The optimality of ad valorem contracts

Andrei Hagiu* and Julian Wright†

July, 2018

Abstract

We provide a new theory of ad valorem contracts (i.e., contracts that vary with the value of the transaction) which can explain why such contracts are widely used between vertically related parties (e.g., in franchising and licensing). Ad valorem contracts allow upstream firms (principals) to preserve their own incentives to make ongoing investments in the channel, deal with pricing distortions caused by channel coordination problems, while at the same time adjusting their investment based on demand shocks that are only observed by the downstream firms (agents). We show that the optimal ad valorem contract allows the principal to achieve the same profits as if it could observe the demand shocks and control price. This optimal contract makes use of revenue sharing (to balance investment incentives and make the principal's investment responsive to demand through price), upfront fixed fees (to extract the agents' expected profit) and an additional term that depends non-linearly on either price or demand (to correct for remaining pricing distortions). Our results are robust to the introduction of competition between agents, production costs, and imperfect monitoring of the agents' prices.

Keywords: revenue sharing, channel coordination, moral hazard, private information

1 Introduction

The use of price-dependent contracts by vertically related parties is widespread. For example, revenue sharing (or royalty payments) in which the fee paid per unit is a constant fraction of the price is common in franchising contracts, in licensing contracts, in shopping mall rental contracts, in the contracts online platforms offer sellers, and sometimes in the wholesale prices charged by manufacturers or suppliers to retailers (e.g., retailers sometimes require markdowns on the wholesale price when they put items on sale, or obtain trade promotions or buybacks that are equivalent to revenue sharing).

*Boston University Questrom School of Business, Boston, MA 02215, E-mail: ahagiu@bu.edu

†Department of Economics, National University of Singapore, Singapore 117570, E-mail: jwright@nus.edu.sg

In this paper we provide a new theory explaining why such ad valorem[1]) contracts (i.e., contracts that vary with the value of the transaction can be optimal. In particular, our theory can explain why the fees charged by a franchisor, a licensor, a shopping mall, an online platform, or a manufacturer may be conditioned on the price set by the corresponding franchisee, licensee, retail shop, online supplier, or distributor. To do so, we use a principal-agent model, in which demand is determined by (i) price, (ii) two ongoing costly investments (or effort) decisions, one made by the principal and one made by the agent, and (iii) a random demand shock privately observed by the agent. The model captures that double-sided moral hazard is common in franchising (ongoing brand advertising by the franchisor, sales effort by the franchisee), licensing (ongoing investments in technological improvements by the licensor, sales effort by the licensees), shopping malls (ongoing investments in facilities and promotional activities by malls, sales effort by retail shops), online platforms (ongoing investments in design and online advertising by the platform, investment in quality and service by online suppliers), and manufacturer-retailer contracts (ongoing brand advertising by the manufacturers, sales effort by the retailers). Furthermore, because the respective agents are typically closer to customers, they usually have private information about demand.

To understand why ad valorem contracts are optimal in this setting, note that in order for the principal to have an incentive to continue investing in activities that increase demand, a positive (above cost) wholesale price is required. However, a positive wholesale price gives rise to a pricing distortion because it implies that the agent does not capture the full benefits associated with expanding demand. An ad valorem contract (with wholesale price proportional to price) can be used to address this problem. At the same time, by tying the per-unit wholesale price to the retail price, the ad valorem contract is able to transmit the agent's private information on demand shocks to the principal, so that the principal adjusts its costly investment accordingly.

In our model, a relatively simple ad valorem contract enables the principal to achieve the same expected profits as if it observes the demand shock and can perfectly control the price set by the agent. We call this the second-best outcome, reflecting that the only problem left unsolved is double-sided moral hazard. The ad valorem contract achieving the second-best outcome consists of a per-unit wholesale price that is proportional to the agent's price (i.e., constant revenue sharing), an additional payment made by the agent to the principal that is independent of demand and depends non-linearly on the price set by the agent, and an upfront fixed fee. Revenue sharing splits the total revenue generated between the principal and the agent so as to optimally balance their respective investment incentives. Revenue sharing also mitigates the pricing distortion that arises from a positive wholesale price. However, whenever the principal's moral hazard problem differs from that of the agent's, a constant revenue share alone is not enough to eliminate the pricing distortion. This is why the additional payment dependent on price is needed.

The principal can achieve the same outcome by replacing this additional payment that depends on price with a non-linear quantity term. Specifically, if the per-unit wholesale price depends linearly

[1]In taxation, ad valorem taxes refer to sales taxes levied on the value of sales (i.e., the price paid) as opposed to specific (or unit) taxes which are sales taxes levied on the number of units sold regardless of the price. See, for example, Suits and Musgrave (1953).

on the price and also on the level of demand, then the principal can also implement the second-best outcome. This captures the case in which the wholesale price paid by the retailer depends on the amount it orders from the manufacturer, as is the case with incremental quantity discounting. The key point in either implementation is that whenever the principal's moral hazard problem differs from that of the agent's, the principal needs an additional contractual instrument to correct for the distortion in prices beyond pure revenue sharing. To make this clear, we characterize how much profit the principal loses by being restricted to a simple revenue sharing contract with an upfront fixed fee. The resulting loss is relatively small provided the effect of the agent's and the principal's investments are not too different and/or provided the principal's investment is not very important.

Perhaps remarkably, the implementation of the principal's second-best outcome through the simple ad valorem contract described above is robust to the introduction of competing agents, as well as to the introduction of marginal costs of production for both the principal and the agent. With competing agents, the only significant change is that there is less upward price distortion, which implies the payment from agents to the principal that only depends on price must be adjusted downwards to reduce the penalty for agents setting higher prices. With marginal production costs, the principal's optimal ad valorem contract has two additional components: (i) a constant per-unit wholesale price which adjusts for the difference in marginal costs weighted by the importance of investments between the principal and the agent, and ii) an additional payment from the agent to the principal that depends linearly on price and compensates for the new distortion in price introduced by the constant per-unit wholesale price.

An important element that makes the implementation of our ad valorem contracts work is the principal's ability to observe the price chosen by the agent. Nevertheless, we are able to show that the ad valorem contract still works even if the principal only observes the agent's price with some probability, provided the principal's contract also includes a penalty which the agent would pay when the principal observes a price that differs from the one reported. Going forward, due to better monitoring technologies, manufacturers, franchisors and licensors will be better able to monitor the prices set by their agents, so we expect the use of such ad valorem contracting to become more widespread.

2 Related literature

We directly build upon a substantial literature exploring how more sophisticated contracts than linear wholesale pricing can help address channel coordination problems in principal-agent settings, such as in the context of franchising. Much of the literature focuses on the role of revenue-sharing contracts together with upfront fixed fees (and sometimes with other more sophisticated instruments as well, like price-dependent profit-sharing) in mitigating concerns arising from double marginalization.[2] Relative to this literature, we establish a new channel by which revenue-sharing contracts improve outcomes, one which arises when the principal's moral hazard problem co-exists with the agent's private information.

[2]Some of the related literature assumes away the use of fixed fees (e.g., Cachon and Lariviere, 2005). From a double marginalization perspective, the absence of fixed fees provides an even stronger rationale for using revenue sharing (e.g., it would exist even without any principal moral hazard or private information on the part of agents).

Within this literature, our model setting is closest to that in the seminal work of Lal (1990), who provides a principal-agent model of franchising. In Section 4 of his paper, Lal considers a setting with double-sided moral hazard, and shows that the principal benefits from using revenue sharing in this context. One difference is that Lal models the principal's and agent's investment/effort decisions as binary (either high or low), while we allow for continuous investment levels and associated convex costs, which are more realistic. A more substantive difference is that while Lal allows for a random demand shock, this is observed neither by the principal nor by the agent, so there is no private information in his setting, which makes the channel coordination problem easier to solve. Thus, in contrast to Lal and other related papers, we uncover a new channel through which revenue sharing contracts achieve optimality, which is that revenue sharing facilitates the transmission of the agent's private information on demand shocks to the principal, so the principal can appropriately react to this information in choosing its ongoing investment.

Rao and Srinivasan (1995) extends Lal's model to allow for continuous investments/efforts and quadratic costs for both the principal and the agent, which is also the model setting we adopt. They show that in this set-up there is a role for wholesale per-unit prices in addition to a revenue sharing contract and a fixed fee. Somewhat similarly, Romano (1994) shows that adding resale price maintenance (RPM) to a standard two-part tariff in a setting with double-sided moral hazard and complete information can strictly improve the principal's profits (see also Bhattacharyya and Lafontaine, 1995). Again, none of these papers allow for the agent to have private information about demand. Furthermore, we show that RPM is not necessary for the optimal contract in our model.

In contrast to this earlier literature, Foros et al. (2009) allows for agents to have private information, but does not allow the principal to face a moral hazard problem. Specifically, they analyze price-dependent profit sharing in which the share of the profits extracted by the principal depends on the prices set by the agents. Thus, they provide a theory of ad valorem wholesale fees based on softening downstream competition and transmitting the agents' private information. Making the profit share extracted by the principal decrease in the prices set by the agents softens downstream competition to maximize industry profit. By itself, this could also be achieved with a two-part tariff (i.e., setting the wholesale price to achieve a particular retail price). However, the presence of the agents' private information means that the wholesale price would need to vary with the demand shock in order to maximize industry profit, which is why price-dependent profit sharing is better than two-part tariffs.

The key difference between Foros et al. (2009) and our work is that we allow for the principal to face a moral hazard problem. As a result, the goal of our information transmission mechanism is to provide the principal with the right investment incentives, whereas in their paper information transmission helps the principal provide competing agents with the right pricing incentives. Furthermore, we show that our ad valorem contracts are optimal rather than just being better than two-part tariffs. And this result holds even when we allow for competing agents. Indeed, we show that with linear demands and double-sided moral hazard, price-dependent profit-sharing is not necessary and is in fact dominated by our ad valorem contracts.

In platform contexts, there have been several recent attempts to explain the widespread use of ad valorem fees rather than per-unit fees by platforms like eBay. Shy and Wang (2011) focus on the market power of the platform and the sellers, and argue that making the fee proportional to price is preferred over a constant per-transaction fee because it helps mitigate double marginalization. Several other papers have a similar mechanism at work: see Foros et al. (2014), Gaudin and White (2014), and Johnson (2017). Llobet and Padilla (2016) show related results in the context of an innovator that has licensed its technology to a downstream firm and has to decide whether to use ad valorem or per-unit license fees. These explanations for ad valorem fees work by assuming away the ability of the platform to set an upfront fixed fee to sellers: ad valorem fees then help mitigate double marginalization. By contrast, we show ad valorem fees remain optimal even when an upfront fixed fee is feasible. Wang and Wright (2017) provide another quite distinct mechanism by which ad valorem fees can benefit a platform. Specifically, ad valorem fees help the platform price discriminate across different transactions when the value of a transaction to buyers tends to vary proportionally with the cost of the good traded (and so its price).

3 Model set-up

We assume the demand $D(p, q, Q)$ generated by a principal and an agent is determined by the choice of three decision variables: the price p and two costly actions, q and Q. Throughout the paper, Q is a costly, non-contractible action always chosen by the principal, and q is a costly, non-contractible action always chosen by the agent. These capture on-going investments made by the principal (e.g., brand name for franchisors, technological infrastructure for platforms) and ongoing effort expended by the agent (e.g., sales, customer service), both of which increase demand. Demand also depends on price p, which is chosen by the agent. Throughout the paper, we assume the principal can make its contracts contingent on the variables it can monitor (i.e., some function of price p and demand D).

