (c, 0)$ net of the lost revenue in the cost configuration $(c_1, c_2) = (0, c)$. This is the marginal benefit of raising Δ_2. In the risk neutral case this was set equal to zero. With risk aversion there is another consideration though: raising Δ_2 increases the output of firm 1's low cost rival, decreasing firm 1's worst case revenue when high cost. This is the marginal cost, given by the right hand side of (19). This cost implies that Δ_2^* is strictly less than $\frac{\Delta}{2}$. The basic message is that the upstream firm allows some additional aggregate output volatility in order to reduce firm 1's risk premium. When risk aversion is high enough, the risk premium reduction effect dominates, and exclusive contracts are optimal.

It is also important to point out that even a little bit of risk aversion is sufficient to generate starkly asymmetric optimal contracts in which firm 1 is given a larger share of the market than firm 2 even though they are ex ante identical. We view this as a surprising result, as it shows that asymmetric contracts do not hinge on extreme preferences. More generally, it is the symmetric contract solution that depends on a knife-edge assumption on risk preferences (risk neutrality) rather than the exclusive contract.

4.2 Optimal contracts with linear demand

In order to form crisper predictions on when exclusion arises, we apply the expressions derived in the general demand case to examine the optimal contract in the case of linear demand $P(Q) = 1 - Q$. The following result provides a characterization.

Proposition 4 *With linear demand, whenever* $r < \min\left\{1 - c, \frac{2+\widehat{r}-2(1+\widehat{r})c}{3}\right\}$, $Q_1^{H*} = Q^{H*} > 0$ *and:*

1. $\Delta_2^* = 0$ *whenever* $\widehat{r} > \widehat{r}^* = \frac{r(1+c-r)}{1-c-r+2cr}$
2. $0 < \Delta_2^* < \frac{\Delta^*}{2}$ *whenever* $\widehat{r} \leq \widehat{r}^*$
3. $\frac{\Delta_2^*}{\Delta^*}$ *is declining in* $\widehat{r}$ *on* $\widehat{r} \in (r, \widehat{r}^*)$

The condition on r contained in the result is again made to ensure that positive high cost output is contracted. Conditional on Q^H being positive, the proposition shows that exclusive contracts are optimal *if and only if* firms' risk aversion surpasses a certain threshold, whereas proposition 4 showed the weaker result that exclusive contracts arise when firms are sufficiently risk averse. This new result gives the model some empirical content—at least in the linear demand case. It predicts that in markets in which firms are more risk averse (e.g. because they are smaller, have less access to capital markets, etc.) exclusion is more likely to be observed.

The solution to the linear case also displays another kind of monotonicity. We already know from the general demand case that firm 1 produces all of the high cost output, but this result shows that, in addition, the production share of the aggregate low cost output gap of firm 2 is declining when risk aversion increases. Hence a more refined empirical prediction of the model is that the difference in average output between two risk averse downstream firms is increasing in their risk aversion.

Proposition 4 also provides insights as to how exclusion depends on the heterogeneity of downstream firms as measured by c. One can easily show that $\frac{\partial \widehat{r}^*}{\partial c} > 0$. This means that when heterogeneity increases, the threshold for exclusion becomes higher.

Corollary 3 *In the linear demand case, the probability of exclusion is decreasing in the heterogeneity of downstream firms.*

In fact the model also delivers another surprising result when c takes on low values: $\lim_{c \to 0} \widehat{r}^* = r$; in words, when heterogeneity between downstream firms declines sufficiently far, almost *any* level of risk aversion is sufficient to guarantee the optimality of exclusive contracts. This further reinforces the point that the optimality of exclusive contracts arises in a potentially wide range of circumstances. At least a partial intuition for corollary 3 arises from the tradeoffs identified in the general demand case. When c is lower and downstream firms are more homogenous, the gains from contracting a larger Δ are reduced. At the same time, when Δ is lower, the gains from spreading it across firms is lower, so exclusion becomes more likely.

5 Insurance Contracts, and their Welfare Implications

Since the upstream firm is hurt by downstream risk aversion, one would expect it to seek out contractual arrangements that insured downstream firms against profit volatility. One way that M can implement an insurance contract is to offer each downstream firm a fixed wage in exchange for producing a certain output level, and in so doing turn itself into the residual claimant on the realized profit. This contractual arrangement shifts all the uncertainty in the market onto M, thus freeing it from having to pay out a risk premium. This can be interpreted as M vertically integrating with both downstream firms rather than allowing them to operate as independent entities, so we will use the terms "insurance contracts", "full integration", and "vertical integration" interchangeably in the rest of this section. We now extend the model from section 2 to include a period 0 in which M decides whether or not to vertically integrate (it makes take-it-or-leave-it fixed wage offers to the downstream firms) and analyze the welfare consequences of this decision.[16]

Since $U[w] = w$, the problem for M under vertical integration becomes

$$\max_{\{Q_i(\widehat{c}_i),W_i(\widehat{c}_i)\}_{i=1}^2} \sum_{i=1}^{2} \left(\begin{array}{c} rQ_i(0)\left\{rP\left[Q_i(0)+Q_j(0)\right]+(1-r)P\left[Q_i(0)+Q_j(c)\right]\right\}+ \\ (1-r)Q_i(c)\left\{rP\left[Q_i(c)+Q_j(0)\right]+(1-r)P\left[Q_i(c)+Q_j(c)\right]\right\}- \\ rW_i(0)-(1-r)W_i(c) \end{array} \right) \tag{20}$$

such that

$$W_i(0) \geq 0 \tag{PC_i^L}$$

$$W_i(c) - cQ_i(c) \geq 0 \tag{PC_i^H}$$

$$W_i(0) \geq W_i(c) \tag{IC_i^L}$$

$$W_i(c) - cQ_i(c) \geq W_i(0) - cQ_i(0). \tag{IC_i^H}$$

The value of integration for M is established in the next result.

Lemma 5 *The maximum profit level that M obtains under vertical integration is equal*

[16]Note that if firm M chooses not to integrate with the agents, the rest of the game follows as before. In particular, at the next stage of the game firm M will decide which contracts to post.

to the maximum profit level that it obtains from contracting with two independent risk-neutral firms.

It is important to emphasize that integration does not solve the information problem that M faces, as is sometimes argued in the theory of the firm literature. It continues to pay information rents to low cost firms just as when downstream firms operate independently. Instead, the value of integration lies in changing the distribution of risk between the downstream firms and upstream firm.

It should be clear that vertical integration makes M better off when downstream firms are risk averse, but a priori it is not clear whether downstream firms and consumers are better off. In order to address this issue, we will consider the case in which downstream firms are sufficiently risk averse for an exclusive contract to be optimal without vertical integration. The main result with general demand is the following.

