Franchising as a Nexus of Incentive Devices for Brand Name Maintenance

Chong-en Bai[*] and Zhigang Tao[†]

December 1999

[*]School of Economics and Finance, The University of Hong Kong, Pokfulam Road, Hong Kong. Telephone: (852) 2859-1036; Fax: (852) 2548-1152; email: baic@hku.hk

[†]School of Business, The University of Hong Kong, Pokfulam Road, Hong Kong. Telephone: (852) 2857-8223; Fax: (852) 2858-5614; email: ztao@hku.hk

Abstract

Franchising involves a variety of contractual and ownership arrangements within a single company. In recent years, a great deal of effort has been made to understand this increasingly popular organization (Lafontaine, 1992, 1993, and Lafontaine and Slade, 1997). There are at least two stylized facts that have posed challenges to the existing theories of the firm. One is the co-existence of company-owned and franchised units, and the other is that franchisees make substantial amount of investment highly specific to their franchise companies.

We set out to explain both puzzles based on the importance of the brand name in franchising (Kaufmann and Lafontaine, 1994). The effort to develop and maintain the brand name changes over time and is difficult to verify (Hadfield, 1990), which has two implications. One is that agents who run franchise units need to be given appropriate incentives for the brand-name-maintenance effort. The other is that franchising contracts are incomplete. For incentive purposes, it is optimal to divide the agents into two groups. Those in the first group (managers of company owned units) receive a salary and focus on brand maintenance. Those in the second group (franchisees) receive a share of the revenue in their own unit and focus mainly on unit specific sales effort (Bai and Tao, 2000). However, the franchisees should also be subject to a minimum service standard that is crucial for brand name maintenance. The high-powered incentive for the franchisees to increase sales revenue implies that they have a strong tendency to divert effort from meeting the minimum standard. To discourage the franchisees from doing so, they should be subject to severe penalty when found violating the standard. We show that, to serve this purpose, it is optimal to have the franchisees make investment highly specific to their franchise companies.

Specifically, the investment by the franchisee to buy physical assets (buildings, equipment, etc.) can be viewed as a performance bond for the minimum standard. If the franchisor controls the assets when the franchisee leaves the company, then the franchisor has an incentive to opportunistically accuse the franchisee of violating the standard and fire the franchisee, getting all the profits arising from the assets. If the franchisee controls the assets, such opportunistic behavior of the franchisor will not occur. Furthermore, if the assets are relationship specific so that their value is very low when detached from the brand name, then the franchisee will have strong incentive not to violate the minimum standard, fearing of being deprived the right to use the brand name in the event of violation.

Overall, the plural forms of contractual and control right arrangements in franchising serve as a nexus of incentive devices for production involving brand-name-maintenance effort in an incomplete contract framework.

Field: Contract Theory, Theory of the Firm

1. Introduction

As arguably one of the most important organizational innovations of the last half century, the franchise has motivated and challenged recent developments in the theory of the ¯rm.[1] A franchise company typically has both company-owned units and franchised units, which di®er in at least two important aspects: (1) Managers in the company-owned units have low-powered incentive contracts, that is, they receive ¯xed wages. Managers in the franchised units, however, have high-powered incentive contracts; they pay a portion of their revenue as a royalty to the company and keep the remainder. (2) Managers in the company-owned units do not own any assets of the units, whereas managers in the franchised units own part or all of the units' physical assets.[2] To explain such intra¯rm heterogeneity of organizational form, we study a multi-agent, multi-task model where one of the tasks generates a public good | the brand name in the context of the franchise. We show that it is optimal to o®er a mix of high-powered and low-powered incentives to ex ante homogeneous agents. In addition, corresponding ownership arrangements are chosen to make the revenue-sharing agreements self-enforcing.

This study is inspired by the history of the development of the business format franchising in the United States. The e±ciency and success of the McDonald brothers' California drive-in made it possible for Ray Kroc (the founder of the McDonald's company) to sell the brand name and the successful business format to would-be entrepreneurs (Love (1986)). Tom Monaghan, the founder of Domino's Pizza, started selling the business of making pizza only after years of perfecting the production process (Dicke (1992)). Indeed, in one of the earliest litigations about business format franchising (Susser v. Carrel), the courts realized that \the cornerstone of the franchise system must be the trademark or trade name of a product."[3] Clearly, what is crucial to a business format franchise is the e®ort to develop and maintain brand name products and services shared by all units of the company (henceforth, goodwill), as well as the e®ort in production and distribution (henceforth, sales).

We consider a company that consists of the headquarters (HQ) and many units. The

[1] The multi-task theory of the ¯rm (Holmstrom and Milgrom (1991, 1994)) rationalizes incentive contracts of various power, while the incomplete-contracts theory of the ¯rm (Grossman and Hart (1986), Hart and Moore (1990), Hart (1995)) focuses on ownership structures.

[2] In 1986, the percentage of franchised units was 76 percent in McDonald's, and 82 percent in Burger King (Lafontaine (1992)). Like most fast food chains, these are examples of the business format franchise, which includes business format as well as (usually generic) goods and services. The other type is the product franchise (such as car dealers) which involves complicated products and services (Dicke (1992)). This paper focuses on the business format franchise, as the aforementioned heterogeneity in organizational form is sharper in the business format franchise than in the product franchise. Furthermore, both the number of the outlets of the business format franchisors and the total nominal sales through them had phenomenal growth between 1972 and 1986 (Lafontaine (1992)).

[3] Case studies reveal that the main reason for franchising is to o®er perspective franchisees a strong brand under which to operate (Dnes (1992)). It has been found that franchised units su®ered signi¯cant loss of revenue when their brand names were revoked (Kaufmann and Lafontaine (1994)).

manager of each unit performs two possible tasks, goodwill and sales.[4] If a manager's payoff is tied to the revenue of her unit with a high-powered incentive contract, she will allocate too little of her effort on goodwill and too much of it on sales. This is because goodwill is a company-wide feature and thus the manager can free ride on the goodwill provided by other units, whereas the sales effort is unit specific and the manager cannot rely on other units for its provision. In contrast, if given a low-powered contract, then she is indifferent to the allocation of her effort between the two tasks and thus is willing to expend some goodwill effort to the extent that it does not bring about disutility.[5] Therefore, in order to induce a goodwill effort, the company offers some of its managers low-powered contracts, despite these contracts' adverse implications on overall effort level. With the goodwill provided by these units, sales effort becomes more important in the remaining units of the company, and high-powered incentive contracts are thus optimal. In a word, contract mix allows the HQ to induce production of goodwill (public good) and at the same time capture its positive externality within the company.

