Sharing revenue and control*

Andrei Hagiu[†] and Julian Wright[‡]

September, 2016

Abstract

We study how much revenue a principal (e.g. a manufacturer) should share with an agent (e.g. a retailer) and how much control it should grant the agent over costly decisions that it can monitor (e.g. the retailer's level of local advertising). These two contracting choices are tightly linked: giving the agent more control over costly decisions goes hand-in-hand with leaving the agent with a higher share of revenue. We study the full range of delegation possibilities facing the principal, and explain why granting the agent control over costly decisions, subject to minimum requirements, is often the best option. Our analysis applies to the contracting choices facing franchisors, manufacturers, shopping malls, online platforms, and movie studios, among other examples. When applied to pricing decisions, it provides a new theory of resale price maintenance, which explains when price ceilings or price floors should be used.

JEL classification: D4, L1, L5

Keywords: partial delegation, revenue sharing, resale price maintenance, platform governance.

1 Introduction

When contracting with franchisees, a business-format franchisor such as a hotel chain, a fast-food restaurant or a car rental company has to decide how much revenue to share with its franchisees and how much to control their advertising choices. The franchisor could decide the level of local advertising itself and write it into the franchise contract, delegate the choice entirely to the franchisees, or let the franchisees decide subject to a minimum advertising expense—we call the latter partial delegation. Similarly, when contracting with a movie theater, a movie studio has to decide whether to control the number of weeks a particular movie will be shown at the theater and how to share the resulting

*Julian Wright gratefully acknowledges research funding from the Singapore Ministry of Education Academic Research Fund Tier 1 Grant No. R122000215-112.

†MIT Sloan School of Management, Boston, MA 02142, E-mail: ahagiu@mit.edu

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

revenue with the theater. The studio could stipulate a particular number of weeks, delegate the choice entirely to the movie theater, or let the movie theater decide subject to a minimum number of weeks.

Similar choices arise in a wide range of other principal-agent settings (e.g. manufacturers and their sales agents, platforms and participating third-party providers of services or applications). We build a theoretical model to evaluate the principal's optimal levels of revenue-sharing and control. In particular, we compare partial delegation—in which the principal retains some control—with both full control and full delegation. We derive conditions under which imposing a minimum threshold for the agent (i.e. minimum requirements) is the optimal form of partial delegation. And we determine when using threshold delegation does better than either full control or full delegation.

In our model, the revenue generated jointly by the principal and the agent depends on three costly decisions. The first is a partially contractible and transferable decision, for which the agent has private information (e.g. how effective is local advertising for a franchisee). The other two are non-contractible and non-transferable ongoing investment decisions, one always undertaken by the principal (e.g. the franchisor's investment in national advertising) and one always undertaken by the agent (e.g. the franchisee manager's effort)—they create a double-sided moral hazard problem. We assume the principal can make use of a two-part tariff—a fixed fee (or payment) and a revenue share—in its contract. To incentivize both the principal and the agent to continue investing in their respective non-transferable activities, the principal will want each party to retain a positive share of revenue. We use this model to compare the principal's expected profit from three types of contract: (i) the principal fixes the level of the transferable decision variable in its contract (which we call the P-mode), (ii) the agent is granted full control over the transferable decision variable (which we call the A-mode), and (iii) a hybrid mode in which the principal restricts the agent's choice of the transferable decision variable according to some rule that still provides the agent with some discretion (which we call the H-mode).

A consequence of revenue sharing is that the agent's choice of the transferable decision will be distorted downwards (e.g. franchisees will not invest as much in local advertising as would be optimal). The smaller the revenue share left to the agent, the larger the magnitude of this distortion (the agent's bias). This consideration suggests that the principal may do best in P-mode, in which it controls the choice of the transferable decision variable by fixing it contractually, thereby avoiding the downward distortion. This provides the principal with a commitment benefit and reduces the need to leave the agent with a high share of revenue. On the other hand, the agent may have better information about the optimal level of the transferable decision variable (e.g. how effective is local advertising, or private revenues from cross-selling other products or services). This consideration suggests that the principal may do best in A-mode, in which it delegates the transferable decision to the agent, but leaves the agent with a higher share of revenue.

Rather than fully controlling (P-mode) or fully delegating (A-mode), we show that the principal often does best using partial delegation (H-mode), which involves the agent's choice of the transferable decision being restricted to be above some minimum threshold. Threshold delegation is a way to get some of the advantages of each of the pure modes: the commitment benefit of the P-mode and the responsiveness to the agent's information that is enabled by the A-mode.

At a high level, our article studies the interaction between two strategic instruments: the sharing of revenues between principal and agent, and the allocation of control over transferable decision variables between them. Both are decided by the principal as part of its contract choice. Revenue sharing endogenously determines the magnitude of the agent's bias (distortion) in choosing the transferable decision variables. In turn, the bias determines the extent to which the principal wishes to delegate control to the agent. We show that this interdependence makes the two instruments strategic complements, i.e. giving the agent more control over transferable decisions (by imposing minimum requirements rather than specifying the exact level of the transferable decision, or by lowering existing minimum requirements, or by removing these requirements altogether) goes hand-in-hand with leaving the agent with a higher share of revenue.

The strategic complementarity implies that the share of revenue the principal should retain is highest in P-mode, lowest in A-mode, and intermediate in H-mode. This in turn implies that when the importance of the principal's moral hazard increases, the P-mode becomes more desirable relative to the H-mode, which becomes more desirable relative to the A-mode. The converse is true when the importance of the agent's moral hazard increases. Further comparative static results can be obtained when we assume private information shocks are drawn from the uniform distribution. With some additional mild assumptions, we show that when the variance of these shocks increases, the principal should delegate more—the A-mode becomes more desirable relative to the H-mode, which in turn becomes more desirable relative to the P-mode.

Our theory of partial delegation can also be applied to provide a new explanation for the use of resale price maintenance (RPM) contracts. To do so, we adapt our model by making price the transferable decision variable. The two non-transferable investment decisions remain unchanged and continue to create double-sided moral hazard. In this context, the principal may be an upstream firm (e.g. a manufacturer) and the agent a downstream firm (e.g. a retailer), although our model of RPM applies much more generally than this (e.g. platforms and third-party application developers, or franchisors and franchisees). We show that when the agent's moral hazard problem is sufficiently important, the principal should adopt the H-mode and set a minimum price, designed to mitigate the agent's moral hazard problem, which is due to the fact that the agent only obtains a fraction of the revenue generated by its investments. On the other hand, when the agent's moral hazard problem is not very important, the principal should set a maximum price, designed to mitigate the double marginalization problem due to the agent not internalizing the revenue retained by the principal. Thus, our theory explains when either minimum or maximum RPM is used, and why manufacturers' contracts that involve RPM oftentimes involve price floors or price ceilings rather than exact price levels.

Finally, we extend our theory to the case in which the principal is a platform and there are network effects across agents, i.e. revenues per agent increase when more agents participate on the platform. We find that principals should use more restrictions on agents (or even switch to full control) when agents' expectations are pessimistic rather than optimistic. And vice versa. This reflects that the principal can weaken the impact of network effects by reducing the share of revenue it leaves to agents, which, as discussed above, goes hand-in-hand with the principal taking more control.

2 Literature review

Our paper combines elements from the literatures on organizational design, agency vs. wholesale pricing, partial delegation, retail channel coordination and double-sided moral hazard.

The choice between P-mode and A-mode in our model is reminiscent of the choice studied by Simon (1951) between contracting on a decision ex-ante (before uncertainty is resolved) vs. giving full authority to the employer (principal) or the employee (agent) to unilaterally choose the decision ex-post. In our model, giving full authority to the principal to unilaterally make the transferable decision ex-post is never optimal. This is because the principal never observes the realization of the agent's private information and can always extract the entire surplus from the agent through its two-part tariffs—this means the principal can always do better by committing to the choice of the transferable action ex-ante.

A related and more recent strand of literature has emerged that studies conditions under which retailers/platforms take control over transferable decisions pertaining to the sale of products to end-consumers or allow their suppliers/complementors to keep control over these decisions. Most of this literature (Gans 2012, Foros et al. 2013, Abhishek et al. 2015, Johnson 2013 and 2014) focuses on price as the main decision that can be controlled by the retailers (wholesale model) or by the suppliers (agency model). Exceptions include Desiraju and Moorthy (1997), in which a supplier decides whether to give a retailer control over price and a costly service decision, price only, service only, or neither; Jerath and Zhang (2010), in which control over price is tied to control over a costly service decision; and Hagiu and Wright (2015, 2016), who study how firms allocate control rights over multiple non-contractible, non-price decisions between them and their agents.

The key novelty that we introduce relative to the articles above is that we allow for an intermediate option between fixing the transferable decision in the principal's ex-ante contract and giving full authority to the agent: the agent can be given authority to choose the transferable action subject to restrictions imposed by the principal's ex-ante contract. This is known as "partial delegation" following the seminal work by Holmstrom (1977, 1984). Several papers have proven that threshold delegation is optimal in similar settings—see for example, Melumad and Shibano (1991), Martimort and Semenov (2006), Alonso and Matouschek (2008), and Amador and Bagwell (2013). We directly show that threshold delegation is optimal in our setting under a relatively weak condition on the distribution of private information, which is closely related to the condition in Martimort and Semenov (2006). There are three key contributions in our model relative to the partial delegation literature: (1) we allow for monetary transfers between the principal and the agent in the form of two-part tariffs set by the principal in the contracting stage; (2) we introduce double-sided moral hazard; and (3) the bias of the agent's objective function relative to the principal's is endogenously determined by the two-part tariff, which in turn depends on the importance of the agent's moral hazard relative to the principal's, and on the importance of the agent's private information. By contrast, the partial delegation literature to date assumes an exogenously given bias, no transfers between principal and agent and no moral hazard for either the principal or the agent.

In retail contexts, partial delegation can be viewed as an additional instrument that can help improve "channel coordination". Our modelling approach is entirely consistent with the principal-agent view of channel coordination taken by the marketing and management literature to date (see Lal, 1990, Gal-Or, 1995, Cachon and Lariviere, 2005, and Foros et al., 2009). However, this literature has focused on improving channel coordination through various payment instruments (revenue sharing, wholesale pricing, quantity discounts, buy-backs) and/or through monitoring, which is modelled as enforcing a specific level of a non-contractible investment in service. We extend this work by showing that the addition of threshold delegation with respect to transferable decision variables provides a more nuanced theory of how channel coordination can be improved, which is also consistent with business practice, given that threshold delegation is commonly observed (see Section 3 below). In particular, the principal's need to retain a positive share of revenues (due to its own moral hazard problem) endogenously determines the agent's bias, i.e. the need for channel coordination, which in turn can be mitigated by using a more restrictive delegation threshold.

Since in our model revenues must be shared between the principal and the agent to incentivize both sides to make non-contractible investments, we also directly build upon principal-agent models with double-sided moral hazard (Lal, 1990, Romano, 1994, and Bhattacharyya and Lafontaine, 1995). The key difference relative to these papers is that our model can explain partial delegation of a transferable action, which is frequently observed in practice. Indeed, the transferable action is entirely absent from the models of Lal (1990) and Bhattacharyya and Lafontaine (1995), and is deterministic in Romano (1994), which means there is no scope for partial delegation. Furthermore, Romano (1994) focuses on price as the transferable action, whereas we cover both the case in which the transferable action is a costly investment and the case in which it is price.

In contrast to most existing theories of resale price maintenance (RPM) such as Deneckere et al. (1996), Jullien and Rey (2007) and Asker and Bar-Isaac (2014), our model can explain the use of both minimum and maximum RPM, and when each would be used. Moreover, the use of price ceilings or floors in our theory of RPM is distinct from theories which predict that the manufacturer uses RPM to fix a particular level of price, e.g. Romano (1994).

Given that our model is applicable to many industry settings featuring platforms, we also contribute to the emerging literature on platform governance rules, i.e. non-price mechanisms employed by platform owners to regulate the access and behavior of platform participants (Boudreau and Hagiu, 2009). Specifically, the minimum requirements placed by the principal on the agent's choice of investment can be viewed as a form of governance whenever the principal is a platform provider (e.g. Uber and Lyft as discussed in Section 3). While several studies have documented various forms of platform governance (e.g. Gawer and Cusumano, 2002, Boudreau and Hagiu, 2009, Boudreau, 2010 and 2012, Hagiu, 2014), the only theoretical models of such non-price governance rules that we are aware of are Casadesus-Massanell and Halaburda (2014) and Parker and Van Alstyne (2014). However, these two papers tackle very different issues than the one we study here: Casadesus-Massanell and Halaburda show that a platform may benefit from restricting the number of applications available on it as a coordination mechanism, while Parker and Van Alstyne study a platform's optimal degree of openness and IP protection for third-party developers.

3 Examples

There are a wide variety of examples that our theory can be applied to. Table 1 summarizes a few key examples, listing decisions that are transferable and potentially subject to restrictions (minimum or maximum requirements), as well as decisions that are non-transferable and subject to moral hazard, and the source of the agent's private information.

Table 1: Examples

	Transferable decisions (possibly subject to restrictions)	*Non-transferable investment decisions made by the agent*	*Non-transferable investment decisions made by the principal*	*Source of agent's private information*
Franchising	local advertising of the outlet; price	outlet manager's effort	national advertising of the brand	effectiveness of local advertising; revenue from cross-selling other products; franchisee's costs
Manufacturer and authorized dealer	investment in quality of outlet; local promotion and advertising	dealer's effort	quality and marketing of the product	local demand; revenue from cross-selling other products or services; dealer's costs
Shopping malls	retailer's opening hours	quality and advertising of retail store	maintenance and advertising of mall	retailer's demand, costs and outside revenues that originate from mall traffic
App stores	app licensing terms; price of app	advertising and upkeep of app	technological upkeep (e.g. payment) and advertising of app store	revenues and cross-selling opportunities outside of app
Transportation platforms	quality of car	customer service	technological upkeep (e.g. payment, dispatch) and advertising of service	repeat business for the driver off the platform; cash tips
Movie studios and theaters	run length	quality and maintenance of the theater's facilities	advertising and promotion of the movie	theater's concession revenue and opportunity costs; local demand for movies;
Manufacturer and retailer	price	sales effort and in-store promotion	quality and marketing of the product	local demand; revenue from cross-selling other products

Consider first "business format" franchising (e.g. hotel, fast-food, car rental, etc.). The franchisor is the principal in our setting, while the franchisee is the agent. One can often distinguish national from local advertising, with the latter being a decision that could be made either by the franchisor or the franchisee. For local advertising, contracts often specify a minimum spending requirement by the franchisee, consistent with our H-mode. Other decisions that are typically chosen by franchisees subject to minimum requirements imposed by franchisors include the number of staff that have to be on-site at various days/times, cleanliness, and opening hours. Consistent with our model, most business format franchisors use a combination of upfront franchise fees (paid at the beginning of the contract period) and sales-based royalties (fixed royalty rate). Blair and Lafontaine (2010, p. 250-253) document that fixed franchisee fees typically range between \$5,000 and \$30,000, while royalty rates (collected by the franchisor) are usually between 3-6% of sales. Franchisees clearly have private information, as indicated in Table 1.