Proposition 5 *The profits of the upstream firm are higher under vertical integration, while the joint profits of the downstream firms are lower.*

The result relies on how the total contracted output is distributed across the different cost types under integration. Moving from the optimal exclusive contract to the optimal contract under integration is equivalent to moving from the optimal contract with one risk neutral firm to the optimal contracts with two. When the upstream firm contracts with an additional firm, it contracts efficient producers (those with $c_i = 0$) to produce more on average and inefficient producers (those with $c_i = c$) to produce less on average: with two firms, meeting at least one low cost firm is more likely, so low cost firms can be contracted to produce more. While this makes the upstream firm better off, it makes the downstream firms worse off. Recall that the source of downstream firms' profits are their information rents, which they collect only if they are efficient. Furthermore the value of these rents is proportional to the amount of output that high cost firms are contracted to produce. Since integration reduces the output of high cost firms, there is a corresponding reduction in the value of the information rents.

There are two additional considerations one must make before deriving any welfare conclusions from proposition 5. First, it says nothing about how consumer surplus under vertical integration compares to surplus under exclusivity. Second, since downstream firms are risk averse, their ex ante expected utility of alternative contracts is in general not equal to their expected profits. In order to incorporate both issues, we turn to the linear demand case. Here the optimal exclusive contract is

$$Q^*_{L1} = \frac{1}{2} \tag{21}$$

$$Q^*_{H1} = \frac{1-c}{2} - \frac{r}{1-r}\frac{c}{2}. \tag{22}$$

where Q^*_{L1} is the amount contracted to the low cost type and Q^*_{H1} to the high cost. One can also easily show that the optimal symmetric contracts with two risk neutral firms[17] are given by[18]

$$Q^*_{L2} = \frac{1}{4} + \frac{c}{4} \tag{23}$$

$$Q^*_{H2} = \frac{1-c}{4} - \frac{r}{1-r}\frac{c}{2}. \tag{24}$$

As one can observe, $2Q^*_{L2} > Q^*_{L1}$, $2Q^*_{H2} < Q^*_{H1}$, and average *aggregate* output is the same in either case. So, when M integrates, low cost firms produce a higher proportion of total output, as highlighted in proposition 5. As average aggregate output stays constant, evaluating consumer welfare is straightforward. Since it is convex in total output (it is given by $\frac{1}{2}Q^2$), consumers are better off when the variance of total output increases, so integration improves their welfare.

In terms of the downstream firms, there are in fact two additional sources of uncertainty for downstream firms at $t = 0$ compared to the baseline model. First, they do not know whether their realized cost type will be high or low; second, under non-integration they do not know whether they will be excluded. We ignore this second consideration by assuming that firm 1 knows that under non-integration it will be offered an exclusive contract and that firm 2 knows it will not be contracted to produce any output by M. We are now ready to ask the following question: when does vertical integration increase welfare? To be precise, here we define welfare as the sum of (1) M's expected profit, (2) expected consumer surplus, and (3) downstream firms' expected utility. Rather than give a general answer, we will instead focus on two illustrative examples. The first is that

Example 1 *When $\widehat{r}$ is sufficiently large, welfare is always higher under integration.*

When downstream firms are sufficiently risk averse, vertical integration unambiguously improves welfare. Independently of M's decision to integrate or not, each downstream firm has an expected utility near zero, which corresponds to the worst-case payoff in both cases. As discussed above, both M and consumers are better off under integration, so welfare must improve. One implication of this example is that banning vertical integration can lower welfare, at least when levels of downstream risk aversion are high.

When $\widehat{r}$ is large, firm 1 places no value on the expected information rents it loses by being integrated with firm 2. When $\widehat{r}$ takes on intermediate values this is no longer the case, and integration has a social cost. We refer to producer surplus as the sum of M's expected profits and downstream firms' expected utility.

[17] Recall that one can focus on symmetric contracts in the risk neutral case without loss of generality.

[18] These are the individual firm cost-type production levels, not the aggregate cost-type production levels.

Example 2 *Let* $\hat{r}'(r) = \frac{r(1+c-r)}{1-c-r+2cr}$ *be downstream firms' weighting functions. There exists a* c^**,* r' *and* $r'' > r'$ *such that, for all* $c < c^*$*:*

1. *When* $r < r'$*, both welfare and producer surplus are higher under integration.*
2. *When* $r' < r < r''$*, welfare is higher under integration and producer surplus is higher under non-integration.*
3. *When* $r > r'$*, both welfare and producer surplus are higher under non-integration.*

This example is constructed in the following way. First we consider the least risk averse type that is offered an exclusive contract under non-integration, the expression for which comes from proposition 4. For low values of c this type approaches the risk neutral type, so computing downstream firms' expected utility is equivalent to computing their expected information rents. When r is high, the value of the lost expected information rents from integration is also high, since downstream firms earn this rent with a high probability.[19] A surprising result is that when r is high enough, downstream firms' lost information rents from integration are greater in value than the increase in M's profit and consumer surplus. This implies that M's integration decision in period 0 is not necessarily in line with welfare: when downstream firms are on average efficient (i.e. r is high), M will integrate to the detriment of society.

The example also has another interesting implication. Suppose we extend the model further to allow firm 1 to offer M a payment in period $t = -1$ in exchange for a commitment to not integrate and firm 2 to offer M a payment in the same period in exchange for a commitment to integrate. For intermediate values of r, firm 1 will be able to offer a payment larger than the combined sum of M's and 2's increased payoff under integration. Such a Coasean bargain would lead M to commit not to integrate in period 0, even though integration would *increase* welfare. In the situation with transfers, whenever vertical integration is observed, it is efficient, but inefficient exclusion can also arise.

The key message here is that a contracting device that on the surface appears to address the friction that leads to exclusion—i.e. vertical integration—in fact has subtle and ambiguous welfare consequences. This is essentially because the integration decision is taken by the principal, and it may harm the downstream agents, which may receive information rents, but not the risk premium, under integration.

[19] One concern is that when r is high, the firm does not what to contract any high cost output, in which case exclusive contracts are not optimal. The bound on r from proposition 4 for the upstream firm to contract high cost output is $r < \min\left\{1-c, \frac{2+\hat{r}-2(1+\hat{r})c}{3}\right\}$. As $c \to 0$, $1-c$ does not put an upper bound on r. At the same time, the second bound limits to $\frac{2+r}{3} > r$.

6 Summary and conclusions

This paper identifies a new rationale for using exclusivity provisions: when agents compete downstream, and do not observe each other's cost type, competition generates uncertainty, and leads risk-averse agents to require a rent (which rises with the degree of risk aversion) as a compensation for the risk of facing a more efficient rival. To save the payment of such rents, the principal may prefer to deal exclusively with one agent.

We also analyze fixed wage contracts (that we interpret as integration between the upstream firm and the downstream agents) and show they do not necessarily improve welfare.

Beyond the specific model used here, we believe that this mechanism offers a general reason why a principal may endogenously restrict the number of agents with whom it wants to deal. Whenever the payoff of one agent depends on the actions or the type of other agents, and there is imperfect information, if agents are risk averse the principal will be obliged to pay risk premia. To save on such rents, the principal may prefer to contract with a strict subset of the potential agents. We plan to show that the same mechanism still holds good in very different settings, such as for instance models of moral hazard where agents are paid according to relative performance schemes.