For simplicity and tractability, we assume that demand is linear in these variables, and can be written as

$$D(p, q, Q) = \theta - \beta p + \phi q + \Phi Q,$$

where β, ϕ and Φ are positive constants, measuring the impact of p, q and Q, respectively, on demand, and θ is an additive demand shock only observed by the agent.[3] Formally, we assume $\theta > 0$ is a random variable drawn from some distribution function with a finite mean $\mathbb{E}(\theta) = \overline{\theta}$ and variance V_θ. We further assume the fixed costs of the respective costly actions are $\frac{1}{2}q^2$ and $\frac{1}{2}Q^2$. The total revenue net of fixed costs generated by the principal and the agent is therefore

$$pD(p, q, Q) - \frac{1}{2}q^2 - \frac{1}{2}Q^2.$$

The timing of the players' moves is as follows. In the first stage, the principal offers its contract and the agent decides whether or not to accept the contract. Without loss of generality, the agent's

[3] Obviously, the particular functional forms we have assumed are important for the exact optimality of the contract we derive in Section 4.1. However, the channels we uncover through which ad valorem contracts help the principal are quite general.

outside option is assumed to be equal to zero. If the contract is accepted, in the second stage the agent observes the realization of θ and decides on its price p. In the final stage, the agent chooses its ongoing effort or investment q, the principal chooses its ongoing investment Q, and payoffs are realized.[4]

We make the following assumption

$$\Omega \equiv \frac{\phi^2 + \Phi^2}{(2\beta - \phi^2)(\phi^2 + \Phi^2) - \Phi^4} > 0. \tag{1}$$

As we will show in Section 4, Ω measures the sensitivity of the optimal price to demand shocks. Assuming $\Omega > 0$ ensures that the optimal price increases with demand even after taking into account the indirect effects through the agent's and the principal's investments. Without any investments (i.e., if $\phi = \Phi = 0$), we would have $\Omega = \frac{1}{2\beta}$, which is the standard sensitivity of the monopoly price to changes in the demand intercept when demand is linear. Note that assuming $\Omega > 0$ is equivalent to assuming

$$\beta > \frac{\phi^4 + \phi^2\Phi^2 + \Phi^4}{2(\phi^2 + \Phi^2)},$$

so that the price sensitivity of demand is sufficiently important relative to the combination of both moral hazard problems. Assumption (1) ensures second-order conditions hold for the relevant optimization problems.

For the principal's problem to be interesting, there must be some principal moral hazard (i.e., $\Phi > 0$). If instead $\Phi = 0$, then the principal could attain the first-best outcome simply by using an upfront fixed fee to extract the agent's entire expected profit in the first stage (and no variable fees, reflecting that the principal's marginal cost is zero). This provides the agent with first-best incentives to invest in q and exploits the agent's private information. Thus, the principal's moral hazard provides a reason to have a positive wholesale price to ensure the principal will have an ongoing incentive to invest in raising demand for the product. Conversely, the agent's moral hazard is not essential for our results to hold: if $\phi = 0$, then the principal would find it optimal to give the agent a vanishingly small share ε of revenue $pD(p, q, Q)$. However, this is not enough to align the agent's pricing incentive with the principal's, because the agent ignores the cost of the principal's investment when setting price. This is why, the principal still needs a term in its ad valorem contract that depends non-linearly on price, as we show below. Finally, if the agent had no private information regarding θ, then, assuming the principal can place restrictions on the price, a second-best solution (in which just the double-sided moral hazard problem remains) would be a standard two-part tariff with an upfront fixed fee and constant per-unit wholesale price, together with RPM, as shown by Romano (1994). However, if RPM is not feasible because it may be deemed illegal by competition authorities (as is the case in many countries), then the standard two-part tariff (with no ad valorem component) is not second-best optimal even in the absence of private information. By contrast, the ad valorem contract we design below is second-best optimal with or without private information. Indeed, as we will show, our ad valorem contract has two advantages over a standard two-part tariff without RPM: the ability

[4] The case in which p, q and Q are all set at the same time, and/or in which the price p set by the agent is not always observed by the principal is analyzed in Section 5.3.

to incorporate the agent's private information about θ and the ability to remove the agent's pricing distortion.

4 Main results

To analyze the optimal contract in our setting, we first solve a hypothetical problem in which both the principal and the agent observe θ *before* contracting. In this case, as noted by Romano (1994), it is optimal for the principal to choose the price $p(\theta)$ and to set a two-part tariff $(w(\theta), T(\theta))$. Thus, for any given θ, the two-part tariff involves a constant per-unit wholesale fee w and an upfront fixed fee T.

With such a contract, the principal's profit is

$$\begin{aligned} & \max_{p(\theta),w(\theta),T(\theta)} \left\{ w(\theta) D(p(\theta), q(\theta), Q(\theta)) - \frac{1}{2} Q(\theta)^2 + T(\theta) \right\} \\ & \text{subject to} \\ T(\theta) \leq & \; (p(\theta) - w(\theta)) D(p(\theta), q(\theta), Q(\theta)) - \frac{1}{2} q(\theta)^2 \\ q(\theta) = & \; \arg\max_{q_0} \left\{ (p(\theta) - w(\theta)) D(p(\theta), q_0, Q(\theta)) - \frac{1}{2} q_0^2 \right\} = \phi (p(\theta) - w(\theta)) \\ Q(\theta) = & \; \arg\max_{Q_0} \left\{ w(\theta) D(p(\theta), q(\theta), Q_0) - \frac{1}{2} Q_0^2 \right\} = \Phi w(\theta). \end{aligned}$$

The first constraint is the agent's participation constraint and is binding. The second and third constraints reflect that in the second stage, the agent and principal choose their non-contractible decisions optimally for the given θ. Thus, the principal's program simplifies to

$$\begin{aligned} & \max_{w(\theta),p(\theta)} \left\{ p(\theta) D(p(\theta), q(\theta), Q(\theta)) - \frac{1}{2} q(\theta)^2 - \frac{1}{2} Q(\theta)^2 \right\} \\ & \text{subject to} \\ q(\theta) = & \; \phi (p(\theta) - w(\theta)) \text{ and } Q(\theta) = \Phi w(\theta). \end{aligned} \tag{2}$$

For a given θ, the solution is

$$p(\theta) = \theta \Omega \text{ and } w(\theta) = \left(\frac{\Phi^2}{\phi^2 + \Phi^2} \right) p(\theta), \tag{3}$$

which implies

$$q(\theta) = \left(\frac{\phi^3}{\phi^2 + \Phi^2} \right) \theta \Omega \text{ and } Q(\theta) = \left(\frac{\Phi^3}{\phi^2 + \Phi^2} \right) \theta \Omega. \tag{4}$$

Note (1) implies the second-order conditions hold for this case. Furthermore, (3) reveals that at the second-best solution, the optimal wholesale price w should be proportional to price p for any θ.[5]

[5] We are extremely grateful to an anonymous referee for providing us with this important insight, and for pushing us to refocus the paper on the implementation of ad valorem contracts with wholesale prices proportional to price, which is what we do below.

The resulting profits for a given θ are

$$\Pi(\theta) = \frac{\theta^2 \Omega}{2}.$$

If instead θ is not known at the contracting stage but is learnt by both parties after contracting, and the principal offers the contract $(p(\theta), w(\theta), T)$, where T is paid before θ is observed and $p(\theta)$ and $w(\theta)$ are determined ex-post by the actual realization of θ according to (3), then the expected profit that the principal would obtain is

$$\Pi^{SB} = \mathbb{E}_\theta[\Pi(\theta)] = \frac{\left(\overline{\theta}^2 + V_\theta\right)\Omega}{2}. \tag{5}$$

This profit is second-best, in the sense that the only problem left unsolved is double-sided moral hazard. Indeed, if $\phi = \Phi = 0$ so that $\Omega = \frac{1}{2\beta}$, then the expected profit corresponds to the usual monopoly profit, after taking expectations over different values of θ.

4.1 Implementation through an ad valorem contract

Consider now the actual setup in our model, in which θ is only observable to the agent after contracting, and the price is set by the agent in the second stage. Moreover, suppose the principal can now only contract on variables it can observe, namely the price set by the agent and demand (i.e., the number of units the agent sells). In this section we will show how the second-best outcome can be implemented through an ad valorem contract in which fees depend on the agent's price. We choose this implementation because it affords a simple characterization of the optimal contract, and allows us to explain the mechanisms by which ad valorem pricing achieves the best outcome. In Section 4.2 we will show the same outcome can also be implemented with an alternative contract, that involves price-dependent and quantity-dependent wholesale prices.

Suppose the principal offers an ad valorem contract in which the per-unit wholesale price depends on the price set according to $w(p) = wp$, and there is an additional payment made from the agent to the principal after the contract is accepted which depends only on p (i.e., it is fixed with respect to the number of units sold) and is denoted $F(p) = fp^2$. Here w and f are constant parameters of the contract chosen by the principal, with $w > 0$ being the share of revenue that the principal extracts. To extract the agent's expected profit, the principal's contract also involves an upfront fixed fee, denoted T.

In the final stage, after θ is realized and the agent has set a price p, the agent and the principal simultaneously choose

$$\begin{aligned} q &= \arg\max_{q_0} \left\{ ((1-w)p) D(p, q_0, Q) - \frac{1}{2} q_0^2 \right\} \\ Q &= \arg\max_{Q_0} \left\{ wpD(p, q, Q_0) - \frac{1}{2} Q_0^2 \right\}. \end{aligned}$$

This implies

$$q(p) = \phi(1-w)p \text{ and } Q(p) = \Phi w p.$$

Then in the second stage, the agent chooses

$$p = \arg\max_{p_0} \left\{ (1-w) p_0 D(p_0, q(p_0), Q(p_0)) - \frac{1}{2} q(p_0)^2 - f p_0^2 \right\}$$

implying

$$p(\theta) = \frac{(1-w)\theta}{\Delta(w,f)}, \tag{6}$$

where

$$\Delta(w,f) = (1-w)\left(2\beta - \phi^2 - \left(2\Phi^2 - \phi^2\right) w\right) + 2f.$$

It turns out that assumption (1) ensures $\Delta(w,f) > 0$ at the optimal levels of w and f chosen by the principal (determined below). This also ensures the second-order condition for $p(\theta)$ to maximize the agent's payoff holds.

Recall the second-best solution is fully characterized by (3)-(4). Thus, we can replicate the second-best outcome by setting

$$w = \frac{\Phi^2}{\phi^2 + \Phi^2} \tag{7}$$

$$f = \frac{\phi^2 \Phi^2 \left(\Phi^2 - \phi^2\right)}{2\left(\phi^2 + \Phi^2\right)^2}. \tag{8}$$

The agent's ex-post payoff for any given θ is then

$$(1-w) p(\theta) D(p(\theta), q(p(\theta)), Q(p(\theta))) - \frac{1}{2} q(p(\theta))^2 - f p(\theta)^2 = \frac{\phi^2 \theta^2 \Omega}{2\left(\phi^2 + \Phi^2\right)},$$

which is always positive. As a result, the principal can set its upfront fee T so as to extract all of the agent's ex-ante surplus, implying

$$T^* = \frac{\phi^2 \left(\overline{\theta}^2 + V_\theta\right) \Omega}{2\left(\phi^2 + \Phi^2\right)}. \tag{9}$$

Assumption (1) implies T^* is positive. Then, by construction, the principal's expected profit is equal to its expected profit in the second-best solution $\Pi^{SB} = E_\theta[\Pi(\theta)]$ given above. This leads to our main result.