It is well documented that company-owned units are better than franchised units in terms of quality of services and adherence to uniform standards.[6] In addition, most (if not all) franchisors require their units not to reveal to the public the ownership status (i.e., franchised or company-owned). These facts directly support our explanation of contract mix in franchise. Moreover, our model can reconcile the existence of company-owned units with the empirical finding that the profit margin under franchise ownership is much higher than that under company ownership (Shelton (1967)). With the low-powered incentive contracts, managers of company-owned units provide goodwill effort for the whole company at the expense of sales effort that would enhance their own profitability. Finally, when ex ante heterogeneity is properly modeled, our explanation is consistent with the empirical finding that units located along highways are more likely to be franchised (Brickley and Dark (1987)).

With contract mix understood, it remains to be explained why managers with the high-powered contracts own part or all of the units' physical assets whereas those with the low-powered contracts do not. Our analysis on this issue is inspired by a stylized fact of the franchise, namely, the incompleteness of franchise contracts (Hadfield (1990)).[7] When ex ante contracts are incomplete about development and maintenance of brand name products and services, they may become unenforceable ex post.

[4] For a business format franchise, goodwill has to be generated by managers who also engage in sales activities. In contrast, for a product franchise, brand names of complicated goods and services could be established independently. See Section 2 for discussion.

[5] Following Holmstrom and Milgrom (1991), we assume that the manager may exert effort up to some limit without explicit incentives, and incentives are only required to encourage work beyond that limit. See Footnote 12 of Section 2 for discussion.

[6] For example, Love (1986) documents that there were problems of quality and cleanliness in McDonald's franchised units (Chapter 4) and that the company-owned units were set up to "encourage wayward McDonald's franchisees to clean up their act" (Chapter 9). Also see Lewin (1996).

[7] For example, "many of the standards with which a McDonald's franchisee must comply will not even be articulated until well after the contract has been signed" (Hadfield (1990)).

From MacLeod and Malcomson (1993, 1996), we know that an ex ante contract could become self-enforcing if the HQ and the manager have appropriate outside options. In contrast to MacLeod and Malcomson's reliance on contractual remedies for outside options, we focus on such roles of ownership arrangements. When the manager owns the asset, she can deny the HQ access to it. Then, the HQ's outside option is to conduct business with only the brand name, while the manager's is to provide generic goods and services. When the HQ owns the asset, it can deny the manager access to both the brand name and the asset. The manager thus does not have any outside option, while the HQ has the outside option of capturing all the revenue. We show that, for their contracts to be self-enforcing, managers with the high-powered contracts need to own their units' physical assets whereas those with the low-powered contracts should not. We further show that, in our setting, optimally chosen contractual remedies cannot mimic what the ownership arrangements have achieved.

It is generally held that, because the value of franchisees' assets depends crucially on the access to their company brand names, franchisees are extremely vulnerable to franchisors' opportunistic behavior. However, empirical studies by Kostecka (1987) found that in 1985 franchisors terminated 2,651 units, which equals only .87 percent of the estimated 301,689 units of business-format franchise companies existing then. It is interesting to note that this puzzle could be easily explained by our theory of ownership structures in the franchise.

This paper is built upon the multi-task theory and incomplete-contracts theory of the firm. Compared with Holmstrom and Milgrom (1991, 1994), this paper incorporates the possibility that one of the many tasks is of public good nature. Heterogeneity in task importance is then endogenously determined, and contract mix is optimal even for ex ante homogeneous units of the company.[8] Compared with Grossman and Hart (1986) and Hart and Moore (1990), this paper considers revenue-sharing contracts in the setting of contractual incompleteness and explores other roles of ownership. In particular, the optimal ownership structure is chosen to make the ex ante contracts self-enforcing rather than to provide direct ex ante effort incentive. Furthermore, the optimality of the contract mix implies multiple ownership arrangements of complementary assets.[9]

Gallini and Lutz (1992) also offer an explanation of contract mix in franchise. Their signalling theory that the franchisor owns some units to reveal its private information about profitability is probably more relevant to new franchises (Lafontaine (1993) and Lafontaine and Shaw (1996)). Lutz (1995) attempts to explain the ownership structures in a franchise by allowing only short-term contracts. She shows that managers with low-powered contracts should own their units, while those with high-powered contracts should not.

The plan of the paper is as follows. In Section 2, we introduce a multi-task model where there is no contractual incompleteness. In Section 3, the HQ's contract design problem is

[8] When all tasks are of private good nature, optimality of contract mix depends crucially on ex ante heterogeneous task importance (Holmstrom and Milgrom (1991, 1994)).

[9] Hart, Shleifer, and Vishny (1996) discuss incomplete contracts with multiple tasks in a different context. In particular, they exclude the possibility of revenue-sharing contracts, and focus on how ownership structure affects the allocation of an agent's attention among her various tasks.

studied, and the optimality of the contract mix is established. In Section 4, by incorporating contractual incompleteness into our model, we show that incentive contracts of various power can be self-enforcing only in the presence of corresponding ownership structures as observed in a franchise. The paper concludes with Section 5. All proofs are in the appendix.

2. The Model

2.1. Production technology

Consider a company (or chain) that consists of the headquarters (HQ) and many units. To highlight the free-rider problem, we assume that there are infinite identical units. Furthermore, for simplicity, we assume that the units are indexed by $i \in I = [0; 1]$.

The manager of each unit performs two tasks: s and g. s is a unit-specific effort that affects only the revenue of the unit, and g is a general effort that increases the revenue of all units of the company. For a fast food chain, for example, s is the sales effort, and g is the effort to develop customer goodwill towards the brand name of the chain, or to learn about customer tastes, or to develop new products. Other examples concern employee hiring and training. s could be the effort to hire employees who are quick and able to follow existing production procedures exactly, while g is to hire employees who like to develop new production processes. s could also be the effort to train employees to have efficient services, while g is the effort to train employees to pay attention to customers' need and come up with new products. From now on, we will call s the sales effort and g goodwill. The level of the two efforts are not verifiable and hence cannot be contracted on.

We assume, however, that the revenue of each unit is verifiable, as is the case for fast food chains.[10] Furthermore, it is given by

$$x(i) = y(s(i); G) + \epsilon(i);$$

where $\epsilon(i)$ is a normally distributed random variable with mean 0 and variance σ^2, ϵ is independent across units, and $G = \int_I g(i)di$ is the total stock of goodwill possessed by the company. The specification of y assumes that g is a pure public good. We also assume that,

Assumption 1 y is increasing and concave in $(s; G)$.

Though the analysis of the model is made most clear-cut by the assumption of infinite units and the pure public good nature of $g(i)$, the qualitative features of our main results remain when the company has finite units and/or goodwill has local effects. We will discuss the relaxation of these assumptions in Subsection 3.6.