References

Asplund, M. (2002): "Risk-averse firms in oligopoly," *International Journal of Industrial Organization*, 20(7), 995–1012.

Hart, O., and J. Tirole (1990): "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity (Microeconomics)*, 1990, 205–286.

Laffont, J.-J., and D. Martimort (2002): *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press, Princeton, NJ.

Laffont, J.-J., and J. Tirole (1987): "Auctioning Incentive Contracts," *Journal of Political Economy*, 95(5), 921–37.

McAfee, R. P., and J. McMillan (1986): "Bidding for Contracts: A Principal-Agent Analysis," *RAND Journal of Economics*, 17(3), 326–338.

——— (1987): "Competition for Agency Contracts," *RAND Journal of Economics*, 18(2), 296–307.

Nocke, V., and J. Thanassoulis (2010): "Vertical Relations under Credit Constraints," Discussion Paper 7636, Centre for Economic Policy Research.

Quiggin, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior & Organization*, 3(4), 323–343.

Rey, P., and J. Stiglitz (1995): "The Role of Exclusive Territories in Producers' Competition," *RAND Journal of Economics*, 26(3), 431–451.

Rey, P., and J. Tirole (2007): "A Primer on Foreclosure," in *Handbook of Industrial Organization*, ed. by M. Armstrong, and R. Porter, pp. 2145–2220. North-Holland.

Riordan, M. H., and D. E. M. Sappington (1987): "Awarding Monopoly Franchises," *American Economic Review*, 77(3), 375–87.

Segal, I. (1999): "Contracting With Externalities," *The Quarterly Journal of Economics*, 114(2), 337–388.

Yaari, M. E. (1987): "The Dual Theory of Choice under Risk," *Econometrica*, 55(1), 95–115.

A Omitted Proofs

A.1 Proof of Lemma 1

Proof. With risk neutrality the maximization problem (3) becomes

$$\max_{\{Q_i(\widehat{c}_i),T_i(\widehat{c}_i)\}_{i=1}^2} \sum_{i=1}^{2} rT_i(0) + (1-r)T_i(c) \text{ such that}$$

$$Q_i(0)\left\{rP\left[Q_i(0)+Q_j(0)\right]+(1-r)P\left[Q_i(0)+Q_j(c)\right]\right\}-T_i(0)\geq 0 \qquad (PC_i^L)$$

$$Q_i(c)\left\{rP\left[Q_i(c)+Q_j(0)\right]+(1-r)P\left[Q_i(c)+Q_j(c)\right]-c\right\}-T_i(c)\geq 0 \qquad (PC_i^H)$$

$$\begin{aligned} &Q_i(0)\left\{rP\left[Q_i(0)+Q_j(0)\right]+(1-r)P\left[Q_i(0)+Q_j(c)\right]\right\}-T_i(0)\geq \\ &Q_i(c)\left\{rP\left[Q_i(c)+Q_j(0)\right]+(1-r)P\left[Q_i(c)+Q_j(c)\right]\right\}-T_i(c) \end{aligned} \qquad (IC_i^L)$$

$$\begin{aligned} &Q_i(c)\left\{rP\left[Q_i(c)+Q_j(0)\right]+(1-r)P\left[Q_i(c)+Q_j(c)\right]-c\right\}-T_i(c)\geq \\ &Q_i(0)\left\{rP\left[Q_i(0)+Q_j(0)\right]+(1-r)P\left[Q_i(0)+Q_j(c)\right]-c\right\}-T_i(0). \end{aligned} \qquad (IC_i^H)$$

One can ignore the PC_i^L constraints because they are implied by the IC_i^L and PC_i^H constraints. This is because

$$\begin{aligned} Q_i(0)\left\{rP\left[Q_i(0)+Q_j(0)\right]+(1-r)P\left[Q_i(0)+Q_j(c)\right]\right\}-T_i(0) &\geq \\ Q_i(c)\left\{rP\left[Q_i(c)+Q_j(0)\right]+(1-r)P\left[Q_i(c)+Q_j(c)\right]\right\}-T_i(c) &\geq \\ Q_i(c)\left\{rP\left[Q_i(c)+Q_j(0)\right]+(1-r)P\left[Q_i(c)+Q_j(c)\right]-c\right\}-T_i(c) &\geq 0. \end{aligned}$$

This in turn implies that the IC_i^L constraints must bind, since otherwise the upstream firm could increase $T_i(0)$ and strictly increase profit. Now IC_i^H can be rewritten as

$$\begin{aligned} &c\left[Q_i(0)-Q_i(c)\right] \geq \\ &T_i(c)-T_i(0) + \left(\begin{array}{c} Q_i(0)\left\{rP\left[Q_i(0)+Q_j(0)\right]+(1-r)P\left[Q_i(0)+Q_j(c)\right]\right\}- \\ Q_i(c)\left\{rP\left[Q_i(c)+Q_j(0)\right]+(1-r)P\left[Q_i(c)+Q_j(c)\right]\right\} \end{array}\right). \end{aligned} \qquad (A.1)$$

The RHS of (A.1) is greater than zero by IC_i^L. So we know that $Q_i(0) \geq Q_i(c)$. But then if IC_i^L is binding, IC_i^H must be satisfied. So we can drop IC_i^H from the problem as well. This leaves PC_i^H, which must be binding, since otherwise the upstream firm could increase $T_i(c)$ and strictly increase profit. ■

A.2 Proof of Proposition 1

Proof. All that remains to be shown is that the optimal value of Δ is positive. Suppose not, and let the optimal value of Q^H be $Q^{H\prime}$. The total profit of the upstream firm from this solution is $Q^{H\prime}\left[P\left(Q^{H\prime}\right)-c\right]$. Without loss of generality, this can be achieved through contracts in which $Q_1(c) = Q^{H\prime}$ and $\Delta_1 = \Delta_2 = 0$.

Now consider the problem in which the upstream maximizes profit within the restricted class of contracts $Q_1(c) = Q^H$ and $\Delta_1 = \Delta$. The problem becomes

$$\max_{Q^H \geq 0, \Delta \geq 0} r\left(Q^H + \Delta\right) P\left(Q^H + \Delta\right) + (1-r) Q^H P\left(Q^H\right) - cQ^H. \tag{A.2}$$

The optimal values (Q^{H*}, Δ^*) for this problem solve the first order conditions

$$\mathrm{MR}\left(Q^{H*} + \Delta^*\right) \leq 0 \tag{A.3}$$

$$r\mathrm{MR}\left(Q^{H*} + \Delta^*\right) + (1-r)\mathrm{MR}\left(Q^{H*}\right) - c \leq 0 \tag{A.4}$$

where (A.3) holds with equality if $\Delta^* > 0$ and (A.4) holds with equality if $Q^{H*} > 0$. These conditions together imply that $\Delta^* > 0$. Suppose not, and that $Q^{H*} = 0$. Then, from (A.3), it must be the case that MR < 0 which is ruled out by assumption. Suppose not, and that $Q^{H*} > 0$. Then (A.4) gives $\mathrm{MR}\left(Q^{H*}\right) = c > 0$ while (A.3) gives $\mathrm{MR}\left(Q^{H*}\right) < 0$, a contradiction.