Proposition 1. *If the agent sets the price, and contracts can only be contingent on the agent's price and/or realized demand, then the principal can achieve the second-best outcome with the contract defined by $w(p) = wp$, $F(p) = fp^2$ and $T = T^*$, where w is defined in (7), f is defined in (8), and T^* is defined by (9), obtaining the expected profit defined in (5).*

The proposition shows that the principal can achieve the second-best outcome with a relatively simple contract, which has the property that the per-unit wholesale price is proportional to the price (i.e., is an example of an ad valorem contract). Since the wholesale price is paid to the principal on each unit sold, the ad valorem wholesale price is also equivalent to a revenue sharing contract in which the agent shares a constant percentage of its revenue with the principal. The additional terms of the ad valorem contract do not depend on the number of units sold—$F(p)$ is a fee that only varies with the price charged and T is an upfront fixed fee. [6]

To understand why the principal wants to use an ad valorem wholesale price, note that if θ is high so demand is high, then the agent will want to set a high price, which increases the net surplus created by the principal's investment. By setting a high per-unit wholesale price when the agent's price is high, this contract creates an incentive for the principal to invest more when demand is high. This provides a key benefit of using ad valorem wholesale pricing which would not arise with a constant per-unit wholesale price. In the optimal contract, the wholesale price is set so as to optimally balance the two moral hazard problems. Given the demand shock is separate from the extent of moral hazard in our demand specification, this requires a constant ratio (with respect to θ) between q and Q. This explains why having a constant ratio between p and w, that is, a proportional ad valorem wholesale price, is indeed optimal in our specification.

The remaining features of the optimal contract must ensure the agent sets the price p appropriately. In the full information setting of Romano (1994), this was achieved by using an RPM contract in which the principal sets p. In the absence of RPM, an advantage of an ad valorem wholesale price (i.e., revenue sharing) in this regard is that it can eliminate the pricing distortion which would arise with constant per-unit wholesale prices due to the fact that the agent does not internalize the impact of its price on the principal's payoff.[7] Indeed, from the perspective of revenue maximization, revenue sharing ensures that the agent's incentives are aligned with the principal with respect to setting the price to maximize revenue. However, this alone will not necessarily ensure the agent sets the price that is best for the principal. This is because the agent receives a constant share of the revenue but bears the full cost of its investment in q, the level of which increases in p, while at the same time it does not bear any of the cost of the principal's investment in Q, the level of which also increases in p. Depending on which type of moral hazard is more important, the price set by the agent based purely on the ad valorem wholesale price (i.e., a constant revenue share) could therefore be set too high or too low.

To understand this better, let us start with the special case in which $\Phi = \phi$, so both the principal's and agent's moral hazard problems are equally important. Then the optimal w implies 50-50 revenue sharing. The agent collects 50% of the revenue and incurs exactly 50% of the total investment costs associated with increasing its price (since the agent's investment q and the principal's investment Q

[6] An important corollary of Proposition 1 is that RPM is not needed for optimality. This corollary applies regardless of whether the agent holds private information or not.

[7] Without the agent's moral hazard, this distortion always leads to prices being too high from the principal's perspective, which is the standard double marginalization effect. However, if the agent's moral hazard is sufficiently important, then this distortion leads to prices being too low. Specifically, demand is decreasing in price iff $\beta > \phi^2 (1 - w_1) + \Phi^2 w_1$, or equivalently at the optimal second-best contact, $\beta > \frac{\Phi^4 + \phi^4}{\Phi^2 + \phi^2}$. Since $\Phi^2 \phi^2$ can be either bigger or smaller than $\Phi^4 + \phi^4$, it is possible for demand to be increasing or decreasing in price even while assumption (1) holds.

increase in p at exactly the same rate but the agent only takes into account the effect through q). As a result, the agent sets a price that maximizes joint profit and there is no need for any additional payment contingent on p to ensure this.

Now suppose $\Phi > \phi$. Then based on the optimal w (which exceeds 1/2), the agent is left with less than 50% of the revenue and also incurs less than 50% of the total investment costs associated with increasing its price. Since the cost function is convex in the agent's share of revenue, whereas the revenue is linear, the latter effect dominates, meaning the agent sets the price too high from the perspective of joint profit maximization. As a result, the agent's price needs to be adjusted downwards, which is the purpose of the fp^2 term in the principal's contract. The converse is true if $\phi > \Phi$. This explains why the level of f in the optimal contract can be positive or negative, depending on whether the principal's or the agent's moral hazard is more important.

As noted above, the optimal second-best contract (or its implementation through an ad valorem contract) has two advantages relative to a standard two-part tariff (i.e., a constant per-unit wholesale fee and an upfront fixed fee). Firstly, it allows the principal's investment decision to respond in an appropriate way to the agent's private information about θ. Secondly, it removes the distortion in the agent's choice of p. We can isolate these two channels, to see how important each is.

To do so, first denote by Δ_I the expected gain to the principal from being able to adjust a standard two-part tariff to information on demand. This equals $E_\theta\left[\Pi^{2PT}(\theta)\right] - \Pi^{2PT}$, where $\Pi^{2PT}(\theta)$ is the principal's profit under the optimal two-part tariff for a given θ and Π^{2PT} is the principal's expected profit under the optimal two-part tariff given θ is not observed. In the appendix we show

$$\Delta_I = \frac{\Phi^4}{2\left(2\beta - \phi^2\right)\left(\left(2\beta - \phi^2\right)\Phi^2 + \beta^2 - \Phi^4\right)} V_\theta. \tag{10}$$

As expected, the information channel Δ_I disappears as $V_\theta \to 0$.

Second, denote by Δ_{PD} the expected gain to the principal from being able to use ad valorem fees so as to optimally deal with pricing distortions rather than the two-part tariff with full information. Thus, $\Delta_{PD} = \Pi^{SB} - \mathbb{E}_\theta\left[\Pi^{2PT}(\theta)\right]$. In the appendix we show

$$\Delta_{PD} = \left(\frac{\Phi^4\left(\beta - \phi^2\right)^2\Omega}{2\left(2\beta - \phi^2\right)\left(\phi^2 + \Phi^2\right)\left(\left(2\beta - \phi^2\right)\Phi^2 + \beta^2 - \Phi^4\right)}\right)\left(\overline{\theta}^2 + V_\theta\right). \tag{11}$$

Unlike the information channel, the price distortion channel remains even when $V_\theta = 0$, although a higher variance of θ still increases the importance of this channel. It does, however, disappear when $\beta = \phi^2$. To understand this, note that for any given positive level of the wholesale price in a standard two-part tariff contract (assuming θ is known to both parties, so that there is no private information), the agent has an upward bias relative to the principal in choosing the price if $\beta > \phi^2$ and a downward bias in choosing the price if $\beta < \phi^2$ (reflecting that a higher price can expand demand through a higher level of q).[8] Thus, when $\beta = \phi^2$, which is the case in which demand is independent of price under a

[8] Specifically, if the wholesale price is w, after substituting $q = \phi(p - w)$ and $Q = w\Phi$ into the demand function and the cost functions, the agent's optimal price is $p = \frac{\theta + \Phi^2 w + \left(\beta - \phi^2\right) w}{2\beta - \phi^2}$, whereas the principal's optimal price is $p = \frac{\theta + \Phi^2 w}{2\beta - \phi^2}$.

standard two-part tariff, there is no pricing bias in the two-part tariff, so no benefit of the ad valorem wholesale price over the standard two-part tariff in dealing with pricing distortions.

In the appendix we show that both Δ_I and Δ_{PD} are positive if $V_\theta > 0$ and $\beta \neq \phi^2$. Recall that the sum $\Delta_I + \Delta_{PD}$ captures the total expected gain to the principal from moving from the optimal two-part tariff to the optimal second-best contract. The gain from the information channel relative to the total gain is then given by the fraction

$$\frac{\Delta_I}{\Delta_I + \Delta_{PD}} = \frac{1}{1 + \frac{(\beta-\phi^2)^2\Omega}{\phi^2+\Phi^2}\left(\frac{\bar{\theta}^2}{V_\theta} + 1\right)}.$$

This expression can be decomposed into four terms: (i) $\left(\beta - \phi^2\right)^2$ which measures the magnitude of any upward or downward bias in the agent's choice of price; (ii) Ω which measures the sensitivity of the optimal price to demand shocks; (iii) $\phi^2 + \Phi^2$ which is a measure of the total investment effects (or, put differently, the overall extent of moral hazard problems); and (iv) $\frac{V_\theta}{\bar{\theta}^2}$ which measures the square of the coefficient of variation of θ. By focusing on one of these terms while holding the others constant, the result above shows that the information channel is more important when the total investment effects are high and the coefficient of variation of θ is high, while the price distortion channel is more important when there is a large upward or downward bias in the agent's choice of price and the optimal price is highly sensitive to demand shocks.

4.2 Implementation with price-dependent and quantity-dependent wholesale prices

Allowing for a component of the contract offered to the agent that depends only on price and not the number of units sold as we did in the previous section (i.e., allowing for the f term in the contract) provides the most direct implementation of the second-best solution. While such a scheme appears to be close to various forms of trade promotions in which retailers get discounts or rebates based on the price set (e.g., see Ailawadi et al, 1999), these are usually paid on a per-unit basis and thus differ from the f term in Proposition 1. In this section we show that using quantity-dependent wholesale prices in additional to price-dependent wholesale prices (i.e., revenue sharing) can also allow the principal to obtain its second-best expected profit as given in (5).

Consider the contract $w(p) = wp + tD(p, q, Q)$, so that the per-unit wholesale price depends linearly on the price and also on the level of demand. Note in a manufacturer-retail context, $t < 0$ implies the wholesale price paid by the retailer decreases in the amount it orders from the manufacturer, which is an example of incremental quantity discounting. Although this type of contract will introduce a p^2 term as in the previous section, it also affects the way q and Q are chosen, and introduces new interaction terms between q, Q and p. As a result, a priori, it is not obvious that such a contract can still achieve the second-best outcome. In the Online Appendix, we prove the following result.

Proposition 2. *Provided second-order conditions hold, there exists a wholesale contract* $w(p) = wp + tD(p, q, Q)$ *in which the principal obtains its second-best expected profit as defined in (5). The* *optimal contract can involve $w < 0$ or $t < 0$ (although not both).*

In the Online Appendix, we characterize the specific levels of w and t that replicate the second-best solution. These are long and rather messy functions of the underlying parameters ϕ, Φ and β such that the resulting $p(\theta)$, $q(p(\theta))$ and $Q(p(\theta))$ match those in (3)-(4). As a result, it is now much harder to interpret how the parameters β, ϕ and Φ influence the design of the contract. This is why we focused on the easier to interpret ad valorem contracts in Section 4.1.