The manager incurs a private cost of $c(s; g)$ to provide these efforts. We assume that these efforts are perfectly substitutable in the manager's cost function and that effort is not

[10] In Section 4, we will consider scenarios where sales revenue may not be verifiable under certain conditions.

costly before it reaches a certain level.[11]

Assumption 2 The cost of efforts s and g is $c = c(s + g)$ and there exists some positive number T such that $c'(t) = 0$ for $t \le T$, $c(T) = 0$, and $c'(t) > 0$, $c''(t) > 0$ for $t > T$.

It is clear that we adopt a multi-task model pioneered by Holmstrom and Milgrom (1991). However, we emphasize the difference in the scope of influence of the two tasks, whereas they focus on the difference in the measurability of the tasks.

2.2. Incentive contracts

We assume that the manager has constant absolute risk aversion. That is, the manager's utility function is $u(z) = -e^{-rz}$, where r is the coefficient of absolute risk aversion and z is the manager's net (but risky) payoff. The HQ is assumed to be a risk-neutral profit maximizer. The HQ chooses a compensation scheme to induce the manager's efforts.

Although all managers are identical ex ante, the HQ can offer them different contracts. Without losing generality, we consider two types of contracts: (1) a fixed-wage contract, and (2) a high-powered contract that rewards the manager according to the revenue of her own unit.[12]

Let $w_i(x(i))$ be the incentive contract for the manager of unit i. Then the manager's expected utility is assumed to take the form

$$u(CE) \equiv E\{u[w_i(x(i)) - c(s(i) + g(i))]\},$$

where CE is the manager's certainty equivalent money payoff, and E is the expectation operator. Given the incentive contract, the manager chooses $s(i)$ and $g(i)$ to maximize $U(CE)$.

[11] We follow Holmstrom and Milgrom (1991) in using this cost function. They argue that "[i]n one-dimensional agency models, it is typically assumed that the agent will not work without incentive pay. The reason for this is not that the agent dislikes even small amounts of work, but rather that the level of work the agent would provide without explicit incentives does not affect the optimal solution. In multitask models, however, the fact that agents supply inputs even without incentive pay can be quite consequential." Indeed, the managers could be restricted from engaging in other activities. With nothing else to do, they might as well spend their time on sales and goodwill. Furthermore, monitoring could be effective in enforcing effort up to a certain level and incentive pay is only needed to induce effort beyond the level.

[12] One may argue that the HQ could also write: (a) a contract that rewards the manager based on the revenue of other units, or (b) a contract that rewards the manager based on the revenue of other units as well as on that of her own unit. Because there are infinite identical units, each unit is so small relative to the whole company that its level of goodwill does not affect the total stock of goodwill of the company. Furthermore, revenue is stochastically independent across units. Therefore, the revenue of other units does not contain any information about the manager's efforts and thus should not affect her compensation. Refer to Holmstrom (1982).

To summarize, the timing of events is as follows.
(1) At t=0, the HQ chooses contract $w_i(x(i))$ for all $i \in I$.
(2) At t=1, managers choose their sales and goodwill efforts simultaneously.
(3) At $t = 2$, the revenue is realized and the contracts are implemented.

3. Contract Mix

3.1. Contract design problem

In this analysis, we constrain the HQ to the choice of linear contracts. There is no loss of generality though. Holmstrom and Milgrom (1987) show that the optimal incentive contract in suitably stationary dynamic environments in which the agent can continuously monitor her own performance is equivalent to the optimum of a reduced-form static model in which the principal is constrained to linear contracts.

For $w_i = \alpha(i)x(i) + \beta(i)$ where $\alpha(i)$ and $\beta(i)$ are constants unrestricted in sign, we have:

$$CE(w_i) = \alpha(i)y(s(i); G) + \beta(i) - c(s(i) + g(i)) - \frac{1}{2}r\sigma^2\alpha(i)^2.$$

<u>At t = 1</u>, given the contracts offered by the HQ, $\{(\alpha(i); \beta(i)) : i \in I\}$, the managers choose $(s(i); g(i))$. Specifically, taking G as given, the manager of unit i solves

$$\max_{s(i) \geq 0; g(i) \geq 0} CE(w_i).$$

If $\alpha(i) > 0$, the manager of unit i chooses $g(i) = 0$ and

$$s(i) = \arg\max_{s(i) \geq 0} \alpha(i)y(s(i); G) - c(s(i)). \qquad (OP - s)$$

This is because, in this case, the manager has an incentive to increase the revenue of her unit. The optimal way to do so is to allocate no effort to goodwill and all her effort to sales, as her goodwill effort has only an infinitesimal effect on the revenue of the unit while sales effort has a non-trivial positive effect. An incentive contract with $\alpha(i) > 0$ is called a high-powered contract. A manager who receives a high-powered contract is called an H manager, and her unit an H unit.

If the manager of unit i receives a fixed-wage contract ($\alpha(i) = 0$), then she puts a total level of T effort (i.e., $s(i) + g(i) = T$ and $c(T) = 0$) and is indifferent between the sales and goodwill efforts. This is because, in contrast to the case of $\alpha(i) > 0$, the manager's payoff does not depend on the revenue of her unit and consequently not on how her effort is allocated between the two tasks. Therefore, when $\alpha(i) = 0$, the manager is assumed to do what is requested by the HQ among the set of $O_i = \{(s(i); g(i)) : s(i) + g(i) = T\}$. An incentive contract with $\alpha(i) = 0$ is called a low-powered contract. A manager who receives a low-powered contract is called an L manager, and her unit an L unit.[13]

[13] It is never optimal to have a negative α.

The above discussion shows that the HQ can induce a goodwill effort only by offering managers fixed-wage incentive contracts; there is discontinuity in the provision of the goodwill effort. The fundamental reason for this result is that goodwill is of public good nature, which implies that the effect of a manager's goodwill effort on the revenue of her unit is generally smaller than that of her sales effort. It follows that, so long as the manager's income depends on the revenue of her unit, she will not fully take into account the externality of her goodwill effort and will provide too little goodwill. The HQ can overcome the under-provision of goodwill effort only by delinking a manager's income from the revenue of her unit.

Let p be the proportion of L units and $L = [0; p] \subseteq I$ be the set of L units. Then $G = \int_L g(j)dj$ and the HQ's expected total profit is

$$\Pi = \int_L [y(T - g(j); G) - w(j)]dj + \int_{I-L} [(1 - \alpha(i))y(s(i); G) - \beta(i)]di;$$

where $g(j)$ is the level of goodwill effort chosen by the HQ for the L unit with index j and $s(i)$ is the level of sales effort chosen by the manager of the H unit with index i. At $t = 1$ (given p, $\{w(j)\}_{j \in L}$ and $\{\alpha(i); \beta(i)\}_{i \in I-L}$), taking $\{s(i)\}_{i \in I-L}$ as given, the HQ chooses $\{g(j)\}_{j \in L}$ to maximize Π, i.e.,

$$\{g(j)\} = \arg\max_{\{g(j)\}} \Pi: \qquad (OP - G)$$

In summary, at $t = 1$, $\{s(i)\}_{i \in I-L}$ and $\{g(j)\}_{j \in L}$ are jointly determined by $(OP - s)$ and $(OP - G)$.