Now since the contracts $Q^H = Q^{H\prime}$ and $\Delta_1 = \Delta = 0$ are within the set of feasible contracts for (A.2) and are not chosen, we have arrived at a contradiction of their optimality. ■

A.3 Proof of Lemma 2

Proof. For $j \neq i$ let $\max\{Q_j(0), Q_j(c)\} = \overline{Q}_j$. Under infinite risk aversion the constraints in problem (3) become

$$Q_i(0) P\left[Q_i(0) + \overline{Q}_j\right] - T_i(0) \geq 0 \tag{PC_i^L}$$

$$Q_i(c)\left\{P\left[Q_i(c) + \overline{Q}_j\right] - c\right\} - T_i(c) \geq 0 \tag{PC_i^H}$$

$$Q_i(0) P\left[Q_i(0) + \overline{Q}_j\right] - T_i(0) \geq Q_i(c) P\left[Q_i(c) + \overline{Q}_j\right] - T_i(c) \tag{IC_i^L}$$

$$Q_i(c)\left\{P\left[Q_i(c) + \overline{Q}_j\right] - c\right\} - T_i(c) \geq Q_i(0)\left\{P\left[Q_i(0) + \overline{Q}_j\right] - c\right\} - T_i(0). \tag{IC_i^H}$$

Because demand is strictly decreasing in price, the worst profit realization for firm i occurs when firm j produces a larger amount. Following the same steps as in the proof of lemma 1, one can show that a necessary condition for incentive compatibility is that $Q_i(0) \geq Q_i(c)$. Thus one can replace $\overline{Q}_j = Q_i(0)$ in the above expressions. Again following steps in lemma 1, one can show that in the optimal contract menus IC_i^L and PC_i^H bind, while IC_i^H and PC_i^L are slack. This yields that claimed representation of optimal transfers. ■

A.4 Proof of Proposition 2

Proof. In the text we showed that whenever it is optimal to contract any high cost output, it is optimal to contract with only one firm. To establish whether it is optimal to indeed contract one firm to produce a positive level of high cost output, observe again expressions (A.3) and (A.4). We have established that (A.3) always holds with equality. Under this condition (A.4) becomes $(1-r)\mathrm{MR}\left(Q^{H*}\right) - c \leq 0$. In order for the upstream firm to contract high cost output, it must be the case that $r < 1 - \frac{c}{\text{MR}(0)} = r^*$. $r^* > 0$ since by assumption $c < P(0) = \text{MR}(0)$. When $r > r^*$, $Q^{H*} = 0$ and M's problem becomes $\max_{\Delta \geq 0} \Delta P(\Delta)$. Since $\text{MR}(0) = P(0) > 0$ it is optimal to set a positive aggregate low cost output gap. ■

A.5 Proof of Lemma 3

Proof. Suppose that $Q_2(0) \geq Q_2(c)$. The firm 1's participation and incentive compatibility constraints become

$$Q_1(0)\left\{\widehat{r}P\left[Q_1(0) + Q_2(0)\right] + (1-\widehat{r})P\left[Q_1(0) + Q_2(c)\right]\right\} - T_1(0) \geq 0 \qquad (PC_1^L)$$

$$Q_1(c)\left\{\widehat{r}P\left[Q_1(c) + Q_2(0)\right] + (1-\widehat{r})P\left[Q_1(c) + Q_2(c)\right] - c\right\} - T_1(c) \geq 0 \qquad (PC_1^H)$$

$$\begin{aligned} &Q_1(0)\left\{\widehat{r}P\left[Q_1(0) + Q_2(0)\right] + (1-\widehat{r})P\left[Q_1(0) + Q_2(c)\right]\right\} - T_1(0) \geq \\ &Q_1(c)\left\{\widehat{r}P\left[Q_1(c) + Q_2(0)\right] + (1-\widehat{r})P\left[Q_1(c) + Q_2(c)\right]\right\} - T_1(c) \end{aligned} \qquad (IC_1^L)$$

$$\begin{aligned} &Q_1(c)\left\{\widehat{r}P\left[Q_1(c) + Q_2(0)\right] + (1-\widehat{r})P\left[Q_1(c) + Q_2(c)\right] - c\right\} - T_1(c) \geq \\ &Q_1(0)\left\{\widehat{r}P\left[Q_1(0) + Q_2(0)\right] + (1-\widehat{r})P\left[Q_1(0) + Q_2(c)\right] - c\right\} - T_1(0). \end{aligned} \qquad (IC_1^H)$$

Now, proceeding along the same steps as in lemma 1, IC_1^L and IC_1^H together imply that $Q_1(0) \geq Q_1(c)$. By a similar logic, the same would also be true if we had instead supposed that $Q_2(0) < Q_2(c)$ and exchanged $\widehat{r}$ for $1 - \widehat{r}$ in the above expressions and vice versa. But by a symmetric argument it must be the case that $Q_2(0) \geq Q_2(c)$ as well. So it is indeed the case that $\pi_i(\widehat{c}_i, c, c_i) \geq \pi_i(\widehat{c}_i, 0, c_i)$ as claimed in the text. From here one can proceed with the same logic as in lemma 1 to yield the stated representations of the transfers. ■

A.6 Proof of Lemma 4

Proof. Using the expressions from lemma 3, one can write

$$\begin{aligned} &r[T_1(c) + T_2(0)] + (1-r)[T_1(c) + T_2(c)] = \\ &r\left[\begin{array}{c} \widehat{r}[Q_1(c) + \Delta_1]P(Q^H + \Delta) + (1-\widehat{r})[Q_1(c) + \Delta_1]P(Q^H + \Delta_1) - cQ_1(c) + \\ \widehat{r}[Q_2(c) + \Delta_2]P(Q^H + \Delta) + (1-\widehat{r})[Q_2(c) + \Delta_2]P(Q^H + \Delta_2) - cQ_2(c) \end{array}\right] + \\ (1-r)&\left[\begin{array}{c} \widehat{r}Q_1(c)P(Q^H + \Delta_2) + (1-\widehat{r})Q_1(c)P(Q^H + \Delta_2) - cQ_1(c) + \\ \widehat{r}Q_2(c)P(Q^H + \Delta_1) + (1-\widehat{r})Q_2(c)P(Q^H + \Delta_1) - cQ_2(c). \end{array}\right] \end{aligned}$$