4.3 Constant revenue share and upfront fixed fee

Suppose the principal can only charge a constant share of revenue and an upfront fixed fee (thus ruling out the implementations of the second-best solution analyzed in Sections 4.1-4.2). This seems to be the most common contract in practice, such as in many franchise contracts and licensing contracts. In this section, we characterize how much the principal gives up by using this simpler contract compared to the second-best outcome. Specifically, we work out the optimal contract when the principal can only use a wholesale price function $w(p) = wp$ and an upfront fixed fee T, and measure the difference between the principal's expected profit using this contract and the expected profit that the principal can obtain with a fully flexible ad valorem contract.

In the Online Appendix, we prove the following Proposition.

Proposition 3. *If the principal can only charge a constant revenue share w and an upfront fixed fee T, the loss of expected profit relative to the second-best expected profit is*

$$\frac{1}{1+\frac{1+\left(\frac{\phi}{\Phi}\right)^2}{\Phi^2\Omega}}\frac{\left(1-\left(\frac{\phi}{\Phi}\right)^2\right)^2}{\left(1-\left(\frac{\phi}{\Phi}\right)^2\right)^2+\frac{1+\left(\frac{\phi}{\Phi}\right)^2}{\Phi^2\Omega}}.$$

The loss in expected profit relative to the second-best is obtained by taking the difference between the principal's second-best profit given by (5) and the profit that the principal can achieve here (with a constant revenue share w and an upfront fixed fee T), and then dividing by the principal's second-best profit. This relative loss lies between zero (when the simpler contract also achieves the second-best outcome) and one.

As already mentioned in the previous section, if the two moral hazards are equal, there is no loss in profits in using the simpler contract in Proposition 3. If the difference between the two moral hazards is small, then the loss in profits is also small. Second, recall that Ω is decreasing in β, so if β is large relative to the moral hazards ϕ and Φ, then once again the loss in profits is small. Third, if the principal's moral hazard is very small (i.e., $\Phi \to 0$), then the relative loss in profits tends to zero. Indeed, without the principal's moral hazard, we already know that a standard two-part tariff is optimal. Fourth, if the agent's moral hazard is very small (i.e., if $\phi \to 0$), then the relative loss becomes equal to $\left(\frac{\Phi^2}{2\beta}\right)^2$, so the relative effects of β and Φ still apply, reflecting that the agent ignores the cost of the principal's investment when setting price.

Finally, given that the principal can no longer achieve the second-best outcome with this simple contract, there emerges a potential role for the principal to place additional restrictions on the price chosen by the agent. Consistent with this, Hagiu and Wright (2018) show that the principal can obtain higher profits by placing either a maximum or a minimum threshold constraint on the price chosen by the agent in the context of a two-part tariff contract (either a standard two-part tariff, or the one studied in this section based on constant revenue sharing).

5 Extensions

In this section we consider three extensions of our baseline model. In the first we introduce marginal costs for both the principal and the agent, and determine how costs impact the optimal ad valorem contract. In the second extension we introduce n competing agents and show that an ad valorem contract similar to the one determined in Section 4.1 remains optimal. In the final extension we show how the optimal ad valorem contract from Section 4.1 can be modified to allow for the possibility that p, q and Q are all set at the same time, and/or that the price p set by the agent is not always observed by the principal.

5.1 Production costs

In this section we extend the baseline model of Section 3 to allow the principal to incur a marginal cost c_P (e.g., a manufacturing cost) and the agent to incur a marginal cost c_A (e.g., a reselling cost). If both the principal and the agent know θ before contracting, the principal's second-best profit is obtained by solving

$$\max_{p(\theta),w(\theta),T(\theta)} \left\{ (w(\theta) - c_P)(\theta - \beta p(\theta) + \phi q(\theta) + \Phi Q(\theta)) - \frac{1}{2}Q(\theta)^2 + T(\theta) \right\}$$

subject to

$$\begin{aligned}
T(\theta) &\leq (p(\theta) - w(\theta) - c_A)(\theta - \beta p(\theta) + \phi q(\theta) + \Phi Q(\theta)) - \frac{1}{2}q(\theta)^2 \\
q(\theta) &= \arg\max_{q_0} \left\{ (p(\theta) - w(\theta) - c_A)(\theta - \beta p(\theta) + \phi q_0 + \Phi Q(\theta)) - \frac{1}{2}q_0^2 \right\} = \phi(p(\theta) - w(\theta) - c_A) \\
Q(\theta) &= \arg\max_{Q_0} \left\{ (w(\theta) - c_P)(\theta - \beta p(\theta) + \phi q(\theta) + \Phi Q_0) - \frac{1}{2}Q_0^2 \right\} = \Phi(w(\theta) - c_P).
\end{aligned}$$

Following the same steps as for the benchmark model, we obtain that for a given θ, the price $p(\theta)$, the wholesale price $w(\theta)$, and the investments $q(\theta)$ and $Q(\theta)$ are given by

$$\begin{aligned} p(\theta) &= \theta\Omega + (c_P + c_A)(1 - \beta\Omega) \\ w(\theta) &= \frac{\Phi^2 p(\theta) + \phi^2 c_P - \Phi^2 c_A}{\phi^2 + \Phi^2} \\ q(\theta) &= \phi(p(\theta) - w(\theta) - c_A) \\ Q(\theta) &= \Phi(w(\theta) - c_P). \end{aligned} \tag{12}$$

It is worth noting that the ratio between $q(\theta)$ and $Q(\theta)$ is still constant and equal to $\frac{\phi^2}{\Phi^2}$, as in the benchmark model. Meanwhile, the price $p(\theta)$ differs from its expression (3) in the benchmark model by the term $(c_P + c_A)(1 - \beta\Omega)$. Thus, perhaps surprisingly, the optimal price can be decreasing in costs if $\beta\Omega > 1$, which is equivalent to[9]

$$\frac{\phi^2 + \Phi^4\phi^2 + \Phi^4}{\phi^2 + \Phi^2} > \beta.$$

The intuition for this condition is as follows. If the agent's cost increases, the agent's investment will decrease, which in turn decreases demand and thus, on its own, leads to a lower price. Consequently, this effect will outweigh the usual effect of a cost increase on price when aggregate moral hazard (measured by the left-hand side of the condition) is more important than the price elasticity effect (measured by β), in which case the optimal price will be lower. The condition is symmetric in ϕ and Φ, and the same regardless of which party's cost increases because the principal sets $w(\theta)$ to optimally rebalance investment incentives.

The principal's resulting profit for a given θ is

$$\Pi(\theta) = \frac{(\theta - \beta(c_A + c_P))^2 \Omega}{2}.$$

As it turns out, our assumption (1) continues to ensure that second-order conditions hold throughout this section.

The second-best solution involves inducing the same $p(\theta)$, $w(\theta)$, $q(\theta)$ and $Q(\theta)$ given by (12), but since the principal does not observe θ, its second-best profits are simply the expected value of the first-best profits over θ, which equal

$$\Pi^{SB} = \frac{\left(\left(\overline{\theta} - \beta(c_A + c_P)\right)^2 + V_\theta\right)\Omega}{2}. \tag{13}$$

Relegating details of the analysis to the appendix, we show that the principal can implement this second-best through an ad valorem contract.

Proposition 4. *Suppose the principal and the agent incur marginal costs c_P and c_A respectively.*

[9] Note that this condition can hold at the same time as assumption (1).

Then the principal can achieve the second-best outcome with an ad valorem contract defined by $w(p) = w_0 + w_1 p$*,* $F(p) = f_0 + f_1 p + f_2 p^2$ *and* $T = T^*$*, where*

$$
\begin{aligned}
w_0 &= \frac{\phi^2 c_P - \Phi^2 c_A}{\phi^2 + \Phi^2}, \; w_1 = \frac{\Phi^2}{\phi^2 + \Phi^2} \\
f_0 &= \min\left\{\frac{1}{2}\frac{\phi^2 \Phi^2 \left(\Phi^2 - \phi^2\right)(c_A + c_P)^2}{\left(\Phi^2 + \phi^2\right)^2}, 0\right\} \\
f_1 &= \frac{\Phi^2 \phi^2 \left(\phi^2 - \Phi^2\right)(c_A + c_P)}{\left(\phi^2 + \Phi^2\right)^2}, \; f_2 = \frac{\Phi^2 \phi^2 \left(\Phi^2 - \phi^2\right)}{2\left(\phi^2 + \Phi^2\right)^2} \\
T^* &= \frac{\phi^2 \left(\left(\left(\overline{\theta} - \beta(c_A + c_P)\right)^2 + V_\theta\right)\Omega\left(\phi^2 + \Phi^2\right) + \Phi^2\left(\Phi^2 - \phi^2\right)(c_A + c_P)^2\right)}{2\left(\Phi^2 + \phi^2\right)^2} - f_0.
\end{aligned}
$$

Comparing the optimal ad valorem contract here with the one obtained for the baseline setting, we note that w_1 and f_2 are the same as w defined by (7) and f defined by (8). In case the agent's moral hazard is more important than the principal's moral hazard, an additional fixed subsidy is required which is paid to the agent ex-post but extracted by the principal's ex-ante. This ensures that for any realization of θ, the agent will continue to want to invest in effort and make positive profit rather than walking away from the venture. We also now have two new components to the optimal contract, w_0 and f_1, both of which depend on costs and the magnitudes of the moral hazards. To gain some intuition regarding their expressions, consider first the case in which the moral hazard problems are equally important, so $\phi^2 = \Phi^2$. This implies $f_1 = 0$ and $w_0 = \frac{c_P - c_A}{2}$. The reason for this result is that, since the moral hazard problems are equal, from the joint profit maximization perspective, the equilibrium investments should be equal. However, both the principal's and the agent's investments are distorted downwards by the marginal costs, so the constant per-unit wholesale price w_0 has to adjust for differences in marginal costs. Specifically, if the agent's marginal cost is higher, the agent's investment is too low relative to the principal's, so the principal needs to set a negative w_0 such that the effective marginal costs per unit are equalized. And vice-versa when the principal's marginal cost is higher.

Next, consider the case in which marginal costs are equal, so $c_P = c_A = c$. This implies the principal's effective marginal cost per unit is $c - w_0 = \frac{2\Phi^2 c}{\phi^2 + \Phi^2}$, whereas the agent's effective marginal cost per unit is $c + w_0 = \frac{2\phi^2 c}{\phi^2 + \Phi^2}$. Thus, the marginal cost per unit is allocated in proportion to the relative importance of moral hazards in order to minimize the overall distortion caused by the principal and the agent not taking into account the joint marginal cost when choosing their respective investments. This means that, for instance, when the principal's moral hazard is more important, ignoring the agent's marginal cost when the principal makes its investment decision induces a bigger distortionary effect than ignoring the principal's marginal cost when the agent makes its investment decision. This explains why w_0 is increasing in ϕ and decreasing in Φ, where recall w_0 is a payment from the agent to the principal.

To understand why marginal costs only affect the constant per-unit wholesale fee w_0 and not the ad valorem component of the wholesale fee w_1, recall that the role of w_1 is to make the principal's investment respond to the demand shock observed by the agent in proportion to the relative importance of the principal's moral hazard. Here, the principal also needs to make sure that the total marginal cost is also allocated in inverse proportion to the relative moral hazards. This can be achieved through a constant per-unit wholesale fee, which does not interfere with the role played by the ad valorem component in transmitting demand information.