<u>At $t = 0$</u>, the HQ chooses p, $\{w(j)\}_{j \in L}$ and $\{\alpha(i); \beta(i)\}_{i \in I-L}$ to maximize the expected total profit, subject to the incentive compatibility constraints (i.e., $(OP - s)$ and $(OP - G)$) and to the individual rationality constraints that managers are willing to accept the contracts. We normalize the reservation utility of the managers to be 0. Then, the HQ's problem at $t = 0$ is to choose p, $\{\alpha(i); \beta(i)\}_{i \in I-L}$, and $\{w(j)\}_{j \in L}$ to

$$\max \int_L [y(T - g(j); G) - w(j)]dj + \int_{I-L} [(1 - \alpha(i))y(s(i); G) - \beta(i)]di \qquad (OP - HQ)$$

$$\begin{aligned} s.t.: \quad & (OP - s); (OP - G) && (IC) \\ & w(j) \geq 0 \text{ for all } j \in L \\ & \alpha(i)y(s(i); G) + \beta(i) - c(s(i)) - \tfrac{1}{2}r\sigma^2\alpha(i)^2 \geq 0 \text{ for all } i \in I - L: && (IR) \end{aligned}$$

Since $w(j)$ and $\beta(i)$ do not affect the incentive compatibility constraints, the individual rationality constraints must be binding at the optimum. Otherwise, the HQ's expected total profit can be increased by reducing $w(j)$ or $\beta(i)$. By substituting constraints (IR) into the objective function, the HQ's optimization problem at $t = 0$ is simplified as:

$$\max_{p; \{\alpha(i)\}_{i \in I-L}} \int_L y(T - g(j); G)dj + \int_{I-L} [y(s(i); G) - c(s(i)) - \frac{1}{2}r\sigma^2\alpha(i)^2]di \qquad (OP - HQ')$$

$$s.t. \quad (OP - s), (OP - G). \qquad (IC)$$

In order to analyze the HQ's contract design problem, we first characterize some necessary conditions for the optimal incentive contracts (Subsection 3.2) and then investigate the equilibrium outcome given the incentive contracts (Subsection 3.3).

3.2. Uniformity of high-powered and low-powered incentive contracts

Our first result regards the uniformity of L units.

Proposition 1 The HQ should request the same level of goodwill effort from all L managers.

$\{g(j)\}$ is chosen at t=1 to solve program $(OP - G)$. Given any p and $\{\alpha(i)\}_{i \in I - L}$, this can be done in two steps: (1) Given G, choose the optimal allocation $\{g(j)\}$ so that $\int g(j)dj = G$; (2) Choose the optimal G. Note that, for given G, the second term of the objective function of $(OP - G)$ is independent of the allocation $\{g(j)\}$. Therefore, step (1) is equivalent to maximizing $\int_L y(T - g(j); G)dj$. Since $y(T - g; G)$ is concave in g, an application of Jensen's inequality implies that it is optimal to choose the same $g(j)$ for all L units.

To characterize the optimal incentive contracts further, we make additional assumptions about the production and cost functions.

Assumption 3 $\lim_{s \to 0} y_s \equiv \frac{\partial y(s;G)}{\partial s} = \infty$, and $\lim_{G \to 0} y_G \equiv \frac{\partial y(s;G)}{\partial G} = \infty$.

Assumption 4 $\frac{c'(s)}{y_s(s;G)}$ is convex in s.

Assumption 5 The marginal product of goodwill effort, $y_G \equiv \frac{\partial y(s;G)}{\partial G}$, is increasing and concave in s.

Assumption 3 is made to avoid possible complications of corner solutions. Assumption 4 holds if the cost function is sufficiently convex. In the case that the production function is Cobb-Douglas and the cost function is $c = (s + g)^{\kappa}$, it is satisfied if λ is sufficiently large. Assumption 5 says that s and G are complementary and the return to s in enhancing the marginal product of G diminishes.

By Proposition 1, Assumptions 1 and 3, we have

Lemma 1 Program $(OP - HQ')$ is equivalent to

$$\max_{p;\{\alpha(i)\}_{i \in I - L}} \; py(T - g; G) + \int_{I-L} [y(s(i); G) - c(s(i)) - \frac{1}{2} r\sigma^2 \alpha(i)^2] di \qquad (OP - HQ'')$$

$$\begin{aligned} s.t.\quad & p y_G(T - \tfrac{G}{p}; G) - y_s(T - \tfrac{G}{p}; G) + \textstyle\int_{I-L} (1 - \alpha(i)) y_G(s(i); G) di \geq 0 && (IC - G) \\ & \alpha(i) y_s(s(i); G) = c'(s(i)) && (IC - s) \end{aligned}$$

Program $(OP - HQ'')$ can be solved in three steps: (1) Solve for $\alpha(i)$ from constraint $(IC - s)$ and substitute it into the objective function and other constraint; (2) Given G and p, solve for the optimal $\{s(i)\}$; (3) Solve for the optimal G and p. By Lemma 1, Assumptions 4 and 5, we can show that step (2) is a concave program.

Lemma 2 Let $\alpha(i) = c'(s)/y_s(s;G)$. Then, given G, the integrand in the objective function of program $(OP - HQ'')$,

$$y(s(i);G) - c(s(i)) - \frac{1}{2} r\sigma^2 \alpha(i)^2;$$

and the integrand in constraint $(IC - G)$,

$$(1 - \alpha(i)) y_G(s(i);G);$$

are both concave functions of $s(i)$.

Lemma 2 and an application of Jensen's inequality imply,

Proposition 2 The HQ should offer the same high-powered contract to all H managers.

Proposition 2 is more involved than Proposition 1 because, in contrast to the L managers, both the sales and total efforts by the H managers are directly affected by the high-powered incentive contracts, as is the HQ's choice of goodwill level. Instead of offering a single high-powered contract, the HQ could offer two contracts, one less high-powered and the other more high-powered, under which one group of H managers would decrease their sales effort and the other increase their sales effort. When the cost function is not convex enough, the decrease in sales effort is not too much while the increase in the sales effort is quite a lot, and the HQ could benefit by offering different high-powered contracts. However, when the cost function is sufficiently convex (specifically, when Assumption 4 is satisfied), offering the same high-powered contract increases the average sales effort of the H managers. When s and G are complementary as Assumption 5 says, this also increases the HQ's incentive to choose a high level of goodwill.