After removing the term

$$r\widehat{r}(Q^H + \Delta)P(Q^H + \Delta) + (1-r)(1-\widehat{r})Q^H P(Q^H) - cQ^H$$

from this expression one is left with

$$r(1-\widehat{r})\left[(Q_1(c)+\Delta_1)P(Q^H+\Delta_1)+(Q_2(c)+\Delta_2)P(Q^H+\Delta_2)\right]+$$
$$(1-r)\widehat{r}\left[Q_1(c)P(Q^H+\Delta_2)+Q_2(c)P(Q^H+\Delta_1)\right]$$

which equals

$$r(1-\widehat{r})\left[(Q_1(c)+\Delta_1)P(Q^H+\Delta_1)+(Q_2(c)+\Delta_2)P(Q^H+\Delta_2)\right]+$$
$$r(1-\widehat{r})\left[Q_1(c)P(Q^H+\Delta_2)+Q_2(c)P(Q^H+\Delta_1)\right]+$$
$$(1-r)\widehat{r}\left[Q_1(c)P(Q^H+\Delta_2)+Q_2(c)P(Q^H+\Delta_1)\right]-$$
$$r(1-\widehat{r})\left[Q_1(c)P(Q^H+\Delta_2)+Q_2(c)P(Q^H+\Delta_1)\right]$$

and finally

$$r(1-\widehat{r})\left[(Q^H+\Delta_1)P(Q^H+\Delta_1)+(Q^H+\Delta_2)P(Q^H+\Delta_2)\right]+$$
$$(\widehat{r}-r)\left[Q_1(c)P(Q^H+\Delta_2)+Q_2(c)P(Q^H+\Delta_1)\right].$$

■

A.7 Proof of Proposition 3

Proof. By inspection of equation (18) in the text, whenever $\Delta_1 \geq \Delta_2$, $Q_1^H = Q$ and $Q_2^H = 0$ is (weakly) optimal. Substituting these conditions in (18) along with the condition $\Delta_1 = \Delta - \Delta_2$ gives the expression

$$\begin{aligned}\pi(Q,\Delta,\Delta_2) = {} & r\widehat{r}R\left(Q+\Delta\right)+(1-r)(1-\widehat{r})R\left(Q\right)-cQ+ \\ & r(1-\widehat{r})\left[R\left(Q+\Delta-\Delta_2\right)+R\left(Q+\Delta_2\right)\right]+ \\ & (\widehat{r}-r)QP\left(Q+\Delta_2\right)\end{aligned} \tag{A.5}$$

We will solve the relaxed problem

$$\max_{Q\geq 0,\Delta\geq 0,\Delta\geq 0}\pi(Q,\Delta,\Delta_2) \tag{A.6}$$

and ignore the constraint $\Delta_2 \leq \frac{\Delta}{2}$, and show that all solution to the relaxed problem satisfy this ignored constraint.

In general (A.6) is not a concave problem and in general there can be multiple solutions to the Kuhn-Tucker first order conditions, some of which are not global maximum. We can rule out the case where $\Delta = 0$ since (A.5) then becomes $QP(Q) - cQ$, which, by the arguments from the proof of proposition 1, can be improved by some $\Delta_1 > 0$. One can also rule out solutions in which $Q = 0$ and $\Delta_2 = 0$. When $Q = 0$ (A.5) becomes

$$r\widehat{r}R(\Delta) + (1-r)(1-\widehat{r})R(0) + r(1-\widehat{r})\left[R(\Delta-\Delta_2) + R(\Delta_2)\right] \tag{A.7}$$

which is maximized at $\Delta_2 = \frac{\Delta}{2}$.

This solution (with $Q = 0$ and $\Delta_2 = \frac{\Delta}{2}$) satisfies the Kuhn-Tucker first order conditions if

$$\frac{\partial \pi(Q,\Delta,\Delta_2)}{\partial Q} = \begin{pmatrix} r\widehat{r}R'(Q+\Delta) + (1-r)(1-\widehat{r})R'(Q) + \\ r(1-\widehat{r})\left[R'(Q+\Delta-\Delta_2) + R'(Q+\Delta_2)\right] + \\ (\widehat{r}-r)\left[P(Q+\Delta_2) + QP'(Q+\Delta_2)\right] \end{pmatrix} - c < 0 \tag{A.8}$$

$$\frac{\partial \pi(Q,\Delta,\Delta_2)}{\partial \Delta} = r\widehat{r}R'(Q+\Delta) + r(1-\widehat{r})R'(Q+\Delta-\Delta_2) = 0 \tag{A.9}$$

$$\frac{\partial \pi(Q,\Delta,\Delta_2)}{\partial \Delta_2} = \begin{pmatrix} r(1-\widehat{r})\left[-R'(Q+\Delta-\Delta_2) + R'(Q+\Delta_2)\right] + \\ (\widehat{r}-r)QP'(Q+\Delta_2) \end{pmatrix} = 0 \tag{A.10}$$

are all satisfied at $Q = 0$ and $\Delta_2 = \frac{\Delta}{2} > 0$. (A.10) is clearly satisfied, while (A.8) and (A.9) rewrite as[20]

$$\widehat{r}R'(\Delta) + (1-\widehat{r})R'\left(\frac{\Delta}{2}\right) = 0 \tag{A.11}$$

$$(1-r)(1-\widehat{r})R'(0) + r(1-\widehat{r})R'\left(\frac{\Delta}{2}\right) + (\widehat{r}-r)P\left(\frac{\Delta}{2}\right) < c. \tag{A.12}$$

Call the solution to A.11 $\Delta^*(\widehat{r})$. Plugging into A.12 gives the condition on r

$$r > \frac{(1-\widehat{r})R'(0) + \widehat{r}P\left(\frac{\Delta^*(\widehat{r})}{2}\right) - c}{(1-\widehat{r})\left[R'(0) - R'\left(\frac{\Delta^*(\widehat{r})}{2}\right)\right] + P\left(\frac{\Delta^*(\widehat{r})}{2}\right)} = \widetilde{r}(\widehat{r},c). \tag{A.13}$$

$\widetilde{r}(\widehat{r},c) < 1$ since

$$(1-\widehat{r})R'\left(\frac{\Delta^*(\widehat{r})}{2}\right) + (\widehat{r}-1)P\left(\frac{\Delta^*(\widehat{r})}{2}\right) = (1-\widehat{r})\left(R'\left(\frac{\Delta^*(\widehat{r})}{2}\right) - P\left(\frac{\Delta^*(\widehat{r})}{2}\right)\right) < 0 < c. \tag{A.14}$$

$\widetilde{r}(\widehat{r},c) > 0$ if[21]

$$c < (1-\widehat{r})P(0) + \widehat{r}P\left(\frac{\Delta^*(\widehat{r})}{2}\right). \tag{A.15}$$

The right-hand side of this expression is positive since $P\left(\frac{\Delta^*(\widehat{r})}{2}\right) > R\left(\frac{\Delta^*(\widehat{r})}{2}\right) > 0$, where the last inequality follows from (A.11). Now define $c^* \min_{\widehat{r}}(1-\widehat{r})P(0) + \widehat{r}P\left(\frac{\Delta^*(\widehat{r})}{2}\right) \in (0, P'(0))$ and $r' = \min_{\widehat{r}} \widetilde{r}(\widehat{r},c^*)$. So one can conclude that the solution with $Q = 0$ and $\Delta_2 = \frac{\Delta}{2} > 0$ does not exist when $r < r'$ and $c < c^*$.