Finally, just as f_2 compensated for the effect of w_1 on price when there is asymmetry in moral hazards (as explained in the benchmark setting), f_1 compensates for the effect of w_0 on price when there is asymmetry in moral hazards, except with the opposite sign. To see this more clearly, suppose $c_P = c_A$, and $\Phi > \phi$. This implies $w_0 < 0$, so that the principal must bear more than half of joint costs. This in turn implies that the agent sets its price too low due to the subsidy $-w_0 > 0$, a distortion which the principal can correct by providing an incentive to increase price through the payment to the principal $f_1 p$, where f_1 is now negative.

5.2 Competing agents

In this section we extend the baseline model of Section 3 by allowing for multiple agents that each incur effort and that compete in price. Suppose there are n agents and the demand for agent i is

$$D_i = \frac{1}{n}\left(\theta - \frac{\beta}{1-s}p_i + \frac{\beta s}{1-s}\frac{1}{n}\sum_{j=1}^{n} p_j\right) + \phi q_i + \Phi Q, \tag{14}$$

where $\beta > 0$ as before, and $s \in [0,1]$ is the degree of substitutability (competition) among agents. The relationship between demand and prices for the n agents can be derived from the representative consumer problem in Shubik and Levitan (1980), and has been used by many other authors including Foros et al. (2009). Their specification arises if $\beta = 1$, $\Phi = \phi = 0$. A nice feature of the specification is that when all agents' prices are equal to p, agent i's demand is $\frac{1}{n}(\theta - \beta p) + \phi q_i + \Phi Q$, so does not depend on s. Clearly when $n = 1$, this just gives our baseline specification of demand. Note also that when $s = 0$, so there is no price competition, we get our baseline specification but with n independent agents.

We modify assumption (1) to

$$\Omega_n \equiv \frac{\phi^2 + n\Phi^2}{(2\beta - n\phi^2)(\phi^2 + n\Phi^2) - n^3\Phi^4} > 0. \tag{15}$$

Again, Ω_n measures the sensitivity of the optimal price (set by the principal to maximize joint profits) to demand shocks. Assuming $\Omega_n > 0$ ensures that the optimal price increases with demand still, and that second-order conditions hold throughout this section.

As in Section 4, we start by deriving the principal's second-best profits assuming θ is known to both the principal and the agents, and the principal can control the prices p_i. In the appendix, we prove that Romano (1994)'s result that the principal can restrict attention to linear two-part tariffs extends to this case with competing agents and linear demand. Thus, to obtain second-best profits, the principal solves

$$\max_{(w_i,p_i)_{i=1,..,n}} \left\{ \sum_{i=1}^{n} \left(p_i D_i - \frac{1}{2} q_i^2 \right) - \frac{1}{2} Q^2 \right\}$$

$$\text{s.t. } q_i = \phi \left(p_i - w_i \right) \text{ for all } i \text{ and } Q = \Phi \sum_{i=1}^{n} w_i,$$

where D_i is given in (14).

It is easily verified that the solution to this problem is symmetric and involves

$$\begin{aligned} p_i &= p(\theta) = \theta \Omega_n \\ w_i &= w(\theta) = \frac{n\Phi^2}{\phi^2 + n\Phi^2} p(\theta) \\ q_i &= q(\theta) = \phi \left(p(\theta) - w(\theta) \right) \\ Q &= Q(\theta) = n\Phi w(\theta) \end{aligned} \tag{16}$$

for all $i = 1, .., n$. The principal's resulting profits are

$$\frac{\theta^2 \Omega_n}{2}.$$

In terms of addressing the double-sided moral hazard problem, the constant relationship between the retail price and the wholesale price remains essentially the same, with the only difference being that Φ^2 is multiplied by n to reflect that the principal's investment applies to all n agents' demands. Note the price does not depend on s, which reflects the special features of the demand specification (i.e., aggregate demand does not depend on s at symmetric prices), and that without any moral hazard the price is the usual monopoly price $\frac{\theta}{2\beta}$.

The second-best solution involves inducing the same $p(\theta)$, $w(\theta)$, $q(\theta)$ and $Q(\theta)$ given by (16), but since the principal does not observe θ, its second-best profits are simply the expected value of the second-best profits over θ, which equal

$$\Pi^{SB} = \frac{\left(\overline{\theta}^2 + V_\theta \right) \Omega_n}{2}. \tag{17}$$

Relegating the proof to the appendix, we show that the principal can implement this second-best through an ad valorem contract.

Proposition 5. *Suppose that there are n competing agents who each set their own prices and whose demands are given by (14). Then the principal can achieve the second-best outcome with an ad valorem* *contract defined by* $w(p) = wp$, $F(p) = fp^2$ *and* $T = T^*$, *where*

$$\begin{aligned} w &= \frac{n\Phi^2}{\phi^2 + n\Phi^2} \\ f &= \frac{n\phi^2\Phi^2\left(\Phi^2 - \phi^2\right)}{2\left(\phi^2 + n\Phi^2\right)^2} - \frac{s}{1-s}\frac{\beta(n-1)\phi^2}{2n^2\left(\phi^2 + n\Phi^2\right)} \\ T^* &= \frac{\phi^2\left(\overline{\theta}^2 + V_\theta\right)\Omega_n^2}{2n\left(\phi^2 + n\Phi^2\right)^2}\left(\frac{\phi^2 + n\Phi^2}{\Omega_n} + n^2(n-1)\Phi^4 + \frac{s\beta(n-1)\left(\phi^2 + n\Phi^2\right)}{(1-s)n}\right). \end{aligned}$$

First, note that if $s = 0$, then w and f represent the generalization of the expressions in the benchmark model to the case with n independent agents. When $s > 0$, the expression of w remains unchanged because its role is to balance the moral hazard of the principal against the moral hazard of each agent. Meanwhile, f is modified by subtracting a term that is increasing in s, so that f is decreasing in s. Thus, when there is more intense competition among agents, the principal must provide stronger incentives to agents to increase their prices in order to offset competition so as to achieve the second best outcome.

5.3 Alternative timing and misreporting of price

In this section we consider an alternative setting to that of Section 3 which allows for the possibility that the principal does not observe the agent's price and/or that the agent's price is set at the same time as q and Q.

Specifically, consider the following alternative timing. In stage 1, as before, the principal offers its contract and the agent decides whether or not to accept the contract. If the contract is accepted, in stage 2 the agent observes the realization of θ and announces its price p to the principal. We denote this announced (or reported) price as p_r. Note the agent does not have to tell the truth.[10]

In the final stage (stage 3), the agent chooses q and the actual price p, while the principal chooses Q. With probability $0 < \eta \leq 1$ the principal observes the actual price, while with probability $1 - \eta$ it does not. If the principal observes the price, then it can potentially extract a penalty from the agent for misreporting, if this is written into the contract. Finally, payoffs are realized.

Since the principal does not observe the agent's actual price until after q and Q are set, nothing would change in this setting if the agent decided the actual price in stage 2, as in our benchmark model. The key difference from our benchmark model is then that the principal does not observe the actual price in stage 2 but rather only observes the reported price, with the principal observing the actual price after stage 3 with probability η.

With this alternative setup, in the Online Appendix we prove the following result.

Proposition 6. *Suppose the price* p *is set in stage 3, at the same time as* q *and* Q*, but the agent can make a price announcement* p_r *in stage 2. And suppose that with probability* $0 < \eta \leq 1$ *the actual price*

[10] A high profile example is Saks Fifth Avenue chain which was found to be inappropriately collecting "markdown money" from vendors by misreporting its actual sales prices. See Rozhon (2005).

is observed by the principal after it has set Q in stage 3, while with probability $1-\eta$ the actual price is not observed by the principal. Provided second-order conditions hold, the principal can achieve the second-best outcome and expected profit (5) with the contract defined by $w(p_r) = wp_r$, $F(p_r) = fp_r^2$ and $T = T^*$, together with the penalty $\frac{\Phi^2(\beta-\phi^2)}{2\eta(\phi^2+\Phi^2)}\left(p^2-p_r^2\right)$ paid from the agent to the principal when p is observed by the principal, where w is defined in (7), f is defined in (8), and T^* is defined by (9).

The proposition states that even under the alternative timing specified above, and even if the principal does not always monitor the actual price set by the agent, the same ad valorem contract as defined in Proposition 1 continues to deliver the second-best outcome to the principal, provided it is applied using the price reported by the agent, and provided the principal imposes the right penalty in its contract for misreporting.[11] The penalty function is designed to ensure the agent will always truthfully self-report the price it intends to set (or has set, if it actually sets the price in stage 2). With the price reported correctly in stage 2, then the optimal ad valorem contract specified in Proposition 1 applies to the reported price, and the same logic of Section 4.1 continues to hold.

To understand how the penalty works, recall from the discussion in Section 4.1 that if $\beta > \phi^2$ then the agent suffers from an upward bias in setting the price when the wholesale price it faces is independent of the actual price it sets, which is the case here. So without any penalty, the agent will prefer to set a higher price than the one reported, which is bad for the principal since demand is then too low from its perspective (it extracts less wholesale revenue). The penalty function corrects for this bias. Taking into account the probability that the principal will observe the agent's actual price, it results in the agent wanting to set the same price that it reports. Obviously, the lower the chance the actual price is observed by the principal, the higher the penalty needs to be. In case, $\beta < \phi^2$, the logic is just the reverse. The agent now suffers from a downward bias in setting its price, and so to offset this bias, the agent faces a penalty whenever it sets its price below its reported price. This implies that the optimal contract defined in Proposition 6 never involves a negative penalty in which the principal pays the agent for misreporting.[12] Furthermore, it implies in the special case in which $\beta = \phi^2$, no penalty is needed to ensure truthful self-reporting.

6 Concluding remarks

The emergence of digital monitoring and data analytics technologies has created more opportunities for firms to enforce ad valorem contracts (contracts that depend on prices) in a cost effective way. The key contribution of this paper is to show how such contracts can solve the channel coordination problem in a natural setting in which demand depends on ongoing investments by both the upstream and downstream firms, and in which the downstream firms (i.e., the agents) have better information

[11] In the absence of limited liability, the principal could obtain the same outcome by setting a very large (infinite) penalty for misreporting, so that the agent always reports truthfully (given $\eta > 0$).

[12] Thus, the penalty in the optimal contract can alternatively be defined so it equals $\max\left\{\frac{\Phi^2(\beta-\phi^2)}{2\eta(\phi^2+\Phi^2)}\left(p^2-p_r^2\right),0\right\}$ if $\beta > \phi^2$ and equals $\max\left\{\frac{\Phi^2(\phi^2-\beta)}{2\eta(\phi^2+\Phi^2)}\left(p_r^2-p^2\right),0\right\}$ if $\beta < \phi^2$.

on demand. It therefore adds to the growing literature trying to explain the widespread use of ad valorem fees (in particular, revenue sharing).

At a high level, ad valorem contracts allow upstream firms (principals) to preserve their own incentives to make ongoing investments in the channel, deal with pricing distortions caused by channel coordination problems, while at the same time adjusting their investment based on demand shocks that are only observed by the agents (but which can be revealed through the agents' prices). To achieve these objectives, such ad valorem contracts can make use of revenue sharing (to balance investment incentives and make the principal's investment responsive to demand shocks through price), upfront fixed fees (to extract the agents' expected profit) and an additional term that depend non-linearly on either price or demand to correct for remaining pricing distortions. We find the loss in profits from ignoring this additional non-linear term (i.e., from only relying on a constant revenue share with an upfront fixed fee) is relatively small provided the importance of the principal's and the agents' investments are not too different or when the principal's ongoing investment is not so important. Moreover, we show that if the principal does make use of this additional non-linear term, then it cannot do better controlling the price set by agents by using resale price maintenance.