Branded manufacturers that distribute their products through authorized dealers provide another large set of related examples. For instance, manufacturers oftentimes impose minimum standards for retail premises and minimum advertising or promotion levels by the retailers, but these same transferable decision variables can also sometimes be stipulated by the manufacturer or left unrestricted. Likewise, a manufacturer that uses sales representatives would face a similar situation—the extent to which it controls their transferable decisions (and therefore the extent to which sales representatives would be considered employees or independent contractors). Consistent with our model, as Anderson (1985) documents, compensation contracts between manufacturers and their dealers or sales representatives typically include revenue-sharing (for an employee, a bonus tied to sales revenue) and a fixed payment (for an employee, a fixed salary component). Dealers and sales representatives have similar types of private information as franchisees.

Our theory also applies to an increasing number of platforms, both offline and online. Consider three examples: shopping malls, digital app stores (e.g. Apple's App Store for iPhone apps and Google's Play Store for Android apps) and ride-hailing apps (e.g. Lyft and Uber). First, all three types of platforms place minimum requirements on important transferable decisions. In the case of shopping malls, the lease agreements often specify minimum opening hours for retailers (those hours could be set by the mall or by each respective retailer). Apple and Google place minimum requirements on the terms of the licensing agreement provided by app developers to their users.[1] For instance, both Apple and Google require developers to assume sole responsibility for any defects or performance issues related to their apps and Google requires developers to respond to customer support inquiries within three business days. UberX and Lyft drivers have to use cars that satisfy a minimum age requirement (e.g. 2001 or newer in many cities for UberX, and 2004 or newer in many cities for Lyft). The two companies also impose minimum requirements on the cars' functionality (e.g. 4 doors, at least 5 seat belts) and on their state of maintenance (e.g. fully functioning A/C and heating, no major cosmetic damage). In contrast, traditional taxi companies can be viewed as functioning in our P-mode, since they completely control and incur the costs corresponding to the choice of cars used by their drivers. Second, the fees charged by the platforms in two of the three examples are consistent with our model. Shopping malls lease agreements typically specify both a fixed fee (monthly rent) and a small revenue share to be paid by retailers to the mall owner. Apple and Google charge application developers fixed fees ($99/year for a developer account in Apple's App Store; $25 one-time fee for a developer account in Google's Play Store) along with a 30% share of revenues generated by the applications. Lyft and Uber do not currently charge drivers any fixed fees, but keep roughly 20% of the revenues generated by each driver. Finally, in all three platform examples agents have private information, as indicated in Table 1.

A further application of our theory is to the movie industry. A movie studio is the principal in our setting, with the movie theater the studio's agent for the "distribution" of its movies to consumers.

[1] See for instance http://www.apple.com/legal/internet-services/itunes/appstore/dev/minterms/

One important transferable variable is the run length of the movie, i.e. the number of weeks it will be shown. The studio could fix the number of weeks in its contract (P-mode), leave it unrestricted (A-mode), but most often it imposes a minimum run length on the theater (H-mode).[2] The contracts between the studio and the theater are broadly consistent with our theory, usually specifying a fixed fee (the "house nut") along with a percentage share of revenue.[3] A movie theater's private information includes the significant additional revenue the theater obtains from concession sales (popcorn, candy, soda), information on local demand, and the theater's opportunity costs of continuing to show a given movie instead of others.

Finally, another key application of our theory is to the case of resale price maintenance (RPM). In RPM contracts, the principal (e.g. a manufacturer, producer or franchisor) exercises control over the price set by the agent (e.g. a retailer, distributor or franchisee). A classic case is Albrecht v. Herald Co. (1968).[4] The (local) newspaper owner (the Herald) granted the distributor and delivery agent (Albrecht) an exclusive territory for selling the Herald's newspaper, the St. Louis Globe-Democrat. Rather than Herald requiring Albrecht to set a particular price, or leaving Albrecht free to set the price, Herald imposed a price ceiling below which Albrecht could choose any price it wanted. The Herald only removed this maximum RPM contract due to a 1968 U.S. Supreme Court decision, which ruled that fixing a maximum price was illegal. In other countries, where maximum RPM is explicitly allowed (e.g. India), retailers routinely set prices below the price ceilings imposed by manufacturers, consistent with our theory.[5] With the shift away from per-se illegality of price fixing in the U.S., business-format franchisors and manufacturers will increasingly face the issue of whether and how to restrict the prices set by their franchisees. The issue of who sets prices is also highly relevant for modern platforms, which make use of all three modes featured in our theory. Most often, developers or suppliers have complete freedom to set prices (e.g. sellers on eBay), but sometimes restrictions are imposed on them (e.g. Apple's App Store or Google's Play Store[6]), or the platform may even set the prices itself (e.g. Uber and Lyft).

4 Model set-up

We assume the demand $R(a, q, Q)$ generated by a principal and an agent is determined by the choice of three decision variables: (i) a is a costly, transferable and partially contractible action[7] chosen either by the principal or the agent; (ii) q is a costly, non-transferable and non-contractible action always chosen by the agent (e.g. effort); and (iii) Q is a costly, non-transferable and non-contractible action always chosen by the principal (e.g. on-going investments). For simplicity, we assume demand is linear

[2] M. Fahey, "Why movies are sometimes here and gone in theaters," CNBC.com, November 17, 2015. http://www.cnbc.com/2015/11/17/why-movies-are-sometimes-here-and-gone-in-theaters.html

[3] See https://en.wikipedia.org/wiki/Film_distributor

[4] See https://supreme.justia.com/cases/federal/us/390/145/case.html

[5] See http://www.fullstopindia.com/mrp-maximum-retail-price/ and https://en.wikipedia.org/wiki/Maximum_retail_price

[6] App Store developers cannot charge more than $999.99, while Play Store app prices are capped at $400 in the U.S. The vast majority of apps sell below these price ceilings, but some do price at the maximum allowed level.

[7] The action a is partially contractible in the sense that the principal can restrict the agents' choice of a but it cannot make payments conditional on it. We discuss this assumption below.

in these variables, and so can be written as

$$R(a, q, Q) = \beta a + \phi q + \Phi Q, \tag{1}$$

where β, ϕ and Φ are positive constants, measuring the impact of a, q and Q, respectively, on demand. The fixed costs of the respective actions are $\frac{1}{2}a^2$, $\frac{1}{2}q^2$ and $\frac{1}{2}Q^2$.

We assume that the price is exogenously determined and normalized to one. This implies demand is equal to revenue, and henceforth we will refer to R as revenue. This assumption may reflect that the price is determined by market norms or regulations. For example, in the movie example, movie producer and theaters do not typically set the price of each particular movie, which is pinned down by the standardized price of movies in the theater in question. These prices are remarkably uniform across titles and over time, as documented by Einav and Orbach (2007). Similarly, a chain store will typically maintain the same prices across multiple outlets, meaning prices may not depend significantly on the contract offered by a particular shopping mall or the demand in that mall. Alternatively, we can allow the principal to set the price in its contract. Specifically, we can write revenue as $p(\theta - p + R(a, q, Q))$ where p is the price and $\theta - p + R(a, q, Q)$ is the level of demand. In an online appendix we show that the comparisons between the different modes considered in the paper remain unchanged with this specification. The normalization of $p = \theta = 1$ is therefore primarily done for expositional convenience. The case in which p is a transferable decision variable that can be chosen by either party is considered separately in Section 6.1.

In order to model the agent's private information, we assume that the agent also derives a private benefit $bR(a, q, Q)$, where only the agent observes b. This private benefit is assumed proportional to the underlying demand (equivalently, revenue). A private benefit reflects that when the agent has a higher level of sales through its contract with the principal, the agent may also have increased opportunities to sell complementary services or products to the same customers, but these opportunities fall outside the scope of the contractual relationship with the principal. For example, movie theaters can sell more popcorn and soda when a movie attracts more viewers, but do not share the resulting revenues with the movie producer. A hotel franchisee may enjoy more private revenue through associated restaurants and tour sales when there are more guests staying. Agents may also obtain reputational benefits that are proportional to sales (e.g. a franchisee manager's résumé will be more impressive if the manager has handled more sales). It is natural that the agent has private information about the value of such benefits.

Formally, we assume b is a random variable drawn from the distribution function G, with positive density $g(.)$ over $[b_L, b_H]$, with finite mean $\mathbb{E}(b) = \bar{b} > 0$ and variance V_b. We assume $0 \leq b_L < b_H$, but do not require that b_H is finite, so we allow for distributions which have unbounded support on the right tail (such as the exponential or normal distributions). The distribution function $G(.)$ is twice continuously differentiable.

Although we focus on private benefits, the logic of our analysis also extends to other sources of private information. One could interpret private benefits as instead being the agent's private costs associated with handling a higher level of demand, so total revenue would become $(1 - b)(\beta a + \phi q + \Phi Q)$.

An analytical problem with uncertain private costs is that the agent's ex-post payoff may turn out to be negative for high draws of the private cost, in which case ex-post the agent would not want to produce. This creates an additional constraint which renders the analysis intractable. Another source of the agent's private information might be the effectiveness of the choice of the transferable decision variable, i.e. the parameter β. For instance, in the franchising example, the effectiveness of local advertising may be uncertain, and something about which the franchisee has better information than the franchisor. In the online appendix we explore this alternative specification, with uncertainty over β and no private benefits ($b = 0$), and show that most of our results continue to hold. In Section 6.1 we analyze a modified version of our model in which the transferable action is a price instead of the costly action and in which the agent has private information with respect to a demand shock.

The total revenue net of fixed costs generated by the principal and the agent is therefore

$$(1+b)\, R\,(a,q,Q) - \frac{1}{2}a^2 - \frac{1}{2}q^2 - \frac{1}{2}Q^2 = (1+b)\,(\beta a + \phi q + \Phi Q) - \frac{1}{2}a^2 - \frac{1}{2}q^2 - \frac{1}{2}Q^2.$$

However, the principal can only contract on the baseline revenue $R\,(a,q,Q)$ given that the realization of private benefits is not observed by the principal. Specifically, we assume that the principal makes a take-it-or-leave-it offer to the agent, which is in the form of a revenue sharing contract with a fixed participation fee. According to this contract, the principal retains the share of revenue $tR\,(a,q,Q)$, with the agent retaining the remaining share $(1-t)\, R\,(a,q,Q)$, where $0 \leq t \leq 1$. The fixed fee T is set to leave the agent indifferent between participating or not. In practice, this could be negative (the agent may be paid a fixed salary to participate) but without loss of generality we normalize the agent's outside option to zero, which means the T arising in our model will always be positive. Such two-part tariffs are widespread and prevail in all of the examples discussed in Section 3.[8] Since the price is exogenously fixed, in our set-up revenue-sharing contracts (t, T) are equivalent to two-part tariffs in which the variable charge is a wholesale price.

In addition to the two-part tariff (t, T), the contract offered by the principal also specifies the level of control over the transferable action a. If the contract directly specifies the choice of a, we say that the principal has chosen the P-mode to reflect that only the principal determines a. The cost $\frac{1}{2}a^2$ of choosing a is then incurred by the principal. If the contract leaves the agent free to set a with no restrictions, we say that the principal has chosen the A-mode to reflect that only the agent determines a. In this case, the cost of choosing a is incurred by the agent. If instead the contract partially restricts the agent's choice of a, we say that the principal has chosen the H-mode, which is a hybrid of the two pure modes in that both the principal and agent determine a. Here too, the cost of choosing a is incurred by the agent.

The timing of the players' moves is as follows: In the first stage, the principal offers its contract (which includes the level of a in P-mode[9]) and the agent decides whether or not to accept the contract.

[8] Note, however, such two-part tariffs are not necessarily optimal due to the presence of uncertainty. For instance, Foros et al. (2009) show that in some situations the principal might do better with a non-constant revenue share.

[9] In our setup, the fourth logically possible delegation option—in which the principal maintains control over a but only chooses it in the second stage instead of fixing it in its contract—is always dominated. This is because the principal never observes any private information, so there is no benefit in waiting rather than committing to the choice of a ex-ante. On

In the second stage, the agent learns b; then actions a (if in A-mode or H-mode), q and Q are chosen and the corresponding costs are incurred. Finally, payoffs are realized.

Several remarks are in order. First, the agent only learns b after having decided whether or not to accept the contract from the principal. This assumption is natural if the task/service is principal-specific. It is also made for analytical tractability: if the agent knew b at the time he was deciding whether or not to accept the principal's contract, that would create a downward-sloping demand of agent participation for the principal, which would lead to intractable calculations.

Second, in A-mode and H-mode, we have assumed, as is realistic, that the cost of a (i.e. $\frac{1}{2}a^2$) is incurred by the agent in the second stage. In P-mode, whether the cost of a is incurred by the principal or the agent is actually immaterial to the outcome because a is set contractually and the principal can use a two-part tariff to extract the agent's entire expected surplus in excess of a fixed outside option. Thus, our assumption that the principal incurs the cost of a in P-mode is made without any loss of generality. This feature of the model is appealing because in real-world settings, the cost of decisions fixed by contract in P-mode is oftentimes borne by the agent (e.g. local advertising by franchisees) and sometimes by the principal (e.g. choice of cars by taxi companies). The equivalence breaks down if there are positive monitoring costs that the principal must incur in order to make sure the agent complies with the contract when the agent is the one who actually chooses a and incurs the associated costs in P-mode—we explore this issue in Section 5.4.

Third, as is standard in the delegation literature, we assume the principal cannot commit to contracts that make payments conditional on a. Otherwise, the principal would do best by keeping the share of revenues that optimally balances double-sided moral hazard, i.e. $t = \frac{\Phi^2(1+\bar{b})}{\Phi^2+\phi^2}$, but reimburse the agent β for every unit of a he chooses. In this case, the principal would ensure that the agent chooses the first-best level of a, i.e. $a = \beta(1+b)$. This option would dominate both the pure modes and the hybrid mode given our linear specification. However, in the various examples discussed in Section 3, payments based on a are generally not used. This may reflect that principals find it more economical to simply restrict the agent's discretion than to determine and enforce the optimal complete contract (see also Alonso and Matouschek, 2008 for several different justifications for the assumption).

In this model, the bias between the principal's and agent's preferred choice of a arises endogenously due to the revenue sharing contract. The share $(1-t)$ of revenue retained by the agent will typically be less than 100% (because the principal needs to be incentivized to invest in Q), so whenever the agent chooses a, his choice will be biased downward. If we compare the agent's choice of a and the first-best choice that the principal wants to induce with its contract, these are in general $a = (1-t+b)\beta$ and $a = (1+b)\beta$ respectively, so the difference (the bias of the agent) does not depend on the agent's private benefits b at all. This is because the principal can extract the total expected value of private benefits through the fixed fee. Thus, private benefits do not inherently create a misalignment of objectives between principal and agent, given that the principal can use a fixed fee. Note that when $t = 0$ (which is optimal if $\Phi = 0$, so ongoing investment on the part of the principal is not required), there is no bias, and as t increases the bias endogenously increases.

the other hand, committing to a in the contract avoids the additional moral hazard in a that would arise if $t < 1$.

5 Results

We now analyze this benchmark model, first considering whether the principal prefers to set the level of a in its contract (P-mode) or entirely delegate that choice to the agent (A-mode), before considering whether the principal instead prefers some form of partial delegation (H-mode).

5.1 Comparison between pure modes

Consider first the P-mode. The principal solves

$$\begin{aligned} & \max_{t,T,a} \left\{ \mathbb{E}\left[t\left(\beta a+\phi q+\Phi Q\right)\right]+T-\frac{1}{2}a^2-\frac{1}{2}Q^2 \right\} \\ & \text{subject to} \\ T \leq \; & \mathbb{E}\left[\left(1-t+b\right)\left(\beta a+\phi q+\Phi Q\right)-\frac{1}{2}q^2\right] \\ q = \; & \left(1-t+b\right)\phi \\ Q = \; & t\Phi. \end{aligned}$$

The first constraint is the agent's participation constraint—recall the agent must make the participation decision prior to learning his private benefit b. The second and third constraints reflect that the agent and the principal's incentives to invest in q and Q respectively are driven by the share of variable revenues that each extracts: $(1-t)$ for the agent and t for the principal.

At the optimum, the first constraint is also binding (the principal uses the fixed fee T to extract the entire net expected payoff from the agent), so the principal's program can be re-written

$$\begin{aligned} & \max_{t,a} \mathbb{E}\left[\left(1+b\right)\left(\beta a+\phi q+\Phi Q\right)-\frac{1}{2}a^2-\frac{1}{2}q^2-\frac{1}{2}Q^2\right] \\ & \text{subject to} \\ q = \; & \left(1-t+b\right)\phi \\ Q = \; & t\Phi. \end{aligned}$$

Optimizing over a and plugging in the two constraints, the principal solves $\max_t \Pi^P(t)$, where

$$\Pi^P(t) \equiv \frac{\beta^2}{2}\left(1+\overline{b}\right)^2+\frac{\phi^2}{2}\left(\left(1+\overline{b}\right)^2+V_b-t^2\right)+\frac{\Phi^2}{2}t\left(2\left(1+\overline{b}\right)-t\right). \tag{2}$$

In this expression, the last two terms are the expected profit generated from the choices of (q,Q) for any $t\in[0,1]$. Eliminating the agent's moral hazard (i.e. providing first-best incentives for q) requires $t=0$, whereas eliminating the principal's moral hazard (i.e. providing first best incentives for Q) requires $t=1$. The revenue share t does not create any loss associated with the investment in the transferable action a because it is set contractually:

$$a^{P*}=\beta\left(1+\overline{b}\right).$$

Optimizing (2) over t, we obtain

$$t^{P*} = \frac{\Phi^2 \left(1 + \bar{b}\right)}{\Phi^2 + \phi^2}.$$

We assume

$$\bar{b} < \frac{\phi^2}{\Phi^2},$$

which ensures $t^{P*} < 1$. The resulting optimal profit for the principal in P-mode is

$$\Pi^{P*} = \frac{1}{2}\left(\left(\beta^2 + \phi^2\right)\left(1 + \bar{b}\right)^2 + \phi^2 V_b + \frac{\Phi^4 \left(1 + \bar{b}\right)^2}{\Phi^2 + \phi^2}\right). \tag{3}$$

Consider now the A-mode. Again, the fixed fee (or salary) T is set such that the principal extracts the entire expected payoff in excess of the agent's outside option. Thus, the principal solves

$$\begin{aligned}
& \max_t \left\{\mathbb{E}\left[(1+b)\left(\beta a + \phi q + \Phi Q\right) - \frac{1}{2}a^2 - \frac{1}{2}q^2 - \frac{1}{2}Q^2\right]\right\} \\
& \text{subject to} \\
a &= (1 - t + b)\,\beta \\
q &= (1 - t + b)\,\phi \\
Q &= t\Phi.
\end{aligned}$$

The principal's expected profit as a function of t is then

$$\Pi^A(t) = \frac{\beta^2}{2}\left(\left(1+\bar{b}\right)^2 + V_b - t^2\right) + \frac{\phi^2}{2}\left(\left(1+\bar{b}\right)^2 + V_b - t^2\right) + \frac{\Phi^2}{2} t\left(2\left(1+\bar{b}\right) - t\right). \tag{4}$$

The last two terms in (4) once again capture the profit generated from the choices of (q, Q) for any $t \in [0, 1]$. Relative to the P-mode, here there is a further reduction in the principal's profit when $t > 0$, because that leads to sub-optimal incentives to invest in a by the agent. This is reflected by the term $-\frac{1}{2}\beta^2 t^2$ in (4). Thus, the optimal revenue share in A-mode needs to balance not just the double-sided moral hazard in (q, Q) as in the P-mode, but also the agent's incentives to invest in a. From this point of view, the A-mode is inferior to the P-mode, where a was set contractually and thus not subject to an incentive problem. On the other hand, the advantage of the A-mode is that the agent can incorporate private information about b in the choice of a. This can be seen by the fact V_b multiplies just ϕ^2 in (3) whereas it multiplies both ϕ^2 and β^2 in (4).

The optimal share of revenue extracted by the principal in A-mode is

$$t^{A*} = \frac{\Phi^2 \left(1 + \bar{b}\right)}{\Phi^2 + \phi^2 + \beta^2}.$$

Note that $0 < t^{A*} < t^{P*} < 1$.

The principal's optimal A-mode profit is then

$$\Pi^{A*} = \frac{1}{2}\left(\left(\beta^2+\phi^2\right)\left(\left(1+\bar{b}\right)^2+V_b\right)+\frac{\Phi^4\left(1+\bar{b}\right)^2}{\Phi^2+\phi^2+\beta^2}\right). \tag{5}$$

Comparing Π^{A*} with Π^{P*}, we obtain the following proposition.

Proposition 1 *The principal's profit is higher in A-mode compared to P-mode if and only if the variance of the agent's private benefit is sufficiently large; i.e.*

$$V_b > \frac{\Phi^4\left(1+\bar{b}\right)^2}{\left(\Phi^2+\phi^2\right)\left(\Phi^2+\phi^2+\beta^2\right)}. \tag{6}$$

Let us interpret the effects of the various parameters on the tradeoff between the two modes, i.e. what happens to the difference between the principal's profit in A-mode and its profit in P-mode. First, when the variance of b (i.e. V_b) increases, intuitively, the A-mode does relatively better because it allows the agent to react to this information in his choice of q and a. In contrast, the P-mode only allows for the agent's private information to be incorporated in his choice of q.

Second, when Φ increases, the tradeoff shifts in favor of the P-mode. Giving control over a to the principal results in a larger share of revenues kept by the principal, so that the choice of Q is closer to the first best in P-mode. When the principal's investment Q becomes more important (higher Φ), this effect is more important. The same reasoning explains why, when ϕ increases, i.e. the agent's investment q is more important, Π^{A*} increases more than Π^{P*}, so that the tradeoff is shifted in favor of the A-mode.

Third, when β increases, so that the transferable action becomes more important for contractible revenues, the effect on the tradeoff between the two modes is ambiguous. There are two opposing effects. On the one hand, a higher β benefits the P-mode because the P-mode does not involve any distortion in the choice of a, so the full increase in revenue due to a higher β is captured by the principal. In contrast, in A-mode, a higher β amplifies the distortion caused by the agent choosing a suboptimal level of a. This is captured by the extra term $-\frac{1}{2}\beta^2 t^2$ in (4), where t is evaluated at t^{A*}. On the other hand, a larger β also magnifies the importance of private information, which favors the A-mode. This is captured by the extra term $\frac{1}{2}\beta^2 V_b$ in (4). Thus, whether an increase in β increases the profits of A-mode over P-mode depends on the sign of $V_b - \left(t^{A*}\right)^2$, which can be positive or negative. For example, when β is large, $V_b - \left(t^{A*}\right)^2$ is positive, so further increasing β makes the A-mode relatively more attractive.

Fourth and finally, a larger average private benefit $\bar{b}$ shifts the tradeoff in favor of the P-mode. The reason is that the P-mode creates less total distortion due to revenue-sharing since a is set ex-ante. As a result, this means that demand and contractible revenues are higher in the P-mode, which multiply the expected private benefit.

5.2 Optimal threshold delegation

Now suppose the principal can monitor a and can therefore restrict the agent's choice of a according to some rule (i.e. H-mode). In particular, the principal could restrict the agent's choice of a to a degenerate interval $\{a_0\}$ that only contains one point—this effectively replicates the P-mode where the principal sets $a = a_0$ in its contract. At the other extreme, the principal's restriction rule could be so lax that it places no effective constraint on the agent's choice of a—this replicates the A-mode. For the sake of clarity, we will only use the label H-mode when the principal's restriction rule is neither one of these two extremes but instead involves some partial restriction in the choices facing the agent. Otherwise, we will refer to the contract choice as P-mode or A-mode given the equivalence noted above.

We first determine sufficient conditions for threshold delegation to be optimal in H-mode. In general, threshold delegation means that the agent is free to set a subject to either a minimum requirement, i.e. $a \geq x$ for some x, or a maximum requirement, i.e. $a \leq x$ for some x. Here, the agent always has a downward bias, so the relevant form of threshold delegation is that with a minimum requirement.

Proposition 2 *If $g'(b) \leq g(b)$ for all $b \in [b_L, b_H]$, the optimal contract in H-mode involves threshold delegation with a minimum requirement.*

The condition in the proposition just requires that the density does not increase too fast; it is obviously satisfied by the uniform distribution (which has constant density) and all distributions that have decreasing density in the positive domain (e.g. the normal or log-normal distributions).

While threshold delegation is not the optimal form of partial delegation for all distribution functions (which has also been noted by the existing literature on partial delegation), it does have the advantage of being simple to write down in a contract and relatively easy to monitor (as opposed to, for example, delegation that involves multiple intervals). This explains why threshold delegation is often used in practice.

Given a share $1-t$ of revenue and a minimum threshold x, the agent chooses

$$a = \begin{cases} \beta(1-t+b) & \text{if} \quad b \geq \frac{x}{\beta} - (1-t) \\ \\ x & \text{if} \quad b \leq \frac{x}{\beta} - (1-t). \end{cases}$$

The principal extracts the agent's entire expected payoff, so the principal's profit is

$$\begin{aligned} \Pi^{H*} &= \max_{t,x} \left\{ \mathbb{E}\left[(1+b)(\beta a + \phi q + \Phi Q) - \frac{1}{2}a^2 - \frac{1}{2}q^2 - \frac{1}{2}Q^2 \right] \right\} \\ & \quad \text{subject to} \\ a &= \max\{\beta(1-t+b), x\} \\ q &= \phi(1-t+b) \\ Q &= \Phi t. \end{aligned}$$

For any given (t,x), let then

$$\begin{aligned}\Pi^H(t,x) \quad &\equiv \quad \mathbb{E}\left[\beta(1+b)\max\{\beta(1-t+b),x\} - \frac{1}{2}\max\{\beta(1-t+b),x\}^2\right] \\ &\quad +\frac{\phi^2}{2}\left(\left(1+\bar{b}\right)^2 + V_b - t^2\right) + \frac{\Phi^2}{2}t\left(2\left(1+\bar{b}\right)-t\right).\end{aligned}$$

This expression is very similar to that of A-mode profits as a function of t (see (4) above), except that the choice of a is constrained from below by x.

If $b_L \geq \frac{x}{\beta} - (1-t)$, then x places no effective constraint on the agent, who chooses $a = \beta(1-t+b)$ for all b—this replicates the A-mode. Thus, the principal's profits are the same as in A-mode:

$$\Pi^H(t,x) = \frac{1}{2}\left(\left(\beta^2+\phi^2\right)\left(\left(1+\bar{b}\right)^2 + V_b\right) - \left(\beta^2+\phi^2\right)t^2 + \Phi^2 t\left(2\left(1+\bar{b}\right)-t\right)\right) = \Pi^A(t).$$

Similarly, if $b_H \leq \frac{x}{\beta} - (1-t)$, then the constraint on a is always binding, so $a = x$ for all b. This is equivalent to the principal choosing $a = x$ contractually, i.e. the P-mode, which also avoids any additional monitoring costs associated with H-mode. The principal's resulting profits are

$$\Pi^H(t,x) = \beta\left(1+\bar{b}\right)x - \frac{1}{2}x^2 + \frac{\phi^2}{2}\left(\left(1+\bar{b}\right)^2 + V_b - t^2\right) + \frac{\Phi^2}{2}t\left(2\left(1+\bar{b}\right)-t\right) = \Pi^P(t,x). \quad (7)$$

As a result, the H-mode only refers to the case when (t,x) are "interior", i.e. such that

$$b_L < \frac{x}{\beta} - (1-t) < b_H.$$

The principal's profit is then

$$\begin{aligned}\Pi^H(t,x) \quad = \quad &\frac{\Phi^2}{2}t\left(2\left(1+\bar{b}\right)-t\right) + \frac{\phi^2}{2}\left(\left(1+\bar{b}\right)^2 + V_b - t^2\right) \qquad (8)\\ &+\int_{b_L}^{\frac{x}{\beta}-(1-t)}\left(\beta(1+b)x - \frac{1}{2}x^2\right)dG(b) \\ &+\int_{\frac{x}{\beta}-(1-t)}^{b_H}\beta^2\left((1+b)(1-t+b) - \frac{1}{2}(1-t+b)^2\right)dG(b).\end{aligned}$$

The advantage of delegating to the agent is that the agent will take into account the realized value of his private benefit when choosing a, so will set a closer to the first-best level, and the principal can extract this additional expected payoff through its fixed fee T. But the principal also needs to extract some of the (measurable) variable revenues (i.e. $t > 0$) in order to maintain an incentive to invest in Q to increase these revenues. When the principal sets $t > 0$ and delegates to the agent, the agent's choice of a will be biased downwards compared to the level of a the principal prefers. The principal therefore prefers to stipulate a minimum level of a to help offset this downward bias, although at the cost of having a set too high whenever b turns out to be particularly low.

Thus, in some sense, threshold delegation is a way for the principal to get the best of both worlds—

delegation and control. As a result, one would expect threshold delegation to dominate both pure modes in general. The following proposition confirms that this intuition is correct provided

$$b_H > \bar{b} + \frac{\Phi^2 \left(1 + \bar{b}\right)}{\Phi^2 + \phi^2}, \tag{9}$$

which requires that the upper support of the random variable b be sufficiently high.[10]

Proposition 3 *The H-mode always dominates the A-mode. If (9) holds, then the H-mode also dominates the P-mode.*

The comparison with the A-mode implies that putting some restriction on the agent's choice of a is always optimal. To see this, consider the H-mode with $t = t^{A*}$ and $x = \beta \left(1 - t^{A*} + b_L\right)$. This replicates the outcome of the A-mode. Now consider increasing x by a small amount κ. The only change in the resulting profits occurs when the realized b is in the interval $[b_L, b_L + \kappa]$. In this interval the minimum requirement on a binds. On the one hand, this means the H-mode involves strictly higher costs due to the higher level of a imposed. On the other hand, it also means the H-mode creates higher revenues. The cost effect is quadratic in κ and the revenue effect is linear in κ, so for small enough κ, the net effect on profits is positive. This means that the principal can always do strictly better than the A-mode by adding a minimum requirement that prevents the agent from choosing the lowest values of a.

The comparison with the P-mode implies that leaving some discretion to the agent over the choice of a is always optimal, provided that (9) holds, i.e. that the private benefit is sometimes sufficiently high relative to its expected value. The simple explanation for this is that the principal can always do at least as well as in the P-mode by setting $t = t^{P*}$ and allowing the agent to choose a above the level that the principal would choose in the P-mode (i.e. $\beta \left(1 + \bar{b}\right)$). Given revenue sharing creates a downward bias in the agent's choice of a, the agent would only want to set a higher a than the level chosen in the P-mode when it is also in the principal's interest to have a higher a. The inequality (9) imposed in the proposition can be written as $\beta \left(1 - t^{P*} + b_H\right) > \beta \left(1 + \bar{b}\right)$, which ensures that there are some realizations of b for which the agent would indeed want to set a above $\beta \left(1 + \bar{b}\right)$.

Without (9), even for the highest realization of b, the agent's choice of a (i.e. $\beta \left(1 - t^{P*} + b_H\right)$) would still be below what the principal would like to choose in P-mode, so the simple logic above no longer applies. We analyze the case when (9) does not hold and G is the uniform distribution in Section 5.3. In that case, we show that the P-mode can dominate threshold delegation.

Note that condition (9) holds whenever Φ is close enough to 0 or ϕ is large enough, which means that partial delegation dominates the two pure modes whenever the principal's moral hazard is sufficiently small or the agent's moral hazard is sufficiently large. In these cases, the principal finds it optimal to retain a very low revenue share t^{P*} in P-mode, an outcome which it can improve upon by keeping

[10] The inequality (9) holds for any distribution with support that contains $\mathbb{R}_+$ (e.g. the normal distribution or the exponential distribution).

$t = t^{P*}$ and giving control over a to the agent. Indeed, given the low t^{P*}, the downward bias of the agent is minimal and can be easily removed with a low minimum threshold.

Proposition 3 says that the A-mode can never be optimal. However, this conclusion holds in the absence of any fixed costs that the principal might incur when operating in a particular mode. In reality, the H-mode is likely to incur higher fixed costs than the A-mode, due to the need to monitor the agent in order to ensure he respects the constraint imposed by the threshold x in H-mode. Clearly, if this monitoring cost is sufficiently large and condition (6) holds, then the A-mode will be optimal. In Section 5.4 we examine how the optimal choice of mode depends on the nature of monitoring costs.

In the rest of this section, we provide some additional properties of the optimal solution in H-mode. To do this, we assume that the first-order conditions of $\Pi^H(t,x)$ characterize a unique interior maximum $\left(t^{H*}, x^*\right)$. From (8), the first-order derivatives of H-mode profits in x and t are

$$\begin{aligned}
\frac{\partial \Pi^H(t,x)}{\partial x} &= \int_{b_L}^{\frac{x}{\beta}-(1-t)} \left(\beta(1+b) - x\right) dG(b) \\
\frac{\partial \Pi^H(t,x)}{\partial t} &= \Phi^2\left(1+\bar{b}\right) - t\left(\Phi^2+\phi^2\right) - \int_{\frac{x}{\beta}-(1-t)}^{b_H} t\beta^2 dG(b).
\end{aligned}$$

Differentiating $\frac{\partial \Pi^H(t,x)}{\partial x}$ with respect to t leads to the following result.

Proposition 4 *The share of revenues kept by the principal t and the minimum threshold x are strategic complements. I.e.* $\frac{\partial^2 \Pi^H(t,x)}{\partial t \partial x} > 0$.

This result contains a clear empirical implication: all other things equal, we should observe principals retaining a higher share of variable revenues when they choose higher (or more restrictive) minimum requirements. The intuition is straightforward. A higher minimum threshold for a is more restrictive for the agent, so there is less need to leave the agent with a large share of revenue to reduce the distortion in the agent's choice of the transferable action a. The first result in the next Proposition is consistent with this intuition.

Proposition 5 *The optimal share of revenue retained by the principal is highest in P-mode, lowest in A-mode, and intermediate in H-mode, i.e.* $0 < t^{A*} < t^{H*} < t^{P*} < 1$. *Moreover, the optimal minimum requirement for a in H-mode is below the fixed choice of a in P-mode, i.e.* $x^* < \beta\left(1+\bar{b}\right)$.

Compared to its optimal solution in H-mode or A-mode, the principal should retain a higher share of revenues in P-mode. This reflects that in P-mode, the amount of variable revenue retained by the principal does not distort the choice of the transferable action a, which is fixed in the contract, whereas it does in the other two modes. Compared to the optimal solution in A-mode, the principal should retain a greater share of revenue in H-mode because there is less scope for revenue sharing to distort the agent's choice of the transferable action a when the agent's choice of a is restricted.

For the second part of the proposition, we already know from the discussion after Proposition 3 that in H-mode the principal should always give the agent discretion for choices of a that are above the fixed choice of $a^{P*} = \beta\left(1+\bar{b}\right)$ in P-mode. In this region, the principal's and agent's interests are aligned. However, the second result in Proposition 5 indicates that the principal should also give agents discretion for some choices of a that are below (but not too much below) a^{P*} so that the agent can make use of his information regarding his private benefits in these cases as well, provided that the realized value of his private benefit is not too low.

We can also derive the following comparative static results.

Proposition 6 *A larger Φ shifts the tradeoff between A-mode and H-mode in favor of H-mode, and shifts the tradeoff between H-mode and P-mode in favor of P-mode. A larger ϕ has the opposite effect on these tradeoffs. I.e.*

$$\begin{aligned}\frac{d\Pi^{P*}}{d\Phi} &> \frac{d\Pi^{H*}}{d\Phi} > \frac{d\Pi^{A*}}{d\Phi}\\ \frac{d\Pi^{P*}}{d\phi} &< \frac{d\Pi^{H*}}{d\phi} < \frac{d\Pi^{A*}}{d\phi}.\end{aligned}$$

Moreover, the optimal H-mode solutions t^{H*} and x^* are increasing in Φ and decreasing in ϕ.

When the principal's ongoing investment in quality becomes more important in determining revenue, it makes the P-mode more desirable relative to the H-mode and the H-mode more desirable relative to the A-mode.[11] And vice versa when the agent's ongoing investment in quality becomes more important. This is easily understood: the principal's moral hazard is best addressed in the P-mode, whereas the agent's is best addressed in the A-mode. The H-mode lies in between these two extremes. The same logic also applies to the optimal contract in H-mode. When the principal's (respectively, the agent's) ongoing investment in quality becomes more important in determining revenue, it is natural that t^{H*} increases (respectively, decreases) to help offset the principal's (respectively, the agent's) more important moral hazard problem. Due to the strategic complementarity of t and x in H-mode, this also means a larger Φ or a smaller ϕ will increase x^*, i.e. the principal should impose a stricter restriction on a.

5.3 The effect of uncertainty

In this section we explore the effect of increased uncertainty over b by focusing on the case b follows a uniform distribution. Specifically, let $b_L = \bar{b} - \sigma$ and $b_H = \bar{b} + \sigma$, so that the variance of b is $V_b = \frac{\sigma^2}{3}$. Since we must have $b_L \geq 0$, we assume

$$\sigma \leq \bar{b}.$$

To keep the analysis as streamlined as possible, we also assume

$$\frac{\Phi^2}{\phi^2} < \bar{b} < \frac{\phi^2}{\Phi^2},$$

[11] As pointed out above, if the H-mode involves fixed monitoring costs, then the P-mode or A-mode could still be preferred in practice despite the result in Proposition 3.

which requires that the agent's investment (e.g. effort) is more important than that of the principal.

Throughout this section, we focus on the effect of varying σ while holding $\bar{b}$ constant. We first characterize the interior solution for the H-mode and the effect of σ on the optimal choice of mode.

Proposition 7 *The optimal solution in the H-mode is interior if and only if $\sigma \in \left[\frac{\Phi^2(1+\bar{b})}{\Phi^2+\phi^2}, \bar{b}\right]$ or ($\sigma \in \left[\frac{4\Phi^2\beta^2(1+\bar{b})}{(\Phi^2+\phi^2+\beta^2)^2}, \frac{\Phi^2(1+\bar{b})}{\Phi^2+\phi^2}\right]$ and $\beta^2 > \Phi^2 + \phi^2$). It is then characterized by*

$$\begin{aligned} t^{H*} &= \frac{\sigma}{2\beta^2}\left(\Phi^2+\phi^2+\beta^2-\sqrt{(\Phi^2+\phi^2+\beta^2)^2-\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{\sigma}}\right) \\ x^* &= \beta\left(1+t^{H*}+\bar{b}-\sigma\right), \end{aligned}$$

where both t^{H*} and x^* are decreasing in σ. Furthermore, if the optimal solution in the H-mode is interior, then a larger σ shifts the tradeoff between A-mode and H-mode in favor of A-mode, and shifts the tradeoff between H-mode and P-mode in favor of H-mode, i.e.

$$\frac{d\Pi^{A*}}{d\sigma} > \frac{d\Pi^{H*}}{d\sigma} > \frac{d\Pi^{P*}}{d\sigma}.$$

It is easily verified that both t^{H*} and x^* are decreasing in σ. The intuition behind this result is straightforward: increasing the variance of the private benefit (holding its expectation constant) means increasing the importance of the agent's private information, which makes delegation more attractive. In turn, this implies the principal should leave a greater share of revenue to the agent and give the agent more discretion to choose a (i.e. a lower minimum requirement).

The second part of the proposition says that the larger the variance of the private benefit (holding its expectation constant), the more attractive the A-mode becomes relatively to the H-mode, and the more attractive the H-mode becomes relatively to the P-mode. This reflects that the H-mode leverages the agent's private information to a larger extent than the P-mode, and the A-mode leverages the agent's private information to the highest extent. In the absence of monitoring costs, this result implies that there exists a cutoff $\widehat{\sigma}$, such that the H-mode is optimal for $\sigma \geq \widehat{\sigma}$ and the P-mode is optimal for $\sigma \leq \widehat{\sigma}$ (recall from Proposition 3 that the A-mode is dominated by the highest profit achieved with interior (t, x) in H-mode).

5.4 Monitoring costs

So far we have largely abstracted from monitoring costs. This is reasonable in settings where monitoring the revenues generated by the agent (which is necessary to implement revenue sharing in all three modes) is sufficient to determine whether the agent's choice of a conforms to the principal's minimum requirements. In such cases, no additional monitoring costs are required in H-mode. For instance, a movie studio effectively monitors the number of weeks that a movie theater shows a particular movie simply by monitoring ticket revenues for that movie.

However, in some other cases the H-mode incurs higher fixed costs than either of the pure modes, reflecting the additional monitoring costs necessary to verify whether the minimum requirements on a are respected. To keep things as simple as possible, we assume that the monitoring technology requires a fixed ex-ante investment (e.g. hiring additional managers and staff) and no extra cost ex-post (see Gal-Or, 1995), thus abstracting away from any strategic monitoring game that the principal and the agent might engage in (as in Lal, 1990). We denote the fixed monitoring costs by F and assume for simplicity that there are no other fixed costs in any of the modes. In particular, we first assume monitoring is not required in P-mode because the principal directly sets a and incurs the associated costs (e.g. traditional taxi companies choose the cars for their drivers and bear the associated costs). The existence of such monitoring costs in H-mode explains why A-mode can be optimal over some parameter ranges. Obviously, the A-mode, which does not require a monitoring cost, dominates the H-mode whenever F is large enough. Moreover, we know that the A-mode dominates the P-mode whenever the variance of b is high enough, i.e. from the condition in Proposition 1 in general, which becomes

$$\sigma > \frac{\sqrt{3}\Phi^2\left(1+\bar{b}\right)}{\sqrt{\left(\Phi^2+\phi^2\right)\left(\Phi^2+\phi^2+\beta^2\right)}}$$

when b is drawn from the uniform distribution.

Figure 1 illustrates the optimal choice of mode as a function of $(F, \sigma) \in [0, 0.05] \times [0, 1]$, when b is drawn from the uniform distribution with the other parameter values set at $\bar{b} = 1$, $\beta = 1$, $\phi = 2$ and $\Phi = 1$. We use light gray to show the region for which A-mode is optimal, and likewise dark gray for H-mode and black for P-mode. As σ increases, the optimal mode shifts from P-mode to H-mode and then to A-mode (consistent with the second part of Proposition 7), or directly from P-mode to A-mode when F is high enough (consistent with Proposition 1). Obviously, for a given level of σ, a higher monitoring cost F shifts the optimal mode away from H-mode—towards the A-mode if σ is high or the P-mode if σ is low.

Finally, in other cases, the transferable action fixed by the principal in its P-mode contract is something that the principal cannot actually control directly without monitoring. For example, the principal does not directly control the maintenance or quality of the agent's premises (e.g. at a retailer, dealer or franchisee), the agent's opening hours, or the effort supplied by agents (this may be measured by user ratings, e.g. average driver ratings on Uber and Lyft, average contractor ratings on Upwork, etc.). In these cases, agents always make these choices and incur the associated costs. Thus, if the principal wants to fix a at a certain level in P-mode, it needs to monitor the agent's choice to ensure it matches the contracted level. Formally, this just means that the monitoring cost F must also be incurred in P-mode,[12] so the only way the principal can forgo the monitoring cost is by choosing the A-mode, in which it does not exercise any control over the choice of a.

Figure 2 repeats the analysis of the optimal choice of mode for this case, assuming that the principal incurs the monitoring cost F in both H-mode and P-mode. Comparing Figures 1 and 2, clearly the H-mode is now optimal for a wider range of parameters. While the H-mode vs. A-mode tradeoff is

[12] As shown in (7), the only distinction between P-mode and H-mode with a restricted to a single value comes from any additional monitoring costs incurred in H-mode.

Figure 1: Optimal mode (F incurred in H-mode only)

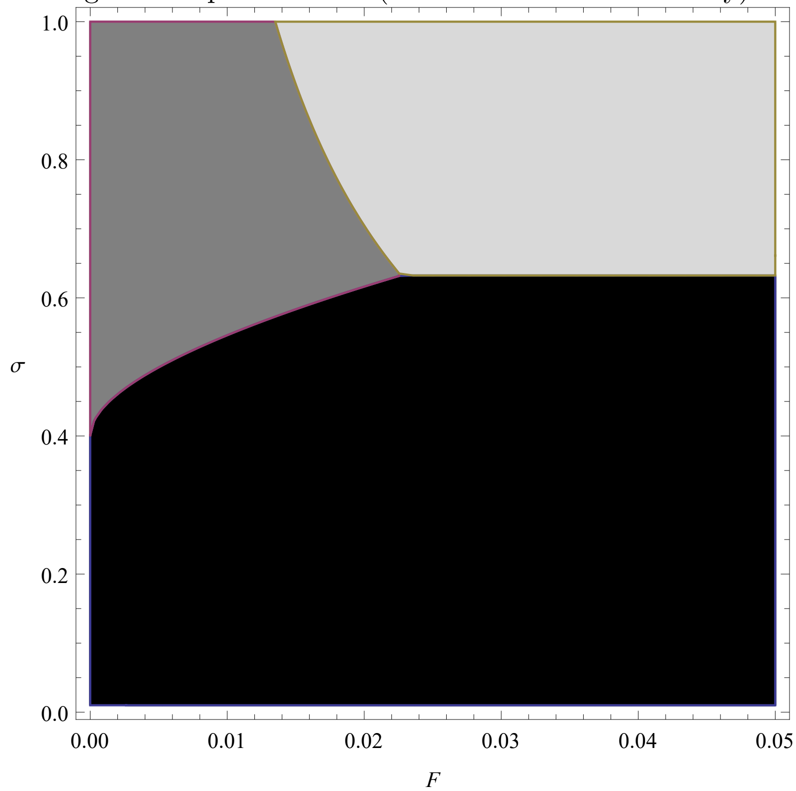
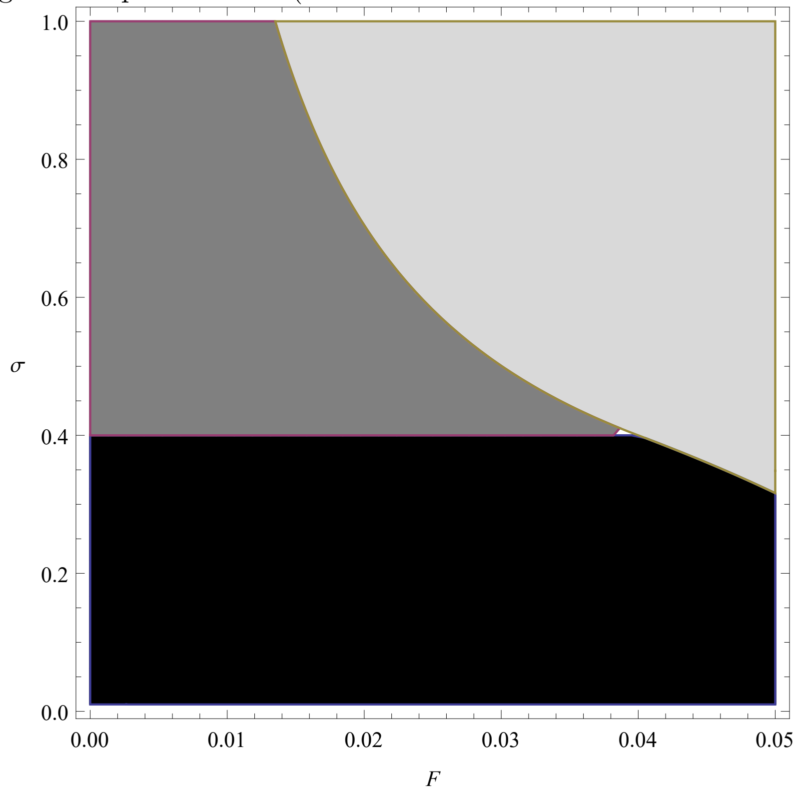


Figure 2: Optimal mode (F incurred in both P-mode and H-mode)



unchanged, the H-mode vs. P-mode tradeoff is now shifted in favor of the H-mode, reflecting that F must now also be incurred in P-mode. For this reason, a higher level of F no longer shifts the optimal mode away from H-mode towards P-mode. Instead, a higher level of F shifts the optimal mode from both H-mode and P-mode towards the A-mode.

6 Extensions

In this section we consider two extensions of our benchmark model from the two previous sections. First, we show how our theory of partial delegation can be modified to provide a new explanation for the use of RPM contracts. Second, we introduce multiple agents and network effects in our benchmark model and explore their effect on the principal's choice of delegation mode for the transferable decision variable a.

6.1 Resale price maintenance

Our theory of partial delegation can be applied to provide a new explanation for the use of resale price maintenance (RPM) contracts. To do so, we adapt our benchmark model of Section 4 by making price the transferable decision variable which may be set either by the principal or the agent. In this context, the principal may be an upstream firm (e.g. a manufacturer) and the agent a downstream firm (e.g. a retailer), although our model of RPM applies much more generally.

Specifically, suppose demand is given by

$$\theta - \beta p + \phi q + \Phi Q,$$

where θ is the demand intercept. We assume θ satisfies the same properties that were satisfied by b in the benchmark model, and denote by $\overline{\theta}$ and V_θ the expected value and variance of θ, respectively. Thus, the agent now has private information regarding the demand shock θ instead of the private benefit b. We also assume

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

which is necessary and sufficient for all second order conditions to hold. Costs are normalized to zero[13] and there are no private benefits. The transferable decision variable is now the price p charged to consumers. The upstream manufacturer charges the downstream agent (retailer) a per-unit wholesale price w and a fixed fee T. Thus, the principal's profit is

$$w\left(\theta - \beta p + \phi q + \Phi Q\right) + T$$

and the agent's payoff is

$$\left(p - w\right)\left(\theta - \beta p + \phi q + \Phi Q\right) - T.$$

[13] Adding a constant marginal cost of production c for the principal (e.g. the manufacturer's production cost), does not change the analysis except that θ is replaced everywhere by $\theta - \beta c$.

The timing assumptions remain unchanged.

The assumptions that the agent's private information is with respect to the demand shock θ and that the principal uses wholesale pricing contracts fit the main motivating example for RPM contracts—the case in which the principal is a manufacturer and the agent is a retailer. These assumptions also ensure the analysis is tractable. Furthermore, the two-part tariff used here is consistent with our approach in the previous sections, where revenue sharing and the use of a per-unit wholesale price were equivalent. The analysis of RPM with private benefits yields very similar results and is provided in the online appendix.

To understand what drives the bias between the principal's and the agent's preferred level of price, consider first the P-mode, in which the principal sets p (together with w and T) in its contract. Taking into account that T is set to extract the agent's expected payoff, the principal solves

$$\begin{aligned} & \max_{w,p} \mathbb{E}\left[p\left(\theta-\beta p+\phi q+\Phi Q\right)-\frac{1}{2}q^{2}-\frac{1}{2}Q^{2}\right] \\ & \text{subject to} \\ q &= (p-w)\,\phi \\ Q &= w\Phi. \end{aligned} \tag{10}$$

Given a wholesale price w, the principal sets

$$p^{P}(w) = \frac{\overline{\theta}+\Phi^{2}w}{2\beta-\phi^{2}} \tag{11}$$

$$Q^{P}(w) = w\Phi, \tag{12}$$

while the agent sets

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

By contrast, in A-mode the agent sets the price p in stage 2, so the principal solves

$$\begin{aligned} & \max_{w}\left\{\mathbb{E}\left[p\left(\theta-\beta p+\phi q+\Phi Q\right)-\frac{1}{2}q^{2}-\frac{1}{2}Q^{2}\right]\right\} \\ & \text{subject to} \\ p &= \frac{w}{2}+\frac{\theta+\phi q+\Phi Q}{2\beta} \\ q &= (p-w)\,\phi \\ Q &= w\Phi. \end{aligned} \tag{13}$$

Solving for (p, q, Q) as functions of w, we obtain

$$p^{A}(w)=\frac{\theta+\left(\Phi^{2}+\beta-\phi^{2}\right)w}{2\beta-\phi^{2}} \tag{14}$$

$$q^A(w) = \left(\frac{\theta + \left(\Phi^2 - \beta\right) w}{2\beta - \phi^2}\right)\phi \tag{15}$$

$$Q^A(w) = w\Phi. \tag{16}$$

Equations (11) and (14) reveal that for the same positive level of w and assuming the same level of θ (e.g. $\overline{\theta}$), the agent has an upward bias in choosing the price if $\beta > \phi^2$ and a downward bias in choosing the price if $\beta < \phi^2$. The existence of a bias (upward or downward) is due to the positive wholesale price, which, like revenue sharing, leads to choices of the agent that are distorted away from the levels preferred by the principal. In textbook models, a manufacturer can eliminate such biases by setting the wholesale price equal to its marginal cost (zero in this case), so that the retailer retains the full margin associated with its decisions. However, this is not possible in our setting, because the principal must keep a positive share of revenues in order to mitigate its own moral hazard problem.

Note also that

$$\mathbb{E}\left[q^A(w)\right] = \left(\frac{\overline{\theta} + \left(\Phi^2 - \beta\right) w}{2\beta - \phi^2}\right)\phi < q^P(w) = \left(\frac{\overline{\theta} + \left(\Phi^2 + \phi^2 - 2\beta\right) w}{2\beta - \phi^2}\right)\phi$$

if and only if $\beta < \phi^2$, i.e. if and only if the agent has a downward bias in choosing the price. Thus, whenever the agent has a downward bias in choosing the price, the strategic complementarity of p and q implies that the expected level of effort q chosen by the agent will also be too low.

Like in the benchmark model, the principal may do better than either of the pure modes by using partial delegation, i.e. H-mode, whereby the principal allows the agent to choose price subject to some constraints. In the following proposition we establish that a sufficient condition for threshold delegation—with a minimum or a maximum requirement, depending on the sign of the agent's bias—to be optimal in H-mode is that the density of θ is non-increasing. Obviously, this includes the case in which θ is drawn from the uniform distribution on $[\theta_L, \theta_H]$.

Proposition 8 *If $g'(\theta) \leq 0$ for all $\theta \in [\theta_L, \theta_H]$, the optimal contract in H-mode involves threshold delegation.*

We can therefore focus on contracts in which the agent is free to set the price subject to either a price ceiling or a price floor. The following proposition provides conditions under which each scenario—maximum RPM (a price ceiling) and minimum RPM (a price floor)—is optimal.

Proposition 9 (Resale price maintenance)

Maximum RPM: *If $\beta > \phi^2$, the H-mode with maximum RPM dominates the A-mode. If in addition $\theta_L < \left(1 - \frac{\Phi^2\left(\beta - \phi^2\right)}{(2\beta-\phi^2)(\phi^2+\Phi^2)-\Phi^4}\right)\overline{\theta}$, the H-mode with maximum RPM also dominates the P-mode.*

Minimum RPM: *If $\beta < \phi^2$, the H-mode with minimum RPM dominates the A-mode. If in addition $\theta_H > \left(1 + \frac{\Phi^2\left(\phi^2 - \beta\right)}{(2\beta-\phi^2)(\phi^2+\Phi^2)-\Phi^4}\right)\overline{\theta}$, the H-mode with minimum RPM also dominates the P-mode.*

The main condition determining whether minimum or maximum RPM is optimal hinges on whether the agent's moral hazard in setting q (as measured by ϕ^2) is sufficiently important (in which case minimum RPM is best) or not very important (in which case maximum RPM is best). To understand the result, note there are two underlying sources of bias, which go in opposite directions. First, absent the agent's moral hazard problem, the price markup set by the agent will be too high since it does not internalize the revenue obtained by the principal. This is the usual pricing distortion that underlies classic double marginalization and its magnitude here is measured by β. The second bias, which runs in the opposite direction, is that the agent under-invests in effort q because it does not internalize the revenues generated from its effort that are captured by the principal. This moral hazard problem can be addressed by leaving the agent with a higher margin $(p - w)$ than the agent would choose itself. When $\beta > \phi^2$, the agent's moral hazard problem is not very important and the first bias dominates, meaning the agent has an upward bias in setting prices, which calls for a price ceiling (maximum RPM). Alternatively, when $\beta < \phi^2$, the agent's moral hazard is more important, so the second bias dominates. This means the agent has a downward bias in setting prices, which calls for a price floor (minimum RPM).

Thus, in contrast to most existing theories of RPM, Proposition 9 can explain the use of both minimum and maximum RPM, and when each would be used. The use of a constraint on prices rather than specifying a specific level of prices reflects the realistic feature that a principal often wants to give the agent discretion to react to its private information on demand shocks, while mitigating the worst pricing biases that can arise when the agent controls the price but only keeps some of the associated variable revenue. In practice, as noted in Section 3, retailers sometimes set prices below price ceilings or above price floors, consistent with our theory but inconsistent with standard RPM theories.

6.2 Network effects and pessimistic expectations

Returning to our benchmark model of Section 4 with private benefits, suppose now that instead of one agent there are N identical agents and that the revenue attributable to each agent (ignoring private benefits) is

$$R(a, q, Q) = R_0(n) + \beta a + \phi q + \Phi Q,$$

where $n \in \{1, .., N\}$ is the total number of agents who actually join. We assume $R_0'(n) > 0$, i.e. there are agglomeration (network) effects. Let

$$\Delta R_0 \equiv R_0(N) - R_0(1) \geq 0,$$

which is a parameter that captures the strength of network effects. With network effects, agents' expectations matter for the outcome. We focus on two polar types of expectations: optimistic and pessimistic (see for example Caillaud and Jullien 2003, and Hagiu and Spulber 2013). With optimistic expectations, according to which each agent expects all other agents to join whenever that is an equilibrium outcome, everything is the same as before, except for the extra constant $NR_0(N)$ which has no impact on the analysis. In this case, the results in Section 5 continue to apply. We therefore focus on the case with pessimistic expectations. This means that agents coordinate on not joining whenever this is an equilibrium. To compare the outcome under pessimistic expectations with the one under optimistic expectations, we will treat $R_0(N)$ as a constant and vary ΔR_0. Clearly, if $\Delta R_0 = 0$, i.e. if there are no network effects, then the profits under the various modes are the same regardless of agents' expectations. Suppose then $\Delta R_0 > 0$.

Under pessimistic expectations the fixed fee that renders agents indifferent between joining and their outside option must now be lower (for any given t) because each agent expects no other agent to join whenever this is an equilibrium. Indeed, the total payoff that each agent expects to receive when no one else joins is $\left(1 - t + \overline{b}\right)\left(R_0(1) + \beta a + \phi q + \Phi Q\right) - T$. Since this total payoff depends on $R_0(1)$ rather than $R_0(N)$, the fixed fee T has to be lower by $\left(1 - t + \overline{b}\right)\Delta R_0$ to attract all agents. With the fixed fee lowered by this amount, all agents join and the principal's profit is equal to its profit under optimistic expectations minus the reduction in the fixed fee, i.e. it is lower by $\left(1 - t + \overline{b}\right)\Delta R_0 N$. This will result in t being chosen differently than under optimistic expectations and therefore will also affect the optimal choice of mode. Note that due to the private benefits $\overline{b}$, even if $t = 1$, network effects do not completely disappear.

In theory, pessimistic expectations could create an incentive to set $t > 1$ and $T < 0$ in order to increase agents' willingness to join even if no one else is expected to join. This would also imply that the optimal choice of q would be zero, and in A-mode, the optimal choice of a would be zero. To rule out any such perverse case we assume

$$\phi^2 > \Phi^2 \overline{b} + \Delta R_0,$$

so that the principal's optimal t will always be within $[0, 1]$.

Under pessimistic expectations, the principal's profit per agent in P-mode, A-mode and H-mode as functions of t is just the same as in (2), (4) and (8) respectively, except for the additional term $\left(1 + \overline{b}\right) R_0(N) - \left(1 - t + \overline{b}\right)\Delta R_0$. The existence of this additional term does not affect the order of revenue shares in Proposition 5, which continues to hold. Specifically, it is straightforward to modify the proof of Proposition 5 to obtain

$$0 < t_{PE}^{A*} = \frac{\Phi^2\left(1 + \overline{b}\right) + \Delta R_0}{\Phi^2 + \phi^2 + \beta^2} < t_{PE}^{H*} = \frac{\Phi^2\left(1 + \overline{b}\right) + \Delta R_0}{\Phi^2 + \phi^2 + \beta^2\left(1 - G\left(\frac{x}{\beta} - (1 - t^{H*})\right)\right)} < \frac{\Phi^2\left(1 + \overline{b}\right) + \Delta R_0}{\Phi^2 + \phi^2} = t_{PE}^{P*} < 1.$$

Denote the resulting optimal profits per agent in each of the three modes by $\pi_{PE}^{P*}(\Delta R_0)$, $\pi_{PE}^{A*}(\Delta R_0)$ and $\pi_{PE}^{H*}(\Delta R_0)$. Using the expressions of t_{PE}^{P*}, t_{PE}^{H*} and t_{PE}^{A*} above and the Envelope theorem, we have

$$\frac{d\pi_{PE}^{H*}}{d(\Delta R_0)} - \frac{d\pi_{PE}^{A*}}{d(\Delta R_0)} = t_{PE}^{H*} - t_{PE}^{A*} > 0 > t_{PE}^{H*} - t_{PE}^{P*} = \frac{d\pi_{PE}^{H*}}{d(\Delta R_0)} - \frac{d\pi_{PE}^{P*}}{d(\Delta R_0)}.$$

Recalling that the principal's per agent profits under optimistic expectations are $\pi_{PE}^{P*}(0)$, $\pi_{PE}^{A*}(0)$ and $\pi_{PE}^{H*}(0)$, we obtain the following proposition.

Proposition 10 *Decreasing the magnitude of network effects under pessimistic expectations or changing expectations from pessimistic to optimistic shifts the tradeoff between P-mode and H-mode in favor of H-mode, shifts the tradeoff between H-mode and A-mode in favor of A-mode, and shifts the tradeoff between P-mode and A-mode in favor of A-mode.*

Thus, if the principal faces pessimistic expectations, stronger network effects shift the tradeoff in favor of the mode in which the principal has more control. The reason is that the principal leaves a lower share of variable revenues to agents in such a mode, which weakens network effects, meaning that the negative effect of pessimistic expectations is also weakened. An implication of this result is that the P-mode (or H-mode) might still dominate even when there is no moral hazard on the part of the principal, i.e. when $\Phi = 0$. Furthermore, due to the strategic complementarity between t and x in H-mode (recall Proposition 4, which continues to hold here), stronger network effects also increase the optimal threshold x^* in H-mode. In other words, if faced with pessimistic expectations, the principal should place more stringent restrictions on the agent when network effects are stronger.

Since the switch from optimistic expectations to pessimistic expectations is equivalent to increasing the magnitude of network effects from 0 to ΔR_0 while holding $R_0(N)$ fixed, the effects of the switch in expectations on the tradeoffs between the three modes and on the optimal restrictions used in H-mode are the same as the effects of increasing network effects described in the previous paragraph.

7 Managerial implications

The emergence of digital monitoring and data analytics technologies has created more opportunities for firms to enforce different degrees of delegation in a cost effective way. As a result, partial delegation is likely to become a contractual instrument that a greater number of firms that act as principals (e.g. franchisors, platforms, manufacturers, movie studios) can consider using when setting the terms for their agents (e.g. franchisees, third-party suppliers, retailers, movie theaters). Our theory provides several lessons for managers in this regard.

At a high level, we have shown that delegation subject to minimum requirements strikes a middle ground between complete delegation to agents and full control by the principals and oftentimes does better than both. It is a way to get the best of both worlds, by leveraging the relevant agents' private information, while also eliminating the more extreme biases that arise when agents only keep some of the revenues they help produce. For platforms/marketplaces, one can view the use of minimum requirements as a non-price governance rule designed to achieve strategic positions that are intermediate along the spectrum between pure marketplace/platform (e.g. relying on independent contractors) and pure vertical integration (e.g. relying on employees). In manufacturer-retailer contexts, threshold delegation is an additional contracting instrument that can improve channel coordination beyond what can be achieved with typical pricing instruments (revenue shares, wholesale pricing, quantity discounts, etc).

We now spell out a few more specific implications of the theory.

First, when principals extract a high share of revenues, they should place more stringent requirements on their agents' actions (e.g. a higher minimum level of advertising or a lower maximum price). Conversely, when giving their agents more autonomy over these actions, principals should extract a smaller share of revenue (Proposition 4). As a result, when moving from full control to partial delegation, the share of revenues left to agents should increase. The same is true when moving from partial delegation to complete delegation, or when minimum requirements are raised (Proposition 5). This is a variation on the principle that high-powered incentives (revenue share) should be aligned with low-powered incentives (control over transferable actions). Thus, for example, platforms that only impose loose or no governance rules on participants should also extract low revenue shares, whereas platforms that impose tight governance rules should extract higher revenue shares.

Second, in settings where the agents' moral hazard is more important relative to the principals' or where there is more uncertainty regarding the agents' private information (i.e. higher variance of private shocks), the principals should give the agents more autonomy, i.e. switch from full control to partial delegation, reduce minimum requirements if they are already in place, or even switch from partial delegation to full delegation to the agents (Propositions 6 and 7) when there are positive monitoring costs.

Third, if network effects are present, then principals should use more restrictions on agents (or even switch to full control) when agents' expectations are pessimistic rather than optimistic (Proposition 10). And vice versa. Pessimistic expectations are particularly relevant for early-stage ventures facing network effects, so such ventures should adopt business models that involve more control or tighter restrictions over transferable actions relative to later-stage ventures, which presumably benefit from optimistic expectations. For example, Uber started off in 2009 with black cars only (i.e. with high minimum requirements on the cars used by their drivers) in a few metropolitan areas. After the company reached critical mass in a number of cities (so that it arguably faced optimistic rather than pessimistic expectations), it launched UberX in 2012, which allowed drivers with any cars subject to far less stringent minimum requirements to participate.

Fourth, principals should consider making use of resale price maintenance (RPM)—where doing so is legal—so as to obtain the benefits from threshold delegation predicted by our theory. Specifically, the principal should use RPM with price ceilings (respectively, price floors) whenever demand is more (respectively, less) price sensitive and the agents' moral hazard is less (respectively, more) important.

8 Conclusions

We have established the key tradeoffs faced by firms that act as principals when choosing the extent to which they should delegate control over transferable decisions to their agents, in contexts in which both the principal and its agents need to be incentivized to make revenue-enhancing on-going investments, and the agent has private information. We have also emphasized the linkage between two key choices that such principals have to make—the extent of decision autonomy and the revenue share given to their agents—thereby providing an integrated theory of both instruments.

There are several promising directions in which our analysis can be extended. One could introduce multiple agents and (positive or negative) spillovers from each agent's choice of the transferable action on the revenues derived by other agents. The question would then be whether the spillovers lead to more or less delegation to the agents. Next, one could allow for multiple transferable decision variables and explore whether the levels of delegation chosen by the principal for these variables are positively correlated. Another avenue would be to introduce risk aversion or wealth constraints for agents, so that they cannot pay large fixed fees upfront. This would increase the principal's optimal share of revenues in all three modes, and so should shift the tradeoff in favor of less delegation. It would also be interesting to extend our model to multiple competing principals. This could possibly generate equilibria in which principals compete with different delegation models.

References

Abhishek, V., K. Jerath and Z.J. Zhang (2015) "Agency Selling or Reselling? Channel Structures in Electronic Retailing," *Management Science*, 62 (8), 2259-2280.

Alonso R. and N. Matouschek (2008) "Optimal Delegation," *Review of Economic Studies*, 75 (1), 259-293.

Amador, M. and K. Bagwell (2013) "The Theory of Optimal Delegation With an Application to Tariff Caps," *Econometrica*, 81 (4), 1541–1599.

Anderson, E. (1985) "The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis," *Marketing Science*, 4 (3), 234-54.

Asker, J. and H. Bar-Isaac (2014) "Raising Retailers' Profits: On Vertical Practices and the Exclusion of Rivals," *American Economic Review*, 104 (2), 672-686.

Blair, R.D. and F. Lafontaine (2010) *The Economics of Franchising*, Cambridge University Press, New York.

Boudreau, K.J. (2010) "Open Platform Strategies and Innovation: Granting Access vs. Devolving Control," *Management Science*, 56 (10), 1849-1872.

Boudreau, K. J., and A. Hagiu (2009) "Platform Rules: Multi-Sided Platforms As Regulators," in *Platforms, Markets and Innovation*, edited by Annabelle Gawer. Cheltenham, UK: Edward Elgar Publishing.

Boudreau, K.J. (2012) "Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation," *Organization Science*, 23 (5), 1409–1427.

Cachon, G. P. and M. Lariviere (2005) "Supply chain coordination with revenue-sharing contracts: Strengths and limitations," *Management Science*, 51 (1), 30–44.

Caillaud, B. and B. Jullien (2003) “Chicken and egg: Competition among intermediation service providers,” *Rand Journal of Economics*, 34 (2), 309–328.

Casadesus-Masanell, R. and H. Halaburda (2014) “When Does a Platform Create Value by Limiting Choice?,” *Journal of Economics & Management Strategy,* 23(2), 258-292.

Deneckere, R., H. Marvel and J. Peck (1996). “Demand Uncertainty, Inventories, and Resale Price Maintenance,” *Quarterly Journal of Economics*, 111 (3), 885-913.

Desiraju, R. and S. Moorthy (1997) “Managing a Distribution Channel under Asymmetric Information with Performance Requirements,” *Management Science*, 43(12), 1628-1644.

Foros, Ø., H.J. Kind and G. Shaffer (2013) “Turning the Page on Business Formats for Digital Platforms: Does Apple’s Agency Model Soften Competition?” CESifo Working Paper 4362.

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

Gal-Or, E. (1995) “Maintaining Quality Standards in Franchise Chains,” *Management Science*, 41 (11), 1774-1792.

Gans, J. (2012) “Mobile Application Pricing,” *Information Economics and Policy*, 24, 52–59.

Gawer A. and M. Cusumano (2002) *Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation*, Boston, MA: Harvard Business School Press.

Hagiu, A. (2014) “Strategic Decisions for Multi-Sided Platforms,” *MIT Sloan Management Review*, 55(2), 71-80.

Hagiu, A. and D. Spulber (2013) “First-Party Content and Coordination in Two-Sided Markets,” *Management Science* 59 (4), 933–949.

Hagiu, A. and J. Wright (2015) “Marketplace or Reseller?” *Management Science*, 61(1), 184–203.

Hagiu, A. and J. Wright (2016) “Controlling vs. Enabling” Harvard Business School working paper 16-002.

Holmstrom, B. (1977), “On Incentives and Control in Organizations,” Ph.D. Thesis, Stanford University.

Holmstrom, B. (1984), “On the Theory of Delegation”, in M. Boyer, and R. Kihlstrom (eds.) *Bayesian Models in Economic Theory* (New York: North-Holland), 115-141.

Jerath, K. and Z.J. Zhang (2010) “Store Within a Store,” *Journal of Marketing Research*, 47 (4), 748–763.

Johnson, J.P. (2013) "The agency and wholesale models in electronic content markets," Cornell University working paper.

Johnson, J.P. (2014) "The agency model and MFN clauses," Cornell University working paper.

Jullien, B. and P. Rey (2007) "Resale price maintenance and collusion," *Rand Journal of Economics*, 38, 983-1001.

Lal R. (1990) "Improving channel coordination through franchising," *Marketing Science*, 9 (4), 299–318.

Martimort, D. and A. Semenov (2006) "Continuity in mechanism design without transfers" *Economics Letters*, 93, 182-189

Melumad, N. and T. Shibano (1991), "Communication in Settings with No Transfers," *Rand Journal of Economics*, 22(2), 173-198.

Orbach, B.Y. and L. Einav (2007) "Uniform prices for differentiated goods: The case of the movie-theater industry," *International Review of Law and Economics*, 27(2), 129-153.

Parker, G. and M. W. Van Alstyne (2014) "Innovation, Openness, and Platform Control," working paper, http://ssrn.com/abstract=1079712.

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

Simon, H. A. (1951) "A Formal Theory of the Employment Relationship," *Econometrica*, 19(3), 293-305.

9 Appendix

9.1 Proof of Proposition 2

Since the principal fixes t in its contract at the same time as deciding on the type and nature of any delegation, we just have to show threshold delegation is optimal for any given t. To do so, we will show that any contract that differs from threshold delegation can be improved upon by a contract with the same t and threshold delegation.

The principal's delegation problem for a fixed choice of t is

$$\begin{aligned}
& \max_{D}\left\{\mathbb{E}\left[(1+b)\left(\beta a+\phi q+\Phi Q\right)-\frac{1}{2}a^2-\frac{1}{2}q^2-\frac{1}{2}Q^2\right]\right\} \\
& \text{subject to} \\
a &= \arg\max_{a'\in D}\left\{(1-t+b)\left(\beta a'+\phi q+\Phi Q\right)-\frac{1}{2}\left(a'\right)^2\right\} \\
q &= \phi\left(1-t+b\right) \\
Q &= \Phi t,
\end{aligned}$$

where D denotes the delegation set to which the principal restricts the agent's choice of a. Due to additive separability in a, q and Q, the program that defines the principal's optimal delegation set $D(t)$ can be re-written more simply as

$$\begin{aligned} & \max_D \left\{ \mathbb{E}\left[(1+b)\beta a - \frac{1}{2}a^2 \right] \right\} \\ & \text{subject to} \\ a \quad = \quad & \arg\max_{a' \in D} \left\{ (1-t+b)\beta a' - \frac{1}{2}(a')^2 \right\} \end{aligned}$$

We want to show that the delegation set $D(t)$ is a threshold interval, i.e. $D(t) = \{a \geq x(t)\}$ for some $x(t)$.

Our proof is based on a more general specification than that needed for Proposition 2. This will allow us to also use it for the proof of Proposition 8 in Section 6.1. Specifically, suppose the principal solves

$$\begin{aligned} & \max_D \mathbb{E}\left[-\alpha_0 a^2 + (\alpha_1 b + \alpha_2) a \right] \\ \text{subject to } a \quad = \quad & \arg\max_{a' \in D} \left\{ -\alpha_0 (a')^2 + (\alpha_1 b + \alpha_2 + \alpha_3) a' \right\} \end{aligned}$$

where α_0, α_1 and α_2 are all positive and we have ignored terms that do not depend on a. In this model, for a given b, the ideal choice of a for the principal is $a = \frac{\alpha_1 b + \alpha_2}{2\alpha_0}$, while for the agent it is $a = \frac{\alpha_1 b + \alpha_2 + \alpha_3}{2\alpha_0}$. We will show that if

$$\frac{|\alpha_3|}{\alpha_1} g'(b) \leq g(b) \tag{17}$$

for all $b \in [b_L, b_H]$, then the optimal delegation set D is a threshold interval. For the model in Section 4 we have $\alpha_0 = \frac{1}{2}$, $\alpha_1 = \beta$, $\alpha_2 = \beta$ and $\alpha_3 = -\beta t$, so that (17) is satisfied given the assumption in Proposition 2 and $0 \leq t \leq 1$.

Consider the case $\alpha_3 < 0$, so that without any restrictions, the agent prefers a lower choice of a than the principal. The principal is therefore interested in restricting the agent from setting a too low. The question remains whether the principal can do better by requiring the agent to choose from some specific values of a or some disjoint intervals that exclude some high values of a.

Suppose first the principal restricts the agent to choose a from some subset of $a \leq a_0 \equiv \frac{\alpha_1 b_0 + \alpha_2 + \alpha_3}{2\alpha_0}$ which includes $a = a_0$, where $b_0 < b_H$. This covers the possibility that the agent can only choose $a = a_0$ or can choose any $a \leq a_0$. In this case, when $b \in [b_0, b_H]$, it is easily seen that the agent chooses $a = a_0$ because the agent's objective function is increasing in a for all $a \leq a_0$. But the principal could strictly improve expected profits by adding the range $a \geq a_0$ to the set of permissible choices of a by the agent. To see this, note that the only change comes from the different choices of a by the agent when $b \in [b_0, b_H]$. The change in expected profits is

$$\begin{aligned} & \int_{b_0}^{b_H} \left(\begin{array}{c} -\alpha_0 \left(\frac{\alpha_1 b + \alpha_2 + \alpha_3}{2\alpha_0} \right)^2 + (\alpha_1 b + \alpha_2) \left(\frac{\alpha_1 b + \alpha_2 + \alpha_3}{2\alpha_0} \right) \\ -\left(-\alpha_0 \left(\frac{\alpha_1 b_0 + \alpha_2 + \alpha_3}{2\alpha_0} \right)^2 + (\alpha_1 b + \alpha_2) \left(\frac{\alpha_1 b_0 + \alpha_2 + \alpha_3}{2\alpha_0} \right) \right) \end{array} \right) dG(b) \\ = \quad & \frac{\alpha_1}{4\alpha_0} \int_{b_0}^{b_H} (b - b_0)\left((b - b_0)\alpha_1 - 2\alpha_3 \right) dG(b) > 0. \end{aligned}$$

Suppose now that the agent is allowed to choose from some set that does not include $a \in (a_0, a_1)$, where $a_0 \equiv \frac{\alpha_1 b_0 + \alpha_2 + \alpha_3}{2\alpha_0}$ and $a_1 \equiv \frac{\alpha_1 b_1 + \alpha_2 + \alpha_3}{2\alpha_0}$ for some $b_L \leq b_0 < b_1 \leq b_H$. In this case, since the agent's objective function is quadratic in a, if the agent's draw of b is in the range $[b_0, b_1]$, then the agent chooses $a = a_0$ when $b \leq \frac{b_1 + b_0}{2}$ and $a = a_1$ when $b > \frac{b_1 + b_0}{2}$. The principal can profitably deviate by adding the range $[a_0, a_1]$ to the set of permissible choices of a. Indeed, the change in profits is

$$\frac{\alpha_1}{4\alpha_0}\int_{b_0}^{\frac{b_1+b_0}{2}}(b-b_0)\left((b-b_0)\alpha_1-2\alpha_3\right)dG(b)+\frac{\alpha_1}{4\alpha_0}\int_{\frac{b_1+b_0}{2}}^{b_1}(b-b_1)\left((b-b_1)\alpha_1-2\alpha_3\right)dG(b)$$
$$=\frac{\alpha_1}{4\alpha_0}\left(\int_{b_0}^{\frac{b_1+b_0}{2}}\left(-2\alpha_3(b-b_0)+\alpha_1(b-b_0)^2\right)g(b)\,db+\int_{\frac{b_1+b_0}{2}}^{b_1}\left(-2\alpha_3(b-b_1)+\alpha_1(b-b_1)^2\right)g(b)\,db\right).$$

Using integration by parts, we have

$$\begin{aligned}\int_{b_0}^{\frac{b_1+b_0}{2}}\left(-2\alpha_3(b-b_0)+\alpha_1(b-b_0)^2\right)g(b)\,db &= -\alpha_3\frac{(b_1-b_0)^2}{4}g\left(\frac{b_1+b_0}{2}\right)\\ &\quad+\int_{b_0}^{\frac{b_1+b_0}{2}}(b-b_0)^2\left(\alpha_1 g(b)+\alpha_3 g'(b)\right)db\\ \int_{\frac{b_1+b_0}{2}}^{b_1}\left(-2\alpha_3(b-b_1)+\alpha_1(b-b_1)^2\right)g(b)\,db &= \alpha_3\frac{(b_1-b_0)^2}{4}g\left(\frac{b_1+b_0}{2}\right)\\ &\quad+\int_{\frac{b_1+b_0}{2}}^{b_1}(b-b_1)^2\left(\alpha_1 g(b)+\alpha_3 g'(b)\right)db.\end{aligned}$$

We can now plug these expressions into the last expression of the profit change above, which becomes equal to

$$\frac{\alpha_1}{4\alpha_0}\left(\int_{b_0}^{\frac{b_1+b_0}{2}}(b-b_0)^2\left(\alpha_1 g(b)+\alpha_3 g'(b)\right)db+\int_{\frac{b_1+b_0}{2}}^{b_1}(b-b_1)^2\left(\alpha_1 g(b)+\alpha_3 g'(b)\right)db\right).$$

This expression is clearly positive under the assumption $-\frac{\alpha_3}{\alpha_1}g'(b)\leq g(b)$ for all b.

Thus, we can conclude that the optimal range of admissible a for the agent must take the form of a threshold interval $a\geq x$. Since $\alpha_3=-\beta t<0$ in the model of Section 4, this completes the proof of Proposition 2.

If on the other hand $\alpha_3>0$, then the same proof applies, except that now the optimal range of admissible a for the agent must take the form of a threshold interval with a maximum requirement $a\leq x$.

9.2 Proof of Proposition 3

We first compare H-mode to A-mode. Set $t^H=t^{A*}>0$ and $x=\beta\left(1-t^{A*}+b_L+\kappa\right)$, where κ is very small. In other words, we choose the same variable fee as in the A-mode and give the agent almost the same discretion as in A-mode, but set the minimum threshold to eliminate the agent's lowest possible choices of a (which occur for the lowest values of b). Then the H-mode profit is strictly higher than the A-mode profit. To see this, note that

$$\begin{aligned}&\Pi^H\left(t^{A*},x=\beta\left(1-t^{A*}+b_L+\kappa\right)\right)-\Pi^{A*}\\ =\;&\Pi^H\left(t^{A*},x=\beta\left(1-t^{A*}+b_L+\kappa\right)\right)-\Pi^H\left(t^{A*},x=\beta\left(1-t^{A*}+b_L\right)\right)\\ =\;&\int_{b_L}^{b_L+\kappa}\beta^2\begin{pmatrix}(1+b)\left(1-t^{A*}+b_L+\kappa\right)-(1+b)\left(1-t^{A*}+b\right)\\ +\frac{1}{2}\left(1-t^{A*}+b\right)^2-\frac{1}{2}\left(1-t^{A*}+b_L+\kappa\right)^2\end{pmatrix}dG(b)\\ =\;&\int_{b_L}^{b_L+\kappa}\beta^2\begin{pmatrix}(1+b)(b_L+\kappa-b)\\ -\frac{1}{2}(b_L+\kappa-b)\left(2\left(1-t^{A*}\right)+b_L+\kappa+b\right)\end{pmatrix}dG(b)\\ =\;&\int_{b_L}^{b_L+\kappa}\frac{\beta^2}{2}(b_L+\kappa-b)\left(2t^{A*}\beta+b-b_L-\kappa\right)dG(b)>0\end{aligned}$$

for κ sufficiently small, because $t^{A*} > 0$. Thus,

$$\Pi^{H*} \geq \Pi^{H}\left(t^{A*}, x = \beta\left(1 - t^{A*} + b_L + \kappa\right)\right) > \Pi^{A*}.$$

This means H-mode strictly dominates A-mode.

Next we compare H-mode to P-mode. Consider the H-mode with $t = t^{P*} = \frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2}$ and $x = a^{P*} = \beta\left(1+\bar{b}\right)$. Then (t, x) is strictly interior because

$$b_L < \frac{a^{P*}}{\beta} - \left(1 - t^{P*}\right) < b_H,$$

where the second inequality follows from the assumption $b_H > \bar{b} + \frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2}$. We then have

$$\begin{aligned}
\Pi^{H}\left(t^{P*}, a^{P*}\right) - \Pi^{P*} &= \int_{\frac{a^{P*}}{\beta}-(1-t^{P*})}^{b_H} \beta^2\left((1+b)\left(1-t^{P*}+b\right) - \frac{1}{2}\left(1-t^{P*}+b\right)^2\right) dG(b) \\
&\quad + \int_{b_L}^{a^{P*}-\left(1-t^{P*}\right)\beta} \left(\beta(1+b)\, a^{P*} - \frac{1}{2}\left(a^{P*}\right)^2\right) dG(b) \\
&\quad - \int_{b_L}^{b_H} \left(\beta(1+b)\, a^{P*} - \frac{1}{2}\left(a^{P*}\right)^2\right) dG(b) \\
&= \int_{\frac{a^{P*}}{\beta}-(1-t^{P*})}^{b_H} \beta^2\left((1+b)\left(-t^{P*}+b-\bar{b}\right) + \frac{1}{2}\left(t^{P*}-b+\bar{b}\right)\left(2-t^{P*}+b+\bar{b}\right)\right) dG(b) \\
&= \int_{t^{P*}+\bar{b}}^{b_H} \frac{\beta^2}{2}\left(b - t^{P*} - \bar{b}\right)\left(b + t^{P*} - \bar{b}\right) dG(b) > 0.
\end{aligned}$$

This implies H-mode strictly dominates P-mode.

9.3 Proof of Proposition 5

From $\frac{\partial \Pi^{H}(t,x)}{\partial t} = 0$, we obtain

$$t^{H*} = \frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2 + \phi^2 + t^{H*}\beta^2\left(1 - G\left(\frac{x^*}{\beta} - \left(1 - t^{H*}\right)\right)\right)}.$$

Recall that

$$t^{P*} = \frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2} \text{ and } t^{A*} = \frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2+\beta^2}.$$

Since $\left(t^{H*}, x^*\right)$ is assumed to be interior, we can conclude $t^{A*} < t^{H*} < t^{P*}$.

For the second part of the proposition, assume the second-order conditions are verified, so we can use the first-order conditions to characterize $\left(t^{H*}, x^*\right)$. Note that $\frac{\partial \Pi^{H}\left(t^{H*}, x^*\right)}{\partial x} = 0$ can be rewritten as

$$x^* = \beta\left(1 + \mathbb{E}\left[b \middle| b < \frac{x^*}{\beta} - \left(1 - t^{H*}\right)\right]\right).$$

Since $\left(t^{H*}, x^*\right)$ is interior, we have $\frac{x^*}{\beta} - \left(1 - t^{H*}\right) < b_H$ and so $\mathbb{E}\left[b|b < \frac{x^*}{\beta} - \left(1 - t^{H*}\right)\right] < \bar{b}$. This implies $x^* < \beta\left(1+\bar{b}\right)$.

9.4 Proof of Proposition 6

Using that $t^{A*} < t^{H*} < t^{P*}$ from Proposition 5, we have

$$\begin{aligned}\frac{d\Pi^{A*}}{d\Phi^2} &= \frac{t^{A*}\left(2\left(1+\bar{b}\right)-t^{A*}\right)}{2} < \frac{d\Pi^{H*}}{d\Phi^2} = \frac{t^{H*}\left(2\left(1+\bar{b}\right)-t^{H*}\right)}{2} < \frac{d\Pi^{P*}}{d\Phi^2} = \frac{t^{P*}\left(2\left(1+\bar{b}\right)-t^{P*}\right)}{2} \\ \frac{d\Pi^{A*}}{d\phi^2} &= \frac{\left(1+\bar{b}\right)^2+V_b-\left(t^{A*}\right)^2}{2} > \frac{d\Pi^{H*}}{d\phi^2} = \frac{\left(1+\bar{b}\right)^2+V_b-\left(t^{H*}\right)^2}{2} > \frac{d\Pi^{P*}}{d\phi^2} = \frac{\left(1+\bar{b}\right)^2+V_b-\left(t^{H*}\right)^2}{2}.\end{aligned}$$

To show the second part of the proposition, note that t^{H*} is defined by $f\left(t^{H*},\Phi^2,\phi^2\right)=0$, where

$$f\left(t,\Phi^2,\phi^2\right) \equiv \Phi^2\left(1+\bar{b}\right) - t\left(\Phi^2+\phi^2\right) - t\beta^2\left(1-G\left(\frac{x\left(t\right)}{\beta}-\left(1-t\right)\right)\right),$$

while $x\left(t\right)$ is defined implicitly by

$$h\left(x,t\right) \equiv \int_{b_L}^{\frac{x}{\beta}-(1-t)} \left(\beta\left(1+b\right)-x\right) dG\left(b\right) = 0.$$

For t^{H*} to maximize $\Pi^H\left(t,x\left(t\right)\right)$, it must be that $\frac{\partial f\left(t=t^{H*},\Phi^2,\phi^2\right)}{\partial t} < 0$. Combined with $\frac{\partial f\left(t=t^{H*},\Phi^2,\phi^2\right)}{\partial \Phi^2} = 1+\bar{b}-t^{H*} > 0$ and $\frac{\partial f\left(t=t^{H*},\Phi^2,\phi^2\right)}{\partial \phi^2} = -t^{H*} < 0$, by the implicit function theorem, this implies that t^{H*} is increasing in Φ and decreasing in ϕ. For x^* to maximize $\Pi^H\left(t^{H*},x\right)$, it must be that $\frac{\partial h\left(x=x^*,t=t^{H*}\right)}{\partial x} < 0$. Combined with $\frac{\partial h(x,t)}{\partial t} > 0$, this implies that $x\left(t\right)$ is increasing around $t=t^{H*}$. Since $x^* = x\left(t^{H*}\right)$, this means that x^* is also increasing in Φ and decreasing in ϕ.

9.5 Proof of Proposition 7

When $G\left(.\right)$ is the uniform distribution, the first-order conditions for interior $\left(t,x\right)$ are

$$\frac{\partial \Pi^H\left(t,x\right)}{\partial x} = \frac{\beta}{4\sigma}\left(1+t+\bar{b}-\sigma-\frac{x}{\beta}\right)\left(\frac{x}{\beta}-\left(1-t\right)-\bar{b}+\sigma\right) = 0 \tag{18}$$

$$\frac{\partial \Pi^H\left(t,x\right)}{\partial t} = \Phi^2\left(1+\bar{b}\right) - t\left(\Phi^2+\phi^2\right) - \frac{t\beta^2}{2\sigma}\left(1-t+\bar{b}+\sigma-\frac{x}{\beta}\right) = 0. \tag{19}$$

Suppose the solution $\left(t^{H*},x^*\right)$ to (18)-(19) is interior, i.e.

$$\bar{b}-\sigma < \frac{x^*}{\beta} - \left(1-t^{H*}\right) < \bar{b}+\sigma. \tag{20}$$

The left-hand side inequality in (20) implies that $\frac{x}{\beta}-\left(1-t\right)-\bar{b}+\sigma > 0$ in (18), so (18)-(19) imply

$$\beta\left(1+t^{H*}+\bar{b}-\sigma\right) = x^* \tag{21}$$

$$\Phi^2\left(1+\bar{b}\right) - t^{H*}\left(\Phi^2+\phi^2\right) - \frac{t^{H*}\beta^2}{2\sigma}\left(1-t^{H*}+\bar{b}+\sigma-\frac{x^*}{\beta}\right) = 0. \tag{22}$$

The right-hand side inequality in (20) is then equivalent to $t^{H*} < \sigma$.

Substituting x^* from (21) into (22), we obtain

$$\Phi^2\left(1+\bar{b}\right)-t^{H*}\left(\Phi^2+\phi^2+\beta^2\right)+\frac{\left(t^{H*}\right)^2\beta^2}{\sigma}=0. \tag{23}$$

This equation has real number solutions for t^{H*} if and only if $\sigma>\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{(\Phi^2+\phi^2+\beta^2)^2}$.

If $\sigma\leq\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{(\Phi^2+\phi^2+\beta^2)^2}$, then the function $f(t)\equiv\Pi^H\left(t,x=\beta\left(1+t+\bar{b}-\sigma\right)\right)$ is weakly increasing in t for all $t\in[0,\sigma]$, so the optimal $\left(t^{H*},x^*\right)$ is non-interior and is such that $\frac{x^*}{\beta}-\left(1-t^{H*}\right)\geq\bar{b}+\sigma$. This also means the H-mode is dominated by the P-mode.

Assume now $\sigma>\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{(\Phi^2+\phi^2+\beta^2)^2}$ and denote the two solutions to equation (23) by

$$t_1 \equiv \frac{\sigma}{2\beta^2}\left(\Phi^2+\phi^2+\beta^2-\sqrt{\left(\Phi^2+\phi^2+\beta^2\right)^2-\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{\sigma}}\right)$$

$$t_2 \equiv \frac{\sigma}{2\beta^2}\left(\Phi^2+\phi^2+\beta^2+\sqrt{\left(\Phi^2+\phi^2+\beta^2\right)^2-\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{\sigma}}\right)>t_1.$$

In this case, the function $f(t)$ is increasing for $t\in[0,t_1]$, decreasing for $t\in[t_1,t_2]$ and increasing again for $t\geq t_2$. Thus, the only candidate interior solution is $t^{H*}=t_1$. This solution is indeed interior if and only if $t_1<\sigma$, which is equivalent to

$$\Phi^2+\phi^2-\beta^2<\sqrt{\left(\Phi^2+\phi^2+\beta^2\right)^2-\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{\sigma}}.$$

The last inequality is in turn equivalent to $\beta^2>\Phi^2+\phi^2$ or ($\beta^2\leq\Phi^2+\phi^2$ and $\sigma>\frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2}$). Combining this with the requirement that $\sigma>\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{(\Phi^2+\phi^2+\beta^2)^2}$ and noting that $\frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2}>\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{(\Phi^2+\phi^2+\beta^2)^2}$ for all parameter values, we obtain that the optimal solution in H-mode is interior if and only if ($\frac{4\Phi^2\beta^2\left(1+\bar{b}\right)}{(\Phi^2+\phi^2+\beta^2)^2}<\sigma<\frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2}$ and $\beta^2>\Phi^2+\phi^2$) or $\sigma>\frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2}$. Finally, recall that we must also have $\sigma\leq\bar{b}$. Since the assumption $\bar{b}>\frac{\Phi^2}{\phi^2}$ implies $\frac{\Phi^2\left(1+\bar{b}\right)}{\Phi^2+\phi^2}<\bar{b}$, we obtain the condition in the text of the proposition.

We have confirmed the Hessian is negative semidefinite, which ensures the local concavity of the unique interior solution $\left(t^{H*},x^*\right)$ determined above, whenever it exists.

To derive the effect of σ on the optimal solutions $\left(t^{H*},x^*\right)$, recall that t^{H*} is defined by

$$f\left(t,\Phi^2,\beta,\sigma\right)=0,\text{ where } f\left(t,\Phi^2,\beta,\sigma\right)\equiv\Phi^2\left(1+\bar{b}\right)-t\left(\Phi^2+\phi^2+\beta^2\right)+\frac{t^2\beta^2}{\sigma}.$$

Clearly, $\frac{\partial f\left(t^{H*},\Phi^2,\beta,\sigma\right)}{\partial\sigma}<0$, so $\frac{dt^{H*}}{d\sigma}<0$. Since $x^*=\beta\left(1+t^{H*}+\bar{b}-\sigma\right)$, we also have $\frac{dx^*}{d\sigma}=\beta\left(\frac{dt^{H*}}{d\sigma}-1\right)<0$.

We now turn to the second part of the proposition. Assuming the optimal solution in H-mode is interior, the principal's profits can be written in the following way:

$$\Pi^{H*}=\max_{t,x}\Pi^H(t,x)=\max_t\Pi^H\left(t,x=\beta\left(1+t+\bar{b}-\sigma\right)\right).$$

We then have

$$\Pi^H\left(t,\beta\left(1+t+\bar{b}-\sigma\right)\right)=\Pi^H\left(t,\beta\left(1-t+\bar{b}+\sigma\right)\right)-\int_{\beta\left(1+t+\bar{b}-\sigma\right)}^{\beta\left(1-t+\bar{b}+\sigma\right)}\frac{\partial\Pi^H\left(t,x\right)}{\partial x}dx$$

$$\begin{aligned} =\ & \frac{\Phi^2}{2}t\left(2\left(1+\bar{b}\right)-t\right)+\frac{\phi^2}{2}\left(\left(1+\bar{b}\right)^2+\frac{\sigma^2}{3}-t^2\right)+\beta^2\left(\left(1+\bar{b}\right)\left(1-t+\bar{b}+\sigma\right)-\frac{1}{2}\left(1-t+\bar{b}+\sigma\right)^2\right) \\ & -\int_{\beta\left(1+t+\bar{b}-\sigma\right)}^{\beta\left(1-t+\bar{b}+\sigma\right)}\frac{1}{4\sigma\beta}\left(t^2\beta^2-\left(\beta\left(1+\bar{b}-\sigma\right)-x\right)^2\right)dx \end{aligned}$$

$$\begin{aligned} =\ & \frac{\Phi^2}{2}t\left(2\left(1+\bar{b}\right)-t\right)+\frac{\phi^2}{2}\left(\left(1+\bar{b}\right)^2+\frac{\sigma^2}{3}-t^2\right)+\beta^2\left(\frac{1}{2}\left(1+\bar{b}\right)^2+\frac{1}{6}\sigma^2-\frac{1}{2}t^2+\frac{1}{3\sigma}t^3\right) \\ \equiv\ & \widetilde{\Pi}^H\left(t,\sigma\right). \end{aligned}$$

Consequently, we have $\Pi^{H*}=\max_t\widetilde{\Pi}^H\left(t,\sigma\right)$. We can then use the envelope theorem to obtain

$$\frac{d\Pi^{H*}}{d\sigma}=\frac{d\widetilde{\Pi}^H\left(t=t^{H*},\sigma\right)}{d\sigma}=\frac{\partial\widetilde{\Pi}^H\left(t=t^{H*},\sigma\right)}{\partial\sigma}=\phi^2\frac{\sigma}{3}+\frac{\beta^2}{3\sigma^2}\left(\sigma^3-\left(t^{H*}\right)^3\right).$$

Since $t^{H*}<\sigma$, this implies

$$\frac{d\Pi^{P*}}{d\sigma}=\phi^2\frac{\sigma}{3}<\frac{d\Pi^{H*}}{d\sigma}<\left(\phi^2+\beta^2\right)\frac{\sigma}{3}=\frac{d\Pi^{A*}}{d\sigma}.$$

9.6 Proof of Proposition 8

Since the principal fixes w in its contract at the same time as deciding on the type and nature of any delegation, we just have to show threshold delegation is optimal for any given w. The principal's delegation problem for a fixed choice of w is

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

which can be rewritten as

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

Ignoring terms that do not depend on p, the program that defines the principal's optimal delegation set $D\left(t\right)$ can be re-written more simply as

$$\max_D \mathbb{E}\left[-\alpha_0 p^2 + (\alpha_1\theta + \alpha_2)\, p\right]$$
$$\text{subject to } p = \arg\max_{p' \in D}\left\{-\alpha_0 (p')^2 + (\alpha_1\theta + \alpha_2 + \alpha_3)\, p'\right\},$$

where $\alpha_0 = \beta - \frac{\phi^2}{2}$ (recall this is assumed to be positive), $\alpha_1 = 1$, $\alpha_2 = w\Phi^2$ and $\alpha_3 = w\left(\beta - \phi^2\right)$. Note that α_3 can be positive or negative. Then applying the proof of Proposition 2 to this specification, we know that threshold delegation is optimal provided

$$w\left|\beta - \phi^2\right| g'(\theta) \leq g(\theta). \tag{24}$$

Since $w > 0$ and $g(\theta) > 0$ for all $\theta \in [\theta_L, \theta_H]$, a sufficient condition for (24) to hold is simply that $g'(\theta) \leq 0$.

9.7 Proof of Proposition 9

Substituting (11)-(12) into (10), we find in P-mode that

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

Similarly, substituting (14)-(16) into (13), we find in A-mode that

$$w^{A*} = \frac{\overline{\theta}\Phi^2}{\beta(2\Phi^2 + \beta) - \Phi^2(\phi^2 + \Phi^2)} \text{ and } \Pi^{A*} = \frac{\overline{\theta}^2\left(\beta\left(2\Phi^2 + \beta\right) - \Phi^2\phi^2\right)}{2(2\beta - \phi^2)\left(\beta(2\Phi^2 + \beta) - \Phi^2(\phi^2 + \Phi^2)\right)} + \frac{V_\theta}{2(2\beta - \phi^2)}.$$

We consider the H-mode for the two cases in the proposition.

9.7.1 Case (i): $\beta > \phi^2$

From (14), the agent is not constrained in its choice of price for a given w if and only if

$$\theta < \left(2\beta - \phi^2\right)x - w\left(\Phi^2 - \phi^2 + \beta\right).$$

The agent therefore sets

$$p(w, x) = \begin{cases} \frac{\theta + \left(\Phi^2 - \phi^2 + \beta\right)w}{2\beta - \phi^2} = p^A(w) & \text{if} \quad \theta < \left(2\beta - \phi^2\right)x - w\left(\Phi^2 - \phi^2 + \beta\right) \\ x & \text{if} \quad \theta \geq \left(2\beta - \phi^2\right)x - w\left(\Phi^2 - \phi^2 + \beta\right). \end{cases}$$

The principal's profit is then

$$\Pi^H(w, x) = \int_{\theta_L}^{\left(2\beta - \phi^2\right)x - w\left(\Phi^2 - \phi^2 + \beta\right)} \left(p^A(w)\left(\theta - \beta p^A(w) + w\Phi^2\right) + \frac{\phi^2}{2} p^A(w)^2\right) dG(\theta)$$
$$+ \int_{(2\beta - \phi^2)x - w(\Phi^2 - \phi^2 + \beta)}^{\theta_H} \left(x\left(\theta - \beta x + w\Phi^2\right) + \frac{\phi^2}{2}x^2\right) dG(\theta) - \frac{\Phi^2 + \phi^2}{2} w^2.$$

We first compare H-mode to A-mode. Set $w^H = w^{A*}$ and $x = \frac{\theta_H + w^{A*}\left(\Phi^2 - \phi^2 + \beta\right)}{2\beta - \phi^2} - \frac{\kappa}{2\beta - \phi^2}$, for some small

$\kappa > 0$. Following the same steps as in the proof of Proposition 3, we get that

$$\begin{aligned}
&\Pi^H\left(w^{A*}, x = \frac{\theta_H + w^{A*}\left(\Phi^2 - \phi^2 + \beta\right)}{2\beta - \phi^2} - \frac{\kappa}{2\beta - \phi^2}\right) - \Pi^{A*} \\
= &\int_{\theta_H - \kappa}^{\theta_H} \left[\left(\frac{1}{2}\left(p^A\left(w^{A*}, \theta\right) - p^A\left(w^{A*}, \theta_H - \kappa\right)\right)\left(2w^{A*}\left(\beta - \phi^2\right) + \theta_H - \theta - \kappa\right)\right)\right] dG\left(\theta\right) > 0,
\end{aligned}$$

for κ sufficiently small because $p^A\left(w, \theta\right) > p^A\left(w, \theta_H - \kappa\right)$ for all w and $\beta > \phi^2$. Thus,

$$\Pi^{H*} \geq \Pi^H\left(w^{A*}, x = \frac{\theta_H + w^{A*}\left(\Phi^2 - \phi^2 + \beta\right)}{2\beta - \phi^2} - \frac{\kappa}{2\beta - \phi^2}\right) > \Pi^{A*}.$$

This implies H-mode dominates A-mode.

Next we compare H-mode to P-mode. Consider the H-mode with $w = w^{P*} = \frac{\overline{\theta}\Phi^2}{(2\beta-\phi^2)(\phi^2+\Phi^2)-\Phi^4}$ and $x = p^{P*} = \frac{\overline{\theta}+\Phi^2 w^{P*}}{2\beta-\phi^2}$. Then (w, x) is strictly interior because

$$\theta_L < \left(2\beta - \phi^2\right) p^{P*} - w^{P*}\left(\Phi^2 - \phi^2 + \beta\right) = \overline{\theta} - w^{P*}\left(\beta - \phi^2\right) < \theta_H,$$

where the first inequality follows from the assumption $\theta_L < \overline{\theta}\left(1 - \frac{\Phi^2\left(\beta-\phi^2\right)}{(2\beta-\phi^2)(\phi^2+\Phi^2)-\Phi^4}\right)$ and the second inequality from $\beta > \phi^2$. Then following the same steps as in the proof of Proposition 3, we obtain

$$\Pi^H\left(w^{P*}, p^{P*}\right) - \Pi^{P*} = \int_{\theta_L}^{\overline{\theta} - w^{P*}\left(\beta - \phi^2\right)} \frac{1}{2\left(2\beta - \phi^2\right)}\left(\overline{\theta} - \theta - w^{P*}\left(\beta - \phi^2\right)\right)\left(\overline{\theta} - \theta + w^{P*}\left(\beta - \phi^2\right)\right) dG\left(\theta\right) > 0,$$

which implies H-mode strictly dominates P-mode.

9.7.2 Case (ii): $\beta < \phi^2$

Using the same steps as in the previous case, we can show that

$$\Pi^{H*} \geq \Pi^H\left(w^{A*}, x = \frac{\theta_L + w^{A*}\left(\Phi^2 - \phi^2 + \beta\right)}{2\beta - \phi^2} + \frac{\kappa}{2\beta - \phi^2}\right) > \Pi^{A*}$$

for κ sufficiently small, which implies that H-mode dominates A-mode.

And similarly, we can show that

$$\Pi^H\left(w = w^{P*} = \frac{\overline{\theta}\Phi^2}{\left(2\beta - \phi^2\right)\left(\phi^2 + \Phi^2\right) - \Phi^4}, x = p^{P*} = \frac{\overline{\theta} + \Phi^2 w^{P*}}{2\beta - \phi^2}\right) > \Pi^{P*},$$

where $\left(w = w^{P*}, x = p^{P*}\right)$ is strictly interior because

$$\theta_L < \left(2\beta - \phi^2\right) p^{P*} - w^{P*}\left(\Phi^2 - \phi^2 + \beta\right) = \overline{\theta} + w^{P*}\left(\phi^2 - \beta\right) < \theta_H$$

due to the assumptions $\phi^2 > \beta$ and $\theta_H > \overline{\theta}\left(1 + \frac{\Phi^2\left(\phi^2-\beta\right)}{(2\beta-\phi^2)(\phi^2+\Phi^2)-\Phi^4}\right)$. Thus, H-mode strictly dominates P-mode under these assumptions.