Another potential solution to (A.6) is an exclusive contract in which $\Delta_2 = 0$. By the above arguments an exclusive contract can only be optimal if $Q > 0$. Such an exclusive contract

[20] Here we have also plugged (A.8) into (A.9).

[21] Recall that $R'(0) = P(0)$.

satisfies the Kuhn-Tucker first order conditions if

$$\left[\frac{\partial \pi(Q,\Delta,\Delta_2)}{\partial Q}\right]_{\Delta_2=0} = 0 \tag{A.16}$$

$$\left[\frac{\partial \pi(Q,\Delta,\Delta_2)}{\partial \Delta}\right]_{\Delta_2=0} = 0 \tag{A.17}$$

$$\left[\frac{\partial \pi(Q,\Delta,\Delta_2)}{\partial \Delta_2}\right]_{\Delta_2=0} = 0. \tag{A.18}$$

which simplifies to

$$rR'(Q+\Delta) + (1-r)R'(Q) = c \tag{A.19}$$

$$rR'(Q+\Delta) = 0 \tag{A.20}$$

$$r(1-\widehat{r})\left[-R'(Q+\Delta) + R'(Q)\right] + (\widehat{r}-r)QP'(Q) < 0. \tag{A.21}$$

and further to

$$(1-r)R'(Q) = c \tag{A.22}$$

$$rR'(Q+\Delta) = 0 \tag{A.23}$$

$$r(1-\widehat{r})R'(Q) + (\widehat{r}-r)QP'(Q) < 0. \tag{A.24}$$

Let $Q^*(r)$ be the solution to (A.22). Since $R'(Q^*) = P(Q^*) + Q^*P'(Q^*)$ (A.24) is satisfied whenever

$$\widehat{r} > \frac{rP(Q^*(r))}{rR'(Q^*(r)) - Q^*(r)P'(Q^*(r))} = f(r) \in (r,1). \tag{A.25}$$

Let $\widehat{r}' = \max_r f(r)$. Now, because marginal revenue is decreasing, $Q^*(r)$ is decreasing in r. $Q^*(r) > 0$ for r near 0 since by assumption $R'(0) = P'(0) > c$. But, assuming that $R'(0) < \infty$, there will exist some point $r'' > 0$ at which $Q^*(r'') = 0$. Define $r^* = \min\{r', r''\}$.

The final solution to consider is one in which no boundary solutions to (A.6) exist. The resulting system of equations simplifies to

$$\begin{pmatrix} (1-r)(1-\widehat{r})R'(Q) + r(1-\widehat{r})R'(Q+\Delta_2) + \\ (\widehat{r}-r)\left[P(Q+\Delta_2) + QP'(Q+\Delta_2)\right] \end{pmatrix} = c \tag{A.26}$$

$$\widehat{r}R'(Q+\Delta) + (1-\widehat{r})R'(Q+\Delta-\Delta_2) = 0 \tag{A.27}$$

$$r(1-\widehat{r})\left[-R'(Q+\Delta-\Delta_2) + R'(Q+\Delta_2)\right] + (\widehat{r}-r)QP'(Q+\Delta_2) = 0. \tag{A.28}$$

Since $P' < 0$, (A.28) implies that $R'(Q+\Delta_2) > R'(Q+\Delta-\Delta_2)$ which in turn implies $\Delta_2 < \Delta - \Delta_2$ and $\Delta_2 < \frac{\Delta}{2}$. So the original claim that the solution to the relaxed problem in (A.6) is also the solution to the problem with the constraint $\Delta_1 \geq \Delta_2$ is validated. Also notice that as $\widehat{r}$ approaches 1, the left hand side of (A.28) must be strictly negative. So there exists some value $\widehat{r}''$ such that this solution does not exist for $\widehat{r} > \widehat{r}''$. Let $\widehat{r}^* = \max\{\widehat{r}', \widehat{r}''\}$.

We have show that for $r < r^*$ and $c < c^*$ only the exclusive contract and above solution exist. Moreover, within this parameter space, when $\widehat{r} > \widehat{r}^*$, only the exclusive contract solution exists. When $\widehat{r} \leq \widehat{r}^*$, either the exclusive contract or above solution exist. For both solutions we find that $\Delta_2^* < \frac{\Delta^*}{2}$. ■

A.8 Proof of Proposition 4

Proof. The strategy for the first part of the proof is to utilize the expressions for the existence of the three solutions derived in the proof of proposition 3. First consider the solution in which $Q = 0$ and $\Delta_2 = \frac{\Delta}{2} > 0$. Equation (A.11) becomes

$$\widehat{r}(1 - 2\Delta) + (1 - \widehat{r})(1 - \Delta) = 0 \tag{A.29}$$

or $\Delta = \frac{1}{1+\widehat{r}}$. Plugging this expression in (A.13) gives

$$\begin{aligned}\widetilde{r}(\widehat{r}, c) = \frac{(1 - \widehat{r}) + \widehat{r}\left(1 - \frac{1}{2(1+\widehat{r})}\right) - c}{(1 - \widehat{r})\left[1 - \left(1 - \frac{1}{1+\widehat{r}}\right)\right] + \left(1 - \frac{1}{2(1+\widehat{r})}\right)} = \frac{1 - \frac{\widehat{r}}{2(1+\widehat{r})} - c}{(1 - \widehat{r})\left(\frac{1}{1+\widehat{r}}\right) + 1 - \frac{1}{2(1+\widehat{r})}} = \\ \frac{2(1 + \widehat{r}) - \widehat{r} - 2(1 + \widehat{r})c}{1 + 2(1 + \widehat{r}) - 2\widehat{r}} = \frac{2 + \widehat{r} - 2(1 + \widehat{r})c}{3}.\end{aligned} \tag{A.30}$$

Next consider the exclusive contract solution. (A.22) gives $Q^*(r) = \frac{1-r-c}{2(1-r)}$ which is positive as long as $r < 1 - c$. So whenever $r < \min\left\{1 - c, \frac{2+\widehat{r}-2(1+\widehat{r})c}{3}\right\}$ the exclusive contract solution exists and the solution above does not. Plugging $Q^*(r)$ into (A.25) gives

$$f(r) = \frac{r\left[\frac{2-2r-(1-r-c)}{2(1-r)}\right]}{r\left[\frac{1-r-(1-r-c)}{(1-r)}\right] + \frac{1-r-c}{2(1-r)}} = \frac{r(1 - r + c)}{1 - r - c + 2cr}. \tag{A.31}$$

Now finally consider the solution with partial exclusion. Expressions (A.26)-(A.28) solve as

$$Q^* = \frac{r(1 - \widehat{r})(2 - 3r + \widehat{r} - 2c(1 + \widehat{r}))}{4r - 2r\widehat{r} - \widehat{r}^2 + r^2(-5 + 4\widehat{r})} > 0 \tag{A.32}$$

$$\Delta^* = \frac{-r^2 + r(2 + 6c - 6c\widehat{r}) + \widehat{r}(-2 + 2c(1 - \widehat{r}) + \widehat{r})}{8r - 4r\widehat{r} - 2\widehat{r}^2 + 2r^2(-5 + 4\widehat{r})} > 0 \tag{A.33}$$

$$\Delta_2^* = \frac{-r^2 - (1 - c)\widehat{r} + r(1 + c + \widehat{r} - 2c\widehat{r})}{4r - 2r\widehat{r} - \widehat{r}^2 + r^2(-5 + 4\widehat{r})} > 0. \tag{A.34}$$