We have explored the robustness of our optimal ad valorem contract to competition between agents, production costs, and imperfect monitoring of the agents' prices. While we focused on a simple linear-quadratic model of demand and costs for tractability, future work could explore other demand specifications and other types of private information. Obviously, the exact details of the optimal ad-valorem contract will be affected, but we expect the underlying mechanisms we uncovered (in correcting for pricing distortions and providing an information channel) to remain relevant. Future work could also explore robustness to the introduction of risk aversion and/or wealth constraints for agents, which would limit the upfront fees that can be extracted by the principal. This should have an effect similar to the principal's moral hazard on the remaining terms of the contract. Another extension would be to allow the agent to know its private information at the time of deciding whether to accept or reject the contract offered by the principal. This would mean the principal faces a screening problem and so is no longer able to extract all (expected) profit from the channel. It would be interesting to explore the extent to which the benefits of ad valorem contracts remain at work in such a setting.

References

Bhattacharyya, S. and F. Lafontaine (1995) "Double-Sided Moral Hazard and the Nature of Share Contracts," *RAND Journal of Economics*, 26(4), 761-781.

Cachon, G. P. and M. Lariviere (2005) "Supply Chain Coordination with Revenue-Sharing Contracts: Strengths and Limitations," *Management Science*, 51(1), 30–44.

Foros, Ø., K.P. Hagen, H.J. Kind (2009) "Price-Dependent Profit Sharing as a Channel Coordination Device," *Management Science*, 55(8), 1280-1291.

Foros, Ø., H.J. Kind and G. Shaffer (2014) "Turning the Page on Business Formats for Digital Platforms: Does Apple's Agency Model Soften Competition?" Working Paper.

Gaudin, G. and A. White (2014) "On the Antitrust Economics of the Electronic Books Industry." Working Paper.

Hagiu, A. and J. Wright (2018) "Resale Price Maintenance and Minimum Requirements," Working paper.

Johnson, J.P. (2017) "The Agency Model and MFN Clauses," *Review of Economic Studies*, 84(3), 1151-1185.

Lal R. (1990) "Improving Channel Coordination through Franchising," *Marketing Science*, 9(4), 299–318.

Llobet, G. and J. Padilla (2016) "The Optimal Scope of the Royalty Base in Patent Licensing," *Journal of Law and Economics*, 59(1), 45–73.

Rao, R.C. and S. Srinivasan (1995) "Why Are Royalty Rates Higher in Service-Type Franchises?" *Journal of Economics & Management Strategy*, 4(1), 7-31.

Romano, R. E. (1994) "Double Moral Hazard and Resale Price Maintenance," *Rand Journal of Economics*, 25(3), 455–466.

Rozhon, T. (2005) "First the markdown, then the showdown," New York Times, February 25, C1.

Shubik, M. and R. Levitan (1980). *Market Structure and Behavior*. Harvard University Press, Cambridge, MA.

Shy, O. and Z. Wang (2011) "Why Do Payment Card Networks Charge Proportional Fees?" *American Economic Review*, 101, 1575–1590.

Suits, D.B. and R. A. Musgrave (1953) "Ad Valorem and Unit Taxes Compared," *Quarterly Journal of Economics*, 67(4), 598-604.

Wang, Z. and J. Wright (2017) "Ad Valorem Platform Fees, Indirect Taxes, and Efficient Price Discrimination." *RAND Journal of Economics*, 48(2), 467–484.

7 Appendix

7.1 Decomposition of profits at the end of Section 4.1

We calculate Δ_I and Δ_{PD}. To obtain these, consider first standard two-part tariffs. The upfront fixed fee T is set such that the principal extracts the entire expected payoff in excess of the agent's outside option, so the principal solves

$$\begin{aligned}
& \max_{w}\left\{\mathbb{E}_{\theta}\left[p\left(\theta-\beta p+\phi q+\Phi Q\right)-\frac{1}{2}q^{2}\right]-\frac{1}{2}Q^{2}\right\} \\
& \text{subject to} \\
p &= \arg\max_{p'}\left\{\left(p'-w\right)\left(\theta-\beta p'+\phi q+\Phi Q\right)-\frac{1}{2}q^{2}\right\}=\frac{\theta+\phi q+\Phi Q}{2\beta}+\frac{w}{2} \\
q &= \phi\left(p-w\right)\text{ and }Q=\Phi w.
\end{aligned}$$

Using all three constraints to solve for p as a function of (w,θ), we obtain

$$p\left(\theta\right)=\frac{\theta+\left(\Phi^{2}+\beta-\phi^{2}\right)w}{2\beta-\phi^{2}}.$$

Using $q=\phi\left(p\left(\theta\right)-w\right)$ and $Q=\Phi w$, substituting the three constraints back into the expression of the principal's profits, and maximizing over w, we obtain the principal's optimal profit with two-part tariffs:

$$\Pi^{2PT}=\frac{\overline{\theta}^{2}\left(\beta\left(2\Phi^{2}+\beta\right)-\Phi^{2}\phi^{2}\right)}{2\left(2\beta-\phi^{2}\right)\left(\left(2\beta-\phi^{2}\right)\left(\Phi^{2}+\phi^{2}\right)+\left(\beta-\phi^{2}\right)^{2}-\Phi^{4}\right)}+\frac{V_{\theta}}{2\left(2\beta-\phi^{2}\right)}.$$

Supposing the principal knows θ, then its profit is

$$\Pi^{2PT}\left(\theta\right)=\frac{\theta^{2}\left(\beta\left(2\Phi^{2}+\beta\right)-\Phi^{2}\phi^{2}\right)}{2\left(2\beta-\phi^{2}\right)\left(\left(2\beta-\phi^{2}\right)\left(\Phi^{2}+\phi^{2}\right)+\left(\beta-\phi^{2}\right)^{2}-\Phi^{4}\right)},$$

which we obtain by setting $V_{\theta}=0$ in Π^{A*} and $\overline{\theta}=\theta$. Then given $\Delta_{I}=\mathbb{E}_{\theta}\left[\Pi^{2PT}\left(\theta\right)\right]-\Pi^{2PT}$, we get

$$\Delta_{I}=\frac{\Phi^{4}}{2\left(2\beta-\phi^{2}\right)\left(\left(2\beta-\phi^{2}\right)\Phi^{2}+\beta^{2}-\Phi^{4}\right)}V_{\theta},$$

which is positive due to (1).

Moreover, using the definition of Π^{SB} from the main text, and that $\Delta_{PD}=\Pi^{SB}-\mathbb{E}_{\theta}\left[\Pi^{2PT}\left(\theta\right)\right]$, we get

$$\Delta_{PD}=\left(\frac{\Phi^{4}\left(\beta-\phi^{2}\right)^{2}}{2\left(2\beta-\phi^{2}\right)\left(2\beta\left(\phi^{2}+\Phi^{2}\right)-\phi^{4}-\Phi^{2}\phi^{2}-\Phi^{4}\right)\left(\left(2\beta-\phi^{2}\right)\Phi^{2}+\beta^{2}-\Phi^{4}\right)}\right)\left(\overline{\theta}^{2}+V_{\theta}\right),$$

which is also positive due to (1).

The expression in the denominator of the large brackets is positive, which follows from (1), together with the requirement that $\left(2\beta-\phi^{2}\right)\Phi^{2}+\beta^{2}-\Phi^{4}>0$, which is necessary for the second-order condition to hold in the principal's profit maximization problem with respect to the wholesale price.

Note that optimal wholesale price set by the principal for given θ in the problem with the standard two-part tariff is

$$w^* = \frac{\Phi^2 \theta}{\left(2\beta - \phi^2\right)\Phi^2 + \beta^2 - \Phi^4}.$$

Since this should be positive, we get $\left(2\beta - \phi^2\right)\Phi^2 + \beta^2 - \Phi^4 > 0$. This together with (1) imply Δ_I and $\Delta_{PD} > 0$.

7.2 Proof of Proposition 4

We prove that the principal can obtain its second-best profits (13) by using an ad valorem contract of the form $w(p) = w_0 + w_1 p$ and $F(p) = f_0 + f_1 p + f_2 p^2$. In the final stage, after θ is realized and the agent has set a price p, the agent and the principal simultaneously choose

$$\begin{aligned} q &= \arg\max_{q_0} \left\{ \left((1-w_1)p - w_0 - c_A\right)\left(\theta - \beta p + \phi q_0 + \Phi Q\right) - \frac{1}{2} q_0^2 \right\} \\ Q &= \arg\max_{Q_0} \left\{ \left(w_0 + w_1 p - c_P\right)\left(\theta - \beta p + \phi q + \Phi Q_0\right) - \frac{1}{2} Q_0^2 \right\}. \end{aligned}$$

This implies

$$\begin{aligned} q(p) &= \phi\left((1-w_1)p - w_0 - c_A\right) \\ Q(p) &= \Phi\left(w_0 + w_1 p - c_P\right). \end{aligned}$$

Then in the second stage, the agent chooses

$$p = \arg\max_{p_0} \left\{ \left((1-w_1)p_0 - w_0 - c_A\right)\left(\theta - \beta p_0 + \phi q(p_0) + \Phi Q(p_0)\right) - \frac{1}{2} q(p_0)^2 - f_1 p_0 - f_2 p_0^2 \right\}.$$

It is easily verified that assumption (1) in the main paper ensures that second-order conditions hold. We obtain

$$p(\theta) = \frac{\theta(1-w_1) - f_1 + \left(\beta + \Phi^2 - \phi^2(1-w_1) - 2\Phi^2 w_1\right) w_0 + \left(\beta - \phi^2(1-w_1) - \Phi^2 w_1\right) c_A - \Phi^2 (1-w_1) c_P}{2(1-w_1)\beta - \phi^2(1-w_1)^2 - 2\Phi^2 w_1 (1-w_1) + 2 f_2}.$$

Thus, we have

$$\begin{aligned} p(\theta) &= \frac{\theta(1-w_1) - f_1 + \left(\beta + \Phi^2 - \phi^2(1-w_1) - 2\Phi^2 w_1\right) w_0 + \left(\beta - \phi^2(1-w_1) - \Phi^2 w_1\right) c_A - \Phi^2 (1-w_1) c_P}{2(1-w_1)\beta - \phi^2(1-w_1)^2 - 2\Phi^2 w_1 (1-w_1) + 2 f_2} \\ w(\theta) &= w_0 + w_1 p(\theta) \\ q(\theta) &= \phi\left((1-w_1)p(\theta) - w_0 - c_A\right) \\ Q(\theta) &= \Phi\left(w_0 + w_1 p(\theta) - c_P\right). \end{aligned}$$

The contract specified in Proposition 4 then ensures the above expressions match those of the second-best solution as given by (12).