Empirical studies by Lafontaine and Shaw (1996) show that, while they vary from one franchisor to another, franchise contracts are extremely uniform across franchisees and stable over time within any particular franchise company. McAfee and Schwartz (1994) offer a market-based explanation for this phenomenon. Specifically, in their model, after signing a contract with one franchisee, a franchisor is tempted to offer another franchisee a contract with a lower royalty rate to undercut the first franchisee and therefore obtain a higher lump-sum fee. Knowing the franchisor has such opportunistic behavior, the first franchisee is reluctant to accept the contract, and the franchisor is thus better off by committing to a uniform contract for all franchisees. Note that in our set-up, each manager is a local monopoly and McAfee and Schwartz's argument is no longer applicable. Our model thus offers an alternative and purely technological explanation for the uniformity of franchise contracts.

3.3. Equilibrium outcomes given the uniform contracts

Propositions 1 and 2 greatly simplify the contract design problem (Subsection 3.1). Let (w) be the low-powered contract offered by the HQ to all L managers, and g is the goodwill effort. Let $(\alpha;\beta)$ be the high-powered contract offered by the HQ to all H managers, and s is the corresponding sales effort.

Recall that, at $t = 1$ (given p, (w) and $(\alpha;\beta)$), H managers choose s and the HQ picks g. For $\alpha > 0$, H manager's optimal choice of s is characterized by

$$\alpha y_s(s;G) - c'(s) = 0. \qquad (FOC - s)$$

The HQ's objective function at $t = 1$ becomes

$$\Pi_1 = p[y(T - g;G) - w] + (1 - p)[(1 - \alpha)y(s;G) - \beta]; \quad \text{where } G = pg.$$

For $p > 0$, the HQ's optimal choice of g is characterized by

$$p y_G(T - \frac{G}{p};G) - y_s(T - \frac{G}{p};G) + (1 - p)(1 - \alpha)y_G(s;G) = 0; \qquad (FOC - G)$$

where the left hand side of the equation is equal to $\frac{\partial \Pi_1}{\partial G}$. Thus, at $t = 1$ (given p, (w) and $(\alpha;\beta)$), equilibrium $(s;G)$ are determined simultaneously by $(FOC - s)$ and $(FOC - G)$.

By the concavity of y, it is a standard exercise to show that, for $\alpha > 0$ and $p > 0$, the Jacobian matrix of $(FOC - s)$ and $(FOC - G)$ has a positive determinant. Then, we have,

Lemma 3 For $\alpha > 0$ and $p > 0$: (1) Equilibrium s and G are differentiable functions of $(\alpha;p)$. (2) Equilibrium s and G both increase (or decrease) with p if $\frac{\partial^2 \Pi_1}{\partial p \partial G} > 0$ (or $\frac{\partial^2 \Pi_1}{\partial p \partial G} < 0$). (3) The effects of α on equilibrium s and G are usually ambiguous.

Regarding the boundary cases, for $p = 0$, $G = pg = 0$ and s is determined by

$$\alpha y_s(s;0) - c'(s) = 0.$$

For $\alpha = 0$, s and G are defined as:[14]

$$s(0;p) = \lim_{\alpha \to 0} s(\alpha;p) \text{ and } G(0;p) = \lim_{\alpha \to 0} G(\alpha;p).$$

Then, we can show,

Lemma 4 s and G are continuous functions of $(\alpha;p)$ on $[0;1]^2$.

[14] In proving Lemma 4, we also show the existence of these limits.

To conclude our analysis in Subsections 3.2-3.3, note that the HQ's objective function at $t = 0$ becomes:

$$\Pi_0(\alpha, p) = p y(T - \frac{G}{p}, G) + (1 - p)[y(s, G) - c(s) - \frac{1}{2} r\sigma^2\alpha^2], \tag{1}$$

where s and g are the equilibrium defined by $(FOC - s)$ and $(FOC - G)$, and functions of (α, p). By Lemma 4, $\Pi_0(\alpha, p)$ is a continuous function of (α, p) on $[0, 1]^2$. The HQ chooses α and p to maximize $\Pi_0(\alpha, p)$. The Weierstrass Theorem then implies,

Corollary 1 The optimal (α, p) exists.

3.4. Characterization of the high-powered contract

When the HQ chooses α, it considers several factors. First, as α increases, H managers are subject to more risk and therefore need to be compensated more. Second, the value of α affects the HQ's incentives to provide goodwill effort and H manager's incentives to expend sales effort. Differentiate Π_0 with respect to α and simplify the derivative with $(FOC - s)$ and $(FOC - G)$. We have, for $\alpha > 0$,

$$\frac{d\Pi_0}{d\alpha} = -(1 - p)r\sigma^2\alpha + \alpha(1 - p)y_G^H \frac{dG}{d\alpha} + (1 - p)(1 - \alpha)y_s^H \frac{ds}{d\alpha}, \tag{2}$$

where a function with a superscript H means that it is evaluated at (s, G). On the right hand side of (2), the first term captures the risk factor and is negative, and the next two terms describe the effects of α on Π_0 through G and s, respectively.

Proposition 3 For any $p \in [0, 1)$, the optimal $\alpha \in (0, 1)$.[15]

The intuition for this result is quite clear. To induce effort from H managers, the HQ has to offer them a positive α. To commit itself to choosing a high level of goodwill effort, the HQ needs $\alpha < 1$ so that it still cares about the effect of goodwill effort on H units' revenues.

3.5. Characterization of contract mix

By equation (1), the HQ's total profit is $\Pi_0 = p\pi^L + (1 - p)\pi^H$, where $\pi^L \equiv y(T - \frac{G}{p}, G)$ is the HQ's expected profit from a L unit, and $\pi^H \equiv y(s, G) - c(s) - \frac{1}{2} r\sigma^2\alpha^2$ is the HQ's expected profit from a H unit. If π^L and π^H were independent of p, then either a high-powered contract or a low-powered contract would be offered by the HQ to all managers (namely, either $p = 0$

[15] For the case $p = 0$, the proposition needs the additional assumption that $y(s, 0) \neq 0$. Note that $\Pi_0(0, 0)$ is defined to be $\lim_{\alpha \to 0} \Pi_0(\alpha, 0)$. At $p = 0$, Proposition 3 says that there exists some $\alpha' \in (0, 1)$ such that $\lim_{\alpha \to 0} \Pi_0(\alpha, 0) < \Pi_0(\alpha', 0)$.

or $p = 1$). The contract mix (namely, $p \in (0,1)$) would only be optimal with probability zero when $\pi^L = \pi^H$.