Whenever $r < \frac{2+\widehat{r}-2(1+\widehat{r})c}{3}$ the numerator of (A.32) is positive. The condition for the denominator to be positive is that

$$r \leq \widehat{r} \leq -r(1 - 2r) + 2(1 - r)\sqrt{r(1 + r)} < 1. \tag{A.35}$$

(A.33) is positive when

$$r \leq \widehat{r} \leq \frac{1 - c + 3cr - \sqrt{(1 - r)^2(1 - 2c) + c^2(1 + 3r)^2}}{1 - 2c}, \tag{A.36}$$

which is more stringent than the condition in (A.35). Finally, the condition for (A.34) positive is

$$r \leq \widehat{r} \leq \frac{r(1+c-r)}{1-c-r+2cr} \equiv \widehat{r}^*. \tag{A.37}$$

Since $\widehat{r}^* < \frac{1-c+3cr-\sqrt{(1-r)^2(1-2c)+c^2(1+3r)^2}}{1-2c}$, $r \leq \widehat{r} \leq \frac{r(1+c-r)}{1-c-r+2cr} \equiv \widehat{r}^*$ is the condition for the partial exclusion solution to hold.

To prove the final statement, let

$$T = \frac{\Delta_2^*}{\Delta^*} = \frac{2\left[-r^2-(1-c)\widehat{r}+r(1+c+\widehat{r}-2c\widehat{r})\right]}{-r^2+r(2+6c-6c\widehat{r})+\widehat{r}(-2+2c(1-\widehat{r})+\widehat{r})}.$$

Differentiating this expression yields

$$\frac{\partial T}{\partial \widehat{r}} = \frac{-2\left[A(c,r)+B(c,r)\widehat{r}-C(c,r)\widehat{r}^2\right]}{\left[-r^2+r(2+6c-6c\widehat{r})+\widehat{r}(-2+2c(1-\widehat{r})+\widehat{r})\right]^2},$$

where

$$A(c,r) = 4cr - 4c^2r - r^2 - 9cr + 6c^2r^2 + r^3 + 4cr^3,$$

$$B(c,r) = 2r\left(1-c-2c^2-r+2cr\right) > 0,$$

and

$$C(c,r) = (1-2c)\left[(1-c)-(1-2c)r\right] > 0.$$

To show that $\frac{\partial T}{\partial \widehat{r}} \leq 0$ we have to show that $C(c,r)\widehat{r}^2 - B(c,r)\widehat{r} - A(c,r) \leq 0$. This inequality is solved by: $\widehat{r}_1(c,r) \leq \widehat{r} \leq \widehat{r}_2(c,r)$, where: $\widehat{r}_1(c,r) = \frac{B-\sqrt{B^2-4AC}}{2C}$ and $\widehat{r}_2(c,r) = \frac{B+\sqrt{B^2-4AC}}{2C}$. Now

$$\left[\frac{\partial T}{\partial \widehat{r}}\right]_{\widehat{r}=r} = -\frac{1-c-r(1+c)}{8cr(1-r)} < 0.$$

We also know that $T(\widehat{r}^*) = 0$ and that, for $\widehat{r} > \widehat{r}^*$, $T(\widehat{r}) \leq 0$. If $\widehat{r}_2(c,r)$ were lower than $\widehat{r}^*$, then it would also be true that $\left[\frac{\partial T}{\partial \widehat{r}}\right]_{\widehat{r}=\widehat{r}^*} > 0$, which is a contradiction since it would imply that there exists an ε such that $T(\widehat{r}) > 0$ for $\widehat{r} = \widehat{r}^* + \varepsilon$. Hence, it must be that $\left[\frac{\partial T}{\partial \widehat{r}}\right]_{\widehat{r}=\widehat{r}^*} \leq 0$, and that $T(\widehat{r})$ is decreasing on $\widehat{r} \in (r, \widehat{r}^*)$. ■

A.9 Proof of Lemma 5

Proof. Using the same arguments as in lemma 1 one can show that IC_i^L is binding, which gives $W_i(0) = W_i(c)$, and that PC_i^H is binding, which gives $W_i(0) = W_i(c) = cQ_i(c)$. Plugging these values back into M's objective function yields precisely the expression in (7). ■

A.10 Proof of Proposition 5

Proof. That vertical integration increases the profit of the upstream firm is trivial: the exclusive contract is within the set of feasible contracts under integration, yet is not in the set of optimal contracts.

Let Q_E^H and Q_E^L be the total amount of output contracted to the high and low cost types in the optimal exclusive contract, and let Q_I^H and Q_I^L be the total amount of output of output contracted to high and low cost types in the optimal contract under integration. Without loss of generality, Q_I^H and Q_I^L can be taken as twice what each individual cost type produces since symmetric contracts are optimal with risk neutrality. Total expected downstream profits under exclusion are rcQ_E^H and under integration are rcQ_I^H.

The profits of the upstream firm in the exclusive case are

$$\pi_E = rQ_E^L P\left(Q_E^L\right) + (1-r)Q_E^H P\left(Q_E^H\right) - cQ_E^H$$

while profits under integration are

$$\pi_I = r^2 Q_I^L P\left(Q_I^L\right) + (1-r)^2 Q_I^H P\left(Q_I^H\right) + 2r(1-r)\left(\frac{Q_I^H + Q_I^L}{2}\right) P\left(\frac{Q_I^H + Q_I^L}{2}\right) - cQ_I^H$$

which simplifies to

$$rQ_I^L\left[rP\left(Q_I^L\right) + (1-r)P\left(\frac{Q_I^H + Q_I^L}{2}\right)\right] + (1-r)Q_I^H\left[rP\left(\frac{Q_I^H + Q_I^L}{2}\right) + (1-r)P\left(Q_I^H\right)\right] - cQ_I^H.$$

The optimal Q_E^H—Q_E^{H*}—satisfies the first order condition

$$\left[\frac{\partial \pi_E}{\partial Q_E^H}\right]_{Q_E^H = Q_E^{H*}} = (1-r)\text{MR}\left(Q_E^{H*}\right) - c = 0.$$

while the optimal Q_I^H—Q_I^{H*}—satisfies the first order condition

$$\left[\frac{\partial \pi_I}{\partial Q_I^H}\right]_{Q_I^H = Q_I^{H*}} = (1-r)\left[(1-r)\text{MR}\left(Q_I^{H*}\right) + r\text{MR}\left(\frac{Q_I^{L*} + Q_I^{H*}}{2}\right)\right] - c = 0.$$

so that

$$(1-r)\text{MR}\left(Q_I^{H*}\right) + r\text{MR}\left(\frac{Q_I^{L*} + Q_I^{H*}}{2}\right) = \text{MR}\left(Q_E^{H*}\right).$$