Finally, note that the agent's ex-post profit in equilibrium is

$$((1-w_1)\,p(\theta)-w_0-c_A)\,(\theta-\beta p(\theta)+\phi q(\theta)+\Phi Q(\theta))-\frac{1}{2}q(\theta)^2-f_1 p(\theta)-f_2 p(\theta)^2.$$

After plugging in the second-best values of $p(\theta)$, $q(\theta)$ and $Q(\theta)$ and the values of (w_0, w_1, f_1, f_2) from the contract specified in Proposition 4, it can be shown that the agent's equilibrium ex-post equilibrium profit is equal to

$$\pi_A(\theta)=\frac{1}{2}\frac{\phi^2}{(\Phi^2+\phi^2)^2}\left((\theta-\beta(c_A+c_P))^2\,\Omega\left(\phi^2+\Phi^2\right)+\Phi^2\left(\Phi^2-\phi^2\right)(c_A+c_P)^2\right).$$

To ensure that $\pi_A(\theta)$ is positive for all θ (so that the agent continues to participate regardless of the realization of θ), the principal can charge the agent the ex-post fee f_0 given in Proposition 4, which is non-positive. And the principal can then extract the entire expected surplus from the agent ex-ante with an upfront fixed fee $T^*=E_\theta[\pi_A(\theta)]-f_0$, which explains the expression for T^* in Proposition 4.

7.3 Proof of optimality of linear two-part tariffs in Section 5.2

Suppose both the principal and the agents can observe θ and the principal controls prices p_i for all $i=1,..,n$. Denote the second-best contract by $W^*(D_i)$, where $W^*(.)$ can be any function of D_i (given the symmetry of the problem and linearity of demands, the optimal contract must be symmetric, i.e., the same for all agents). We will show that the principal can replicate the outcome of this contract with a linear two-part tariff of the form $w^* D_i + T^*$.

Denote by (p^*, q^*, Q^*, D^*) the symmetric equilibrium induced by the first-best contract $W^*(.)$, where recall

$$D^*=\frac{1}{n}(\theta-\beta p^*)+\phi q^*+\Phi Q^*.$$

First, $W^*(.)$ must be continuous and differentiable at D^*. The argument is the same as in Romano (1994): a discontinuity or non-differentiability at D^* would imply either $q^*=0$ or $Q^*=0$, which cannot be optimal given linear demand and quadratic costs.

Second, let w^* be the derivative of W^* with respect to D, evaluated at D^*. By definition of q^* and Q^* we must have

$$\begin{aligned} q^* &= \arg\max_q \left\{p^*\left(\frac{1}{n}(\theta-\beta p^*)+\phi q+\Phi Q^*\right)-W^*\left(\left(\frac{1}{n}(\theta-\beta p^*)+\phi q+\Phi Q^*\right)\right)-\frac{1}{2}q^2\right\} \\ &= \phi(p^*-w^*). \end{aligned}$$

$$Q^*=\arg\max_Q\left\{nW^*\left(\frac{1}{n}(\theta-\beta p^*)+\phi q^*+\Phi Q^*\right)-\frac{1}{2}Q^2\right\}=n\Phi w^*.$$

Clearly then, the principal can induce the same outcome (p^*, q^*, Q^*, D^*) with the contract $w^* D_i$, and set the upfront fixed fee T to extract the entire profit from the agents.

7.4 Proof of Proposition 5

We prove that the principal can obtain its second-best profits in (17) by using an ad valorem contract of the form $w(p) = wp$ and $F(p) = fp^2$, along with an upfront fixed fee T.

In the final stage, after θ is realized and each agent has set a price p_i, agent i and the principal simultaneously choose

$$\begin{aligned} q_i &= \arg\max_{q_0} \left\{ ((1-w)\,p_i)\,D_i - \frac{1}{2}q_0^2 \right\} = \phi\,(1-w)\,p_i \\ Q &= \arg\max_{Q_0} \left\{ w\sum_{i=1}^{n} p_i D_i - \frac{1}{2}Q_0^2 \right\} = \Phi w \sum_{i=1}^{n} p_i. \end{aligned}$$

Then in the second stage, agent i chooses

$$p_i = \arg\max_{p_i'} \left\{ (1-w)\,p_i' D_i - \frac{1}{2}q_i\left(p_i'\right)^2 - f\left(p_i'\right)^2 \right\}$$

which in the symmetric equilibrium implies

$$p(\theta) = \frac{1}{n}\frac{(1-w)\,\theta}{(1-w)\left(\left(\frac{1}{n^2}\frac{2n-(n+1)s}{1-s}\right)\beta - \phi^2 - \left((n+1)\,\Phi^2 - \phi^2\right)w\right) + 2f}.$$

The contract specified in Proposition 5 then ensures $p(\theta)$ above, along with $w = \frac{n\Phi^2}{\phi^2 + n\Phi^2}$ and the resulting levels of $q(\theta)$ and $Q(\theta)$, match those of the second-best solution as given by (16).

Finally, note that each agent's ex-post profit in equilibrium is

$$(1-w)\,p(\theta)\left(\frac{\theta - \beta p(\theta)}{n} + \phi q(\theta) + \Phi Q(\theta)\right) - \frac{1}{2}q(\theta)^2 - fp(\theta)^2.$$

After plugging in the second-best values of $p(\theta)$, $q(\theta)$ and $Q(\theta)$ and the values of (w, f) from the contract specified in Proposition 5, it can be shown that each agent's equilibrium ex-post equilibrium profit is equal to

$$\pi_A(\theta) = \frac{\theta^2\phi^2\Omega_n^2}{2n\left(\phi^2 + n\Phi^2\right)^2}\left(\frac{\phi^2 + n\Phi^2}{\Omega_n} + n^2(n-1)\,\Phi^4 + \frac{s\beta(n-1)\left(\phi^2 + n\Phi^2\right)}{(1-s)\,n}\right),$$

which is clearly positive. Thus, the principal can extract the agent's entire expected profit ex-ante with an upfront fixed fee $T^* = E_\theta\left[\pi_A(\theta)\right]$, which explains the expression for T^* in Proposition 5.

Online Appendix

This online appendix provides the proofs behind some of the results and claims in the main text.

A Proof of Proposition 2

Consider the case in Section 4.2 of the main paper. In the final stage, after θ is realized and the agent has set a price p, the agent chooses

$$q = \arg\max_{q_0} \left\{ (1-w)\, pD(p, q_0, Q) - \frac{1}{2} q_0^2 - tD(p, q_0, Q)^2 \right\}$$

and the principal simultaneously chooses

$$Q = \arg\max_{Q_0} \left\{ wpD(p, q, Q_0) - \frac{1}{2} Q_0^2 + tD(p, q, Q_0)^2 \right\}.$$

Provided $t \neq 0$ and second-order conditions hold, this leads to the equilibrium choices

$$\begin{aligned} q(p) &= \frac{\phi\left(\left(1 - w + 2t\left(\beta - \Phi^2\right)\right)p - 2t\theta\right)}{1 - 2t\left(\Phi^2 - \phi^2\right)} \\ Q(p) &= \frac{\Phi\left(2t\theta + \left(w - 2t\left(\beta - \phi^2\right)\right)p\right)}{1 - 2t\left(\Phi^2 - \phi^2\right)}. \end{aligned}$$

Evaluating the final stage choices of investments given p is set at the second-best level $p(\theta)$, and setting these equal to the second-best levels of investment we have that

$$q(p(\theta)) = \frac{\phi\left(\left(1 - w + 2t\left(\beta - \Phi^2\right)\right)p(\theta) - 2t\theta\right)}{1 - 2t\left(\Phi^2 - \phi^2\right)} = q(\theta)$$

and

$$Q(p(\theta)) = \frac{\Phi\left(2t\theta + \left(w - 2t\left(\beta - \phi^2\right)\right)p(\theta)\right)}{1 - 2t\left(\Phi^2 - \phi^2\right)} = Q(\theta)$$

provided

$$t(w) = \frac{(1-w)\,\Phi^2 - \phi^2 w}{2\beta\left(\phi^2 + \Phi^2\right) - 2\Phi^2\phi^2}.$$

Thus, provided t is set in this way (for any given w), at the second-best prices, the investments in the last stage will also both equal the second-best levels. Assumption (1) ensures the denominator is positive. Note in case $w < \frac{\Phi^2}{\Phi^2+\phi^2}$, this implies $t > 0$, while if $w > \frac{\Phi^2}{\Phi^2+\phi^2}$, this implies $t < 0$, which shows $w < 0$ and $t < 0$ is not possible.

Now consider the agent's pricing decision in the second stage. The agent solves

$$p = \arg\max_{p_0} \left\{ (1-w)\, p_0 D(p_0, q(p_0), Q(p_0)) - \frac{1}{2} q(p_0)^2 - tD(p_0, q(p_0), Q(p_0))^2 \right\}.$$

We need to make sure that w is set so that given $t(w)$ above, the optimizing choice of p_0 matches the second-best function $p(\theta)$.

Differentiating the agent's objective function with respect to p_0 we get

$$(1-w)\left(D+p_0\frac{dD}{dp_0}\right)-q\frac{\partial q}{\partial p_0}-2tD\frac{dD}{dp_0}=0,$$

which can be evaluated at the second-best solutions $p_0=p(\theta)$, $q=q(\theta)$ and $D=D(p(\theta),q(\theta),Q(\theta))$. Solving this first-order condition for w, after taking into account the relationship $t(w)$, implies

$$w=\frac{\sqrt{Y}-\left(2\Phi^4\phi^2+\beta^2\left(\phi^2+\Phi^2\right)-2\Phi^2\beta\left(\phi^2+\Phi^2\right)\right)}{2\left(\Phi^2-\phi^2\right)\left(\beta\left(\phi^2+\Phi^2\right)-\phi^2\Phi^2\right)}$$

where

$$Y=\beta^4\left(\Phi^2+\phi^2\right)^2+4\Phi^4\phi^4\left(4\beta^2+\Phi^2\phi^2\right)+4\beta\phi^2\Phi^2\left(\beta\left(\phi^4+\Phi^4-2\beta\left(\phi^2+\Phi^2\right)\right)-2\phi^2\Phi^2\left(\phi^2+\Phi^2\right)\right).$$

At this level of w and $t(w)$, the agent will chose p_0 to be equal to the second-best level $p(\theta)$, and the final stage investments q and Q will also replicate the second-best solution. Note using L'Hôpital's rule, when $\phi^2\to\Phi^2$ we can confirm that $w=\frac{1}{2}$ and $t=0$, as expected. Numerical examples show that even under (1), w can be negative or t can be negative.

B Proof of Proposition 3

In the final stage we have

$$\begin{aligned}
q &= \arg\max_{q_0}\left\{(1-w)\,p\,(\theta-\beta p+\phi q_0+\Phi Q)-\frac{1}{2}q_0^2\right\}=\phi(1-w)\,p\\
Q &= \arg\max_{Q_0}\left\{wp\,(\theta-\beta p+\phi q+\Phi Q_0)-\frac{1}{2}Q_0^2\right\}=\Phi wp.
\end{aligned}$$

Then in the second stage, the agent chooses

$$p(w,\theta)=\arg\max_{p_0}\left\{(1-w)\,p_0\,(\theta-\beta p_0+\phi q(p_0)+\Phi Q(p_0))-\frac{1}{2}q(p_0)^2\right\},$$

which implies

$$p(w,\theta)=\frac{\theta}{2\beta-\phi^2+\left(\phi^2-2\Phi^2\right)w}.$$

The second-order condition is $2\beta-\phi^2+\left(\phi^2-2\Phi^2\right)w>0$ and we will verify later that it holds at the optimal value of w determined below. The agent's ex-post payoff for any given θ is

$$\frac{(1-w)\,\theta^2}{2\left(2\beta-\phi^2+w\left(\phi^2-2\Phi^2\right)\right)}>0.$$

Thus, the principal's first stage optimization problem is

$$\max_{w,T}\left\{\mathbb{E}_\theta\left[wp(w,\theta)(\theta-\beta p(w,\theta)+\phi q(p(w,\theta))+\Phi Q(p(w,\theta)))-\frac{1}{2}Q(p(w,\theta))^2\right]+T\right\}$$
subject to
$$T \leq \mathbb{E}_\theta\left[(1-w)p(w,\theta)(\theta-\beta p(w,\theta)+\phi q(p(w,\theta))+\Phi Q(p(w,\theta)))-\frac{1}{2}q(p(w,\theta))^2\right],$$

which reduces to

$$\max_{w}\left\{\mathbb{E}_\theta\left[p(w,\theta)(\theta-\beta p(w,\theta)+\phi q(p(w,\theta))+\Phi Q(p(w,\theta)))-\frac{1}{2}q(p(w,\theta))^2-\frac{1}{2}Q(p(w,\theta))^2\right]\right\}$$
$$= \left(\theta^2+V_\theta\right)\max_w\left\{\frac{\left(2\beta-w^2\left(\Phi^2+\phi^2\right)-2w\left(\Phi^2-\phi^2\right)-\phi^2\right)}{2\left(2\beta-\phi^2-2w\Phi^2+w\phi^2\right)^2}\right\}.$$

The derivative in w is

$$\left(\theta^2+V_\theta\right)\frac{\left(2\beta-\phi^2\right)\Phi^2-w\left(2\beta\left(\Phi^2+\phi^2\right)+2\Phi^4-4\Phi^2\phi^2\right)}{\left(2\beta-\phi^2-2w\Phi^2+w\phi^2\right)^3}$$

and is equal to zero for

$$w^*=\frac{\Phi^2\left(2\beta-\phi^2\right)}{2\beta\left(\phi^2+\Phi^2\right)+2\Phi^4-4\Phi^2\phi^2}.$$

We need to verify three things to ensure that this w^* does indeed maximize the principal's profits: i) $0<w^*<1$, ii) the second-order condition in w for the principal's optimization problem, and iii) the second-order condition in p_0 for the agent's optimization problem.

First, note that (1) implies $2\beta>\phi^2$ and the first inequality in

$$2\beta\left(\phi^2+\Phi^2\right)>\Phi^4+2\Phi^2\phi^2+\phi^4>4\Phi^2\phi^2-2\Phi^4.$$

Thus, $w^*>0$. Furthermore, $w^*<1$ is equivalent to

$$2\beta\phi^2>3\Phi^2\phi^2-2\Phi^4.$$

Assumption (1) implies the first inequality of

$$2\beta\phi^2>\frac{\left(\phi^4+\Phi^4+\phi^2\Phi^2\right)\phi^2}{\phi^2+\Phi^2}>3\Phi^2\phi^2-2\Phi^4,$$

where the second inequality is true for all non-negative ϕ and Φ. Thus, we have proven that $0<w^*<1$ under assumption (1).

Second, we have

$$2\beta-\phi^2-2w^*\Phi^2+w^*\phi^2=\frac{\left(2\beta-\phi^2\right)\left(2\beta\left(\Phi^2+\phi^2\right)-3\Phi^2\phi^2\right)}{2\beta\left(\phi^2+\Phi^2\right)+2\Phi^4-4\Phi^2\phi^2}>0,$$

where $2\beta-\phi^2>0$, $2\beta\left(\Phi^2+\phi^2\right)-3\Phi^2\phi^2>0$ and $2\beta\left(\phi^2+\Phi^2\right)+2\Phi^4-4\Phi^2\phi^2>0$ all follow from (1). Thus, the second-order condition of the principal's optimization problem in w holds when evaluated at $w=w^*$.

Third and finally, the second-order condition of the agent's optimization problem in p evaluated at $p_0=p\left(w^*,\theta\right)$ holds for all θ if and only if

$$2\beta-\phi^2+\left(\phi^2-2\Phi^2\right)w^*>0.$$

This is equivalent to

$$\frac{\left(2\beta-\phi^2\right)\left(2\beta\left(\Phi^2+\phi^2\right)-3\Phi^2\phi^2\right)}{2\beta\left(\phi^2+\Phi^2\right)+2\Phi^4-4\Phi^2\phi^2}>0,$$

which is also implied by assumption (1).

Evaluating the principal's expected profits at w^* we obtain

$$\mathbb{E}_\theta\left[\Pi\left(\theta\right)\right]=\frac{\left(2\beta\left(\phi^2+\Phi^2\right)+\Phi^4-3\phi^2\Phi^2\right)\left(\overline{\theta}^2+V_\theta\right)}{2\left(2\beta-\phi^2\right)\left(2\beta\left(\phi^2+\Phi^2\right)-3\Phi^2\phi^2\right)}.$$

Comparing to the second-best profit given by (5) in the main paper, the difference in expected profit divided by the second-best profit is

$$\frac{\Phi^4\left(\Phi^2-\phi^2\right)^2}{\left(2\beta-\phi^2\right)\left(2\beta\left(\phi^2+\Phi^2\right)-3\phi^2\Phi^2\right)\left(\phi^2+\Phi^2\right)}.$$

This represents the principal's percentage loss relative to second-best. It is then easily verified that this expression is equal to the one provided in the text of Proposition 3.

C Alternative timing and misreporting of price

Consider the alternative specification in Section 5.3 of the main paper. Consider a contract that involves the per-unit wholesale price being equal to wp_r, based on the reported price, and a payment fp_r^2 that only depends on the reported price in stage 2 and not the number of units sold, as well as a penalty $f_p\left(p^2-p_r^2\right)$. The penalty only applies if the price is actually observed, which is assumed to happen with probability η. The principal determines the contract parameters w, f, and f_p, while η is an exogenous parameter.

The equilibrium in stage 3 is defined by

$$\begin{aligned}
p\left(\theta,w,f_p,p_r\right) &= \arg\max_p\left\{\left(p-wp_r\right)\left(\theta-\beta p+\phi q\left(\theta,w,f_p,p_r\right)+\Phi Q\left(w,p_r\right)\right)-\eta f_p\left(p^2-p_r^2\right)\right\} \\
&= \frac{\theta+\phi q\left(\theta,w,f_p,p_r\right)+\Phi Q\left(w,p_r\right)+\beta wp_r}{2\left(\beta+\eta f_p\right)} \\
q\left(\theta,w,f_p,p_r\right) &= \phi\left(p\left(\theta,w,f_p,p_r\right)-wp_r\right) \\
Q\left(w,p_r\right) &= \Phi wp_r.
\end{aligned}$$

The stage 3 equilibrium is then

$$\begin{aligned} p\left(\theta,w,f_p,p_r\right) &= \frac{\theta + wp_r\left(\beta+\Phi^2-\phi^2\right)}{2\left(\beta+\eta f_p\right)-\phi^2} \\ q\left(\theta,w,f_p,p_r\right) &= \frac{\theta\phi - wp_r\phi\left(\beta+2\eta f_p-\Phi^2\right)}{2\left(\beta+\eta f_p\right)-\phi^2} \\ Q\left(w,p_r\right) &= \Phi w p_r. \end{aligned} \tag{C.1}$$

Assuming that f is chosen to induce truthful revelation of p such that $p = p_r$ in equilibrium (which we'll show later is possible), the stage 3 equilibrium as a function of θ, w and f_p is

$$\begin{aligned} p &= p_r = \frac{\theta}{2\left(\beta+\eta f_p\right)-\phi^2-w\left(\beta+\Phi^2-\phi^2\right)} \\ q &= \frac{\theta\phi\left(1-w\right)}{2\left(\beta+\eta f_p\right)-\phi^2-w\left(\beta+\Phi^2-\phi^2\right)} \\ Q &= \frac{\theta\Phi w}{2\left(\beta+\eta f_p\right)-\phi^2-w\left(\beta+\Phi^2-\phi^2\right)}. \end{aligned}$$

Recalling that the second-best solution is fully characterized by (3)-(4), we can match the second-best solution by setting

$$\begin{aligned} w &= \frac{\Phi^2}{\phi^2+\Phi^2} \\ f_p &= \frac{\Phi^2\left(\beta-\phi^2\right)}{2\eta\left(\phi^2+\Phi^2\right)}. \end{aligned}$$

Note that at this value of f_p, the second-order condition necessary for the optimizations above to be valid holds if and only if

$$2\beta+2\eta f_p-\phi^2 = \frac{\left(2\beta-\phi^2\right)\left(\phi^2+\Phi^2\right)-\Phi^4+\Phi^2\left(\beta+\Phi^2-\phi^2\right)}{\phi^2+\Phi^2} > 0.$$

Finally, we need to ensure truthful revelation of the price by the agent in stage 2. The agent's profit from the perspective of stage 2 is

$$\begin{aligned} &\left(p\left(\theta,w,f_p,p_r\right)-wp_r\right)\left(\theta-\beta p\left(\theta,w,f_p,p_r\right)+\phi q\left(\theta,w,f_p,p_r\right)+\Phi Q\left(w,p_r\right)\right) \\ &-\frac{1}{2}q\left(\theta,w,f_p,p_r\right)^2-\eta f_p p\left(\theta,w,f_p,p_r\right)^2-\left(f-\eta f_p\right)p_r^2. \end{aligned}$$

Using (C.1), the level of p_r that maximizes these profits is

$$p_r\left(\theta,w,f_p,f\right) = \frac{\theta w\left(2\eta f_p+\beta-\Phi^2\right)}{w^2\left(\left(\beta-\Phi^2\right)^2-2\eta f_p\left(2\Phi^2-\phi^2\right)\right)-2\left(f-\eta f_p\right)\left(2\beta+2\eta f_p-\phi^2\right)}.$$

To induce truthful revelation, we need to have

$$p\left(\theta, w, f_p, p_r\right) = p_r\left(\theta, w, f_p, f\right)$$

for all θ when $w = \frac{\Phi^2}{\phi^2+\Phi^2}$ and $f_p = \frac{\Phi^2\left(\beta-\phi^2\right)}{2\eta\left(\phi^2+\Phi^2\right)}$. This can be achieved by setting

$$f = \frac{\phi^2\Phi^2\left(\Phi^2-\phi^2\right)}{2\left(\Phi^2+\phi^2\right)^2},$$

which is identical to the level of f in (8), which corresponds to the optimal ad valorem contract in the benchmark setting. Taking into account that $p_r > 0$ and $w > 0$ at the proposed equilibrium, the second-order condition corresponding to the last optimization problem (in p_r) is

$$\beta - \Phi^2 + 2\eta f_p = \frac{\beta\left(2\Phi^2+\phi^2\right) - \Phi^2\left(\Phi^2+2\phi^2\right)}{\phi^2+\Phi^2} < 0.$$

Thus, the solution we have determined above is valid provided

$$\begin{aligned}
\left(2\beta-\phi^2\right)\left(\phi^2+\Phi^2\right) - \Phi^4 + \Phi^2\left(\beta+\Phi^2-\phi^2\right) &> 0 \\
\Phi^2\left(\Phi^2+2\phi^2\right) - \beta\left(2\Phi^2+\phi^2\right) &> 0.
\end{aligned}$$

Since there is no penalty in equilibrium (recall $p = p_r$ with this contract), and since p, q and Q are all set at the second-best levels and the upfront fee T is set according to (9) so as to extract the agent's expected profit, the principal will achieve expected profits as defined in (5), consistent with the second-best outcome.