In this model, however, goodwill is a public good. In addition, only L managers can provide a goodwill effort. This seems to justify the existence of L units as they have positive externality on other units of the company. On the other hand, L managers, because of the low-powered contract, expend a lower level of total effort than H managers ($s(j) + g(j) = T$ for $j \in L$, whereas $s(i) \geq T$ for $i \in I - L$), which makes the L units less profitable. Intuition therefore suggests that the HQ offers a high-powered contract to some managers and a low-powered contract to the others.

To investigate the optimality of contract mix, for $p > 0$, we differentiate Π_0 with respect to p and simplify the derivative by $(FOC - s)$ and $(FOC - G)$.

$$\frac{d\Pi_0}{dp} = (\pi^L - \pi^H) + (1-p)\alpha y_G^H \frac{\partial G}{\partial p} + (1-p)(1-\alpha) y_s^H \frac{\partial s}{\partial p} + g y_s^L; \tag{3}$$

where a function with a superscript H (L, respectively) means that it is evaluated at (s, G) $((T - \frac{G}{p}, G)$, respectively).

When there is one more L unit, the HQ gains π^L but loses π^H. The first term on the right hand side of (3) captures this direct effect. Proposition 4 says that this direct effect is negative.

Proposition 4 For any given $p \in (0,1)$, $\pi^L < \pi^H$ at the corresponding HQ's optimal choice of α.

There are two reasons for the result of Proposition 4. One is that, to induce goodwill effort, the HQ is constrained to offering the low-powered incentive contract to L managers. As a result, L managers expend a lower level of total effort than H managers. The other reason is that H managers free ride on L managers for goodwill and therefore they are able to put more effort on sales. Both reasons stem from the public good nature of goodwill effort. Interestingly, Proposition 4 is supported by empirical findings (Shelton (1967)).

Proposition 4, however, does not imply that there should be no or very few L units. L units are important because they are the only providers of goodwill effort, which not only increases profits of all units directly, but also affects the productivity of sales effort in H units. The second and the third terms on the right hand side of (3) capture, respectively, the effect of p on Π through total stock of goodwill and that through the level of sales effort of H managers. The signs of these two terms are the same as those of $\frac{\partial G}{\partial p}$ and $\frac{\partial s}{\partial p}$ respectively, and are therefore ambiguous by Lemma 3.

When $p \neq 1$, the burden of providing a given level of goodwill, G, is borne entirely by L units. Because the sales effort has decreasing returns to scale, the cost of L units devoting less effort to sales exceeds the benefit of H units being able to exert more effort in sales. Having more L units mitigates such inefficient substitution between sales and goodwill efforts across units. This substitution effect is captured by the fourth term on the right hand side of (3) and is always positive.

Note that, if $y(s,0) = 0$ for all s, the optimality of $p^* > 0$ is straightforward. Specifically, at $p = 0$, $G = 0$ and $\Pi = 0$; but at $p > 0$, $G > 0$ and $\Pi > 0$. Therefore, the optimal p must be positive. Intuitively, if the revenue vanishes in the absence of goodwill, then the HQ must have the L units to provide the essential input.

Our analysis, however, is focused on the scenario that $y(s,0) > 0$. It is then not so apparent that the optimal p has to be positive. In that case, a sufficient condition for $p^* > 0$ is $\frac{d\Pi}{dp} > 0$ as $p \to 0$, which we can show under the following assumption.

Assumption 6 If $y(s,0) > 0$ for $s > 0$, then $y_{sss} \geq 0$ and $\lim_{(s,G)\to(0,0)} y_s(s,G) = \infty$.

Assumptions 1, 3, 5, and 6 (all about the revenue function) are satisfied by the functions of $y(s,G) = c_0 s^a G^b + c_1 s^\delta + c_2 G^\gamma$, with c_0, c_1, and c_2 non-negative as well as a, b, δ and γ between 0 and 1, which include Cobb-Douglas functions as special cases. Note that, in terms of the importance of G to the revenue function, Assumption 6 imposes much fewer restrictions than $y(s,0) = 0$ does. Furthermore, Assumption 4 is satisfied if y takes the above form and the cost function has the form of $c = (s + g)^\lambda$ with a sufficiently large λ.

Lemma 5 If $y(s,0) \neq 0$, then $\lim_{p\to 0} \frac{\partial \Pi}{\partial p} > 0$ for $\alpha < 1$.

This, together with Lemma 4 and Proposition 3, implies that

Proposition 5 It is optimal for the company to have some L units.

Having established the optimality of $p^* > 0$, we turn to the question of when it is optimal for the HQ to have some H units, namely, $p^* < 1$. With L units providing the goodwill, the HQ can offer managers of the remaining units high-powered incentive contracts and thereby elicit high sales effort from them. Whether or not high-powered incentive contracts should be offered to some manager depends on the magnitudes of their costs and benefits.

One cost of high-powered incentive contracts is that they subject the managers to risks, resulting in inefficient risk sharing between the risk neutral HQ and the risk averse managers. The cost is lower when the managers are less risk averse or when there is less uncertainty. The main benefit of high-powered incentive contracts is that they elicit a high sales effort from H managers. When the marginal costs of sales effort are lower, it is easier to induce it and thus the benefit of high-powered contracts is higher. Therefore, we have:

Proposition 6 It is more likely for the HQ to have H units when the managers are less risk averse, or when the uncertainty about the revenue is smaller, or when marginal cost of effort is smaller.

Proposition 6 gives one set of conditions for the HQ to have H units, namely, H units are attractive enough. A complementary set of conditions is that L units do not perform very well. With low-powered incentive contracts, the total level of efforts is T. If T is very small, the profits from L units are very low. Therefore, we have:

Proposition 7 *It is more likely for the HQ to have H units when T is smaller.*

3.6. Discussion

Before concluding this section, we discuss the pure public good assumption and the infinite unit assumption. Then we argue for the consistency of our model with the evidence on the locational distribution of different types of units.

3.6.1. Finite units and local effects of goodwill

When g is not a pure public good (equivalently, g has some local effect), the managers will admittedly have more incentives to develop goodwill. However, they still will not take into account the positive externality of goodwill on other units and thus will not do enough for goodwill. In fact, so long as the local effect of goodwill effort is not extremely strong, it is still optimal to give some units low-powered incentive contracts so that their managers are more willing to provide goodwill than they would when they had high-powered incentive contracts.

When there are only finite units, the optimal reward to a manager could depend on the revenue of other units and this, admittedly, will give the manager some incentives to develop goodwill.[16] However, for such incentives to be strong enough, the share of the manager's income from the revenue of other units must be large enough. Such a high share brings about a high cost of risk and has to be compensated by substantially lowering the share of income from the revenue of the manager's own unit.[17] Therefore, the incentives for sales effort will be greatly reduced. In summary, inducing goodwill effort by linking the manager's reward to the revenue of other units is very costly in terms of sales effort. It is still better to generate goodwill by giving some managers low-powered incentive contracts.

3.6.2. Highway evidence

Since Rubin (1978), the franchise has been considered as a way of dealing with managers' incentive problems rather than raising money from small investors. Assuming that the goodwill is already established, the choice between company units and franchised units is influenced by the following two concerns: (1) company managers are more likely to put in low level of effort (the shirking problem), (2) franchisees have more incentive to substitute low quality products (the free-rider problem). What is implicit in the above discussion is that the shirking problem is much less severe for franchisees as they have high-powered contracts, while

[16] In this context, a high-powered contract links the manager's reward more closely to the revenue of her own unit than to that of other units, whereas a low-powered contract links her reward only to the total revenue of the whole company.

[17] Roughly speaking, a one-percentage-point increase in the share of revenue from other units needs to be compensated by a $\sqrt{N}$-percentage-point decrease in the share of revenue from her own unit, where N is the number of units in the company.

the free-rider problem is much less severe for company managers as they have ¯xed wages.

When a unit is away from its monitoring center (HQ), the shirking problem becomes severe if the unit is company-owned. This implies that company ownership is more likely for units that are closer to the monitoring center. On the other hand, when a unit is situated in an area (such as a highway) where there are a lot of non-repeat customers, then the free-rider problem is serious if the unit is franchised. This suggests that, compared with other units, highway units are more likely to be company owned. Brickley and Dark (1987) formally test the above two hypotheses, but they ¯nd support only for the ¯rst one. In particular, they ¯nd that highway units are more likely to be franchised. Holmstrom and Milgrom (1994) also discuss the tension between their theoretical predictions and the highway evidence.

Note that, in our model, the goodwill is endogeneously generated. Speci¯cally, company-owned units are chosen for the production of goodwill, whereas franchised units are used to reap the bene¯ts of company goodwill. In other words, the HQ is fully aware of the low goodwill e®ort in the franchised units, and it is simply not optimal for the HQ to induce same goodwill e®ort in all units. This implies that the free rider problem (which is a®ected by the percentage of non-repeat customers) itself is not that crucial so long as the HQ can enforce the minimum quality standard (i.e., g is non-negative). More importantly, when units are heterogeneous, our model would predict that company ownership is more likely for those units that have more advantage in generating goodwill and that the HQ is located near those units to alleviate the shirking problem of the company managers.

Consider again that there are many types of customers (highway, residential, downtown, etc.). We want to argue that highway units do not have more advantage in generating goodwill than downtown stores, which could then explain the highway evidence. Our argument is based on how goodwill or brand name is established. Conceivably, for goodwill production, a company needs to ensure that its initial customers like the products and then \spread" favorable impressions to those who have not purchased the products. To achieve maximum impact, it is important that those initial customers are from various submarkets and therefore they can spread their impressions to a wide population of the society. We believe that the downtown units serve more diverse customers than the highway units. For example, for a Boston Common McDonald's, the customers could be tourists or business people of other cities as well as people who live in the vicinity of the Boston Common. Even truck drivers who usually frequent highway McDonald's may bring their kids to the Boston Common on Sundays. The rest of the argument follows immediately.

It is important to our argument that there are many types of customers, only one of which is highway customers. If the whole market were highway customers and downtown customers, then it would be optimal to have some company owned units along highways to spread positive impressions directly among such a large group of customers.

We have argued that the downtown units are in better position to spread positive impressions about existing brands than highway units. The same argument also works for product development. Note that, for the HQ to assess the marketability of a new product, it is important to get feedback from a diverse group of customers at fewest number of stores. As argued above, the downtown units tend to have more diverse customers and therefore they are natural candidates for company ownership.[18]

4. Ownership and Contract Enforceability

The above analysis has addressed the ¯rst distinguishing feature of a franchise, namely, that both high-powered and low-powered incentive contracts are used. What remains to be explained is why managers with the high-powered incentive contract own some or all of their units' physical assets whereas those with the low-powered incentive contract do not. In this section, we ¯rst discuss a stylized fact of franchise, namely, incompleteness and hence unenforceability (by third parties) of ex ante incentive contracts. We then show that the incentive contracts of various power can only be voluntarily enforced, and hence the desirable incentives be protected, in the presence of the corresponding ownership arrangements.

4.1. Contractual incompleteness in a franchise

A basic assumption in Sections 2 and 3 is that contracts are complete. The HQ o®ers the low-powered incentive contract to some managers who provide the goodwill e®ort, and the high-powered incentive contract to others who expend all their e®ort on sales. Once these contracts are written, the HQ is no longer needed for carrying out the business.[19] In this complete-contracting framework, it does not matter whether the units' physical assets are owned by the managers or by the HQ.[20]

In reality, the retail markets for the HQ's brand name products and services are significantly uncertain. The HQ needs to set the standards and policies to be followed by the franchisees in response to market changes in order to develop and maintain its brand name. For example, when new scienti¯c studies reveal that some of the food ingredients are not healthy enough, the HQ (of a fast-food business) may choose to replace those ingredients. To write complete contracts with the managers, it requires the HQ to foresee all possible future contingencies and devise corresponding strategies for the managers, which is very costly if not impossible. As a result, \many of the standards with which a franchisee must comply will not even be articulated until well after the contract has been signed," and \the key characteristic of the franchise contract is its incompleteness" (Had¯eld (1990)).

Rather than writing complete contracts ex ante, the HQ retains the residual rights of control to make business decisions ex post as unforeseen contingencies arise (Grossman and

[18] This argument seems to imply that the signaling theory is also consistent with the highway evidence, as downtown stores serve more diverse customers and would then provide better signal about pro¯tability. However, downtown stores only provide signal about the average pro¯tability of all stores, whereas a potential franchisee prefers signal about stores of the same type as hers. The best way for the company to signal is to own units in each of the submarkets.

[19] Note that contracts could be complete though the goodwill and sales e®orts are not veri¯able. See Hart (1995).

[20] Throughout the paper, we assume that the HQ owns the goodwill stock. This is in fact an optimal arrangement. We will further discuss this after Proposition 10.

Hart (1986), Hart and Moore (1990), Hart (1995)). An examination of the McDonald's franchise contract reveals that the HQ has substantial residual rights of control, and that its franchises are required to \strict adherence to licensor's standards and policies as they exist now and as they may be from time to time modi¯ed." A recent example is the outbreak of mad cow disease and the subsequent decision by McDonald's of not using UK beef.

While the residual rights of control greatly facilitates the HQ to develop and maintain its brand name in response to market changes, it also gives the HQ an opportunity to hold up the managers ex post. A HQ may investigate \a minor or curable contract violation not to promote the quality of its franchisees but to achieve some other, opportunistic goal at the franchisee's expense," or abuse its power \in order to transfer the franchises to more pro¯table franchisees or to convert the outlets to company ownership" (Had¯eld (1990)).

It should be stressed that the objective of this paper is not to probe why franchise contracts are often incomplete. There is an extensive body of legal studies on this issue (see Had¯eld (1990) and references therein). Among the reasons suggested are the importance of brand name products and services, the uncertainty of the retail markets, and the need for quick responses to possible market changes. See also Anderlini and Felli (1994), Hart (1995), MacLeod (1996a, 1996b), Maskin and Tirole (1996), and Segal (1995) for the theoretical foundations of the incomplete-contracts approach. What this paper attempts to show is that, given the contractual incompleteness in a franchise, the ex ante incentive contracts of various power can be voluntarily enforced only in the presence of the corresponding ownership arrangements.

4.2. Unenforceability of ex ante contracts

In the presence of contractual incompleteness, disputes between the HQ and the managers are inevitable. Moreover, contractual incompleteness makes it di±cult to verify which party is at fault and implement penalties for breach of ex ante incentive contracts.[21] This implies that ex ante contracts may become unenforceable ex post, and renegotiation could be initiated by either the HQ or the managers.

Accordingly, the timing of events is modi¯ed as follows (see also Figure 1). At $t = 0$, the HQ designs the high-powered and low-powered contracts and chooses the contract mix ratio (p).[22] It also makes ownership arrangements for the units' physical assets. At

[21] In her seminal paper, Had¯eld (1990) reviews several cases on franchise contracts (Picture Lake Campground v. Holiday Inns, 497 F. Supp. 858, 869 (E.D. Va. 1980). Vaughn v. General Foods Corp., 797 F.2d 1407, 1411, 1415 (7th Cir. 1986)) and concludes that the courts have taken the so-called business judgment approach. Speci¯cally, the courts treat the \franchisor's interest as if it represented the entirety of the relation." As a result, the franchisors could always come up with some business justi¯cation for termination of franchise contracts and therefore do not need to pay any penalty to the franchisees. As for franchisees, they certainly have the freedom to leave the relationship voluntarily; however, under the business judgment approach, they don't expect to receive any payment from the franchisors except for getting back the value that is attached with their asset investment.

[22] In the following analysis, we consider scenarios where it is optimal for the HQ to o®er some

$t = 1$, the managers choose s and g simultaneously. We call this the investment stage of the game. We maintain the assumption that, if a manager is indifferent between some choices, she will choose the one most preferred by the HQ. At $t = 2$, potential revenue becomes known to the HQ and the manager. They then have the option to sever their relationship. They decide whether to sever the relationship, and if not, whether to renegotiate the revenue-sharing contract. We call this stage of the game the renegotiation stage. As there is perfect information at his stage, renegotiation takes a negligible amount of time. After the renegotiation stage, production is carried out. We normalize the length of the entire game to be 1 and assume that the period before $t=2$ is of length $1 - \tau \in [0; 1)$ and that after $t=2$ is thus of length τ.[23]

We adopt MacLeod and Malcomson's (1993) specification and analysis of the renegotiation stage of the game. In MacLeod and Malcomson (1993), both a no-trade and outside option could be the triggers for contract renegotiation. In our model, however, trade is always preferred by both parties to no-trade, and the only possible trigger for contract renegotiation is the outside option. Let u denote the HQ's stage payoff under the ex ante incentive contract, and v denote that of the manager. Let u_0 be the HQ's stage payoff from the outside option, and v_0 be that of the manager. u_0 and v_0 are determined by the ownership arrangements in Subsection 4.3, and they can be specified by contracts in Subsection 4.4.

We obtain the following equilibrium of the renegotiation subgame.

Proposition 8 (MacLeod and Malcomson (1993)) *Suppose $u + v > u_0 + v_0$. The payoffs to the HQ and the manager in any subgame perfect equilibrium of the renegotiation subgame are given by U and V, where ρ is the interest rate,*

$$\rho V = u + v - \rho U; \tag{4}$$

$$\rho U = \begin{cases} u + v - v_0; & \text{when } v_0 > v; \\ u_0; & \text{when } u_0 > u; \\ u; & \text{otherwise.} \end{cases} \tag{5}$$

Proposition 8 shows that, whenever the manager (or the HQ) has higher payoff from the outside option than from the ex ante incentive contract, the renegotiation is initiated and the manager (or the HQ) obtains her outside option payoff under the new incentive contract.

4.3. Ownership and self-enforcement of ex ante contracts

It was shown in Subsection 4.2 that, depending on the outside options of the HQ and the managers, the revenue-sharing contracts (designed at $t = 0$) could be renegotiated (at $t = 2$) thereby distorting the effort incentive of the concerned parties. In this subsection, we assume away any contractual remedies for the event of an outside option. The parties' payoffs under

managers the high-powered contract (i.e., $p^* < 1$).

[23] As long as the incentive contract is not renegotiated after the uncertainty is resolved, the argument for linear contracts still applies (Holmstrom and Milgrom (1987)).

the outside option are thus determined solely by the ownership arrangements of the units' physical assets. We show that appropriate ownership arrangements at $t = 0$ can ensure voluntary self-enforcement of the ex ante contracts at $t = 2$. In the next subsection, we conclude our analysis by showing that the contractual remedies, even optimally chosen, cannot mimic what the ownership arrangements do. Taken together, we can explain not only multiple ownership arrangements of a franchise but also their correspondence with the incentive contracts of various power.

Recall that, in the complete-contracting framework as described in Section 3, the HQ offers the high-powered incentive contract $(\alpha^*; \beta^*)$ to some managers and the low-powered incentive contract (w^*) to others. These contracts induce the second-best sales effort s^* and goodwill effort $G^* = pg^*$. In the incomplete-contracting framework considered here, the same ex ante contracts would generally lead to inefficient equilibrium sales and goodwill efforts, denoted by s^e and G^e respectively, as the ex ante contracts may be renegotiated. Our objective is to find ownership arrangements under which, in the subgame perfect equilibrium, $(s^e; G^e)$ coincides with the second-best $(s^*; G^*)$ and no renegotiation occurs. Such ownership arrangements, if they exist, are said to make the ex ante contracts $(\alpha^*; \beta^*)$ and (w^*) self-enforcing.

Before we analyze individual cases, note that the managers' efforts are not human capital in our model. Once the goodwill and sales efforts are made by the manager of a unit, the former is attached to the company's brand name while the latter is embedded in the unit's physical asset.

Also note that the HQ could make a lump-sum payment w_1 to the manager between the investment stage and the renegotiation stage.[24] Then the promised lump-sum payment to the manager in the production stage becomes $\beta^* - w_1$ for the high-powered contract, or $w^* - w_1$ for the low-powered contract. A proper choice of w_1 may reduce the pressure for renegotiation.

In Section 3, we harmlessly normalized the