Since, by proposition 1, $Q_I^{L*} > Q_I^{H*}$, $\text{MR}\left(\frac{Q_I^{L*}+Q_I^{H*}}{2}\right) < \text{MR}\left(Q_I^{H*}\right)$, so that the above expression can only hold with equality if $Q_E^{H*} > Q_I^{H*}$. But this implies that joint downstream profits are lower under integration than exclusion. ■

A.11 Computations for Examples in Section 5

The expressions for the optimal exclusive contract arise from a straightforward application of the proof of proposition 4. When contracting with two risk neutral firms, M's problem can be written as

$$\max_{Q_{L2}, Q_{H2}} 2\left[rQ_{L2}\left(1 - Q_{L2}\right) + (1-r)Q_{H2}\left(1 - Q_{H2}\right) - cQ_{H2} - \left(rQ_{L2} + (1-r)Q_{H2}\right)^2\right] \quad \text{(A.38)}$$

which yields the first order conditions

$$(1+r)Q^*_{L2} + (1-r)Q^*_{H2} = \frac{1}{2} \tag{A.39}$$

$$rQ^*_{L2} + (2-r)Q^*_{H2} = \frac{1}{2} - \frac{r}{1-r}\frac{c}{2} \tag{A.40}$$

with solutions

$$Q^*_{L2} = \frac{1}{4} + \frac{c}{4} \tag{A.41}$$

$$Q^*_{H2} = \frac{1-c}{4} - \frac{r}{1-r}\frac{c}{2}. \tag{A.42}$$

A more general expression for M's profit in the linear demand case is, for $k = 1, 2$,[22]

$$\mathbb{E}[\,Q\,] - \mathbb{E}\left[\,Q^2\,\right] - kcQ^*_{Hk} = \mathbb{E}[\,Q\,] - \mathbb{E}[\,Q\,]^2 - V[\,Q\,] - kcQ^*_{Hk}$$
$$= \mathbb{E}[\,Q\,] - \mathbb{E}[\,Q\,]^2 - kr(1-r)\,(Q^*_{Lk} - Q^*_{hk}) - kcQ^*_{Hk} \tag{A.43}$$

where Q is aggregate output and the expression for variance in the last equality arises from arguments in section 3.1. Basic calculations reveal that expected output is the same in both cases. The difference in variance between the integrated and independent cases is

$$r(1-r)\left[\frac{1}{2} - \frac{1-c}{2} + \frac{r}{1-r}\frac{c}{2}\right]^2 - 2r(1-r)\left[\frac{1}{4} + \frac{c}{4} - \frac{1-c}{4} + \frac{r}{1-r}\frac{c}{2}\right]^2 =$$
$$-r(1-r)\left[\left(\frac{r}{1-r} + 1\right)\frac{c}{2}\right]^2 = -\frac{r}{1-r}\frac{c^2}{4} \tag{A.44}$$

while

$$2cQ^*_{H2} - cQ^*_{H1} = c\left(\frac{1-c}{2} - \frac{r}{1-r}\frac{c}{2}\right) - 2c\left(\frac{1-c}{4} - \frac{r}{1-r}\frac{c}{2}\right) =$$
$$c\left(\frac{1-c}{2} - \frac{r}{1-r}\frac{c}{2}\right) - c\left(\frac{1-c}{2} - \frac{r}{1-r}c\right) = \frac{0.5rc^2}{1-r}. \tag{A.45}$$

So integration increases M's profit by $\frac{r}{1-r}\frac{c^2}{4}$. A general expression for consumer surplus is

$$\mathbb{E}\left[\frac{1}{2}Q^2\right] = \frac{1}{2}\left(\mathbb{E}[\,Q\,]^2 + V[\,Q\,]\right). \tag{A.46}$$

Using the same steps as above, one can show that consumer surplus increases by $\frac{r}{1-r}\frac{c^2}{8}$ under integration. Information rents for firm 1 with the exclusive contract generate expected utility $[1-\widehat{r}(1-r)]\,c\left(\frac{1-c}{2} - \frac{r}{1-r}\frac{c}{2}\right)$, while firm 2 has an expected utility of 0. Under vertical integration both have expected utility $[1-\widehat{r}(1-r)]\,c\left(\frac{1-c}{4} - \frac{r}{1-r}\frac{c}{2}\right)$. So, total downstream expected

[22] The expressions for Q^*_{H1} and Q^*_{L1} are in the text.

utility under integration decreases by

$$
\begin{aligned}
&[1-\widehat{r}(1-r)]\, c\left(\frac{1-c}{2}-\frac{r}{1-r}\frac{c}{2}\right)-[1-\widehat{r}(1-r)]\, 2c\left(\frac{1-c}{4}-\frac{r}{1-r}\frac{c}{2}\right)= \\
&[1-\widehat{r}(1-r)]\, c\left(\frac{1-c}{2}-\frac{r}{1-r}\frac{c}{2}\right)-[1-\widehat{r}(1-r)]\, c\left(\frac{1-c}{2}-\frac{r}{1-r}c\right)= \\
&\frac{[1-\widehat{r}(1-r)]\, rc^2}{2(1-r)}
\end{aligned} \tag{A.47}
$$

Total welfare is higher with non-integration whenever

$$
\frac{[1-\widehat{r}(1-r)]\, rc^2}{2(1-r)}-\frac{r}{1-r}\frac{c^2}{8}-\frac{r}{1-r}\frac{c^2}{4}>0. \tag{A.48}
$$

When $\widehat{r}\to 1$ this never holds. Now suppose $\widehat{r}(1-r)=\widehat{r}'(1-r)$.

$$
\frac{[1-\widehat{r}'(1-r)]\, rc^2}{2(1-r)}-\frac{r}{1-r}\frac{c^2}{8}-\frac{r}{1-r}\frac{c^2}{4}>0 \tag{A.49}
$$

whenever

$$
\frac{[1-\widehat{r}'(1-r)]}{2}-\frac{3}{8}>0. \tag{A.50}
$$

Using the fact that $1-\widehat{r}'(1-r)=\frac{r(r-c)}{c+r-2cr}$, this condition rewrites as $r(4r-3)-c(3-2r)>0$. For c sufficiently low, this condition will hold with equality for some $r''\in(0.75,1)$. Moreover the left hand side is increasing whenever $r>\frac{3}{8}$, so the condition holds if and only if $r>r''$. Producer surplus is higher under non-integration whenever

$$
\frac{[1-\widehat{r}'(1-r)]}{2}-\frac{1}{4}>0. \tag{A.51}
$$

which rewrites as $r(2r-1)-c>0$, which holds with equality for some $r'\in(0.5,1)$ when c is sufficiently low. Moreover, since the left hand side is increasing whenever $r>\frac{1}{4}$, the condition holds if and only if $r>r'$. Finally, note that $r(2r''-1)-c$ must be bigger than zero since $2r-1>4r-3$ and $1<3-2r$. So we can conclude that $r'<r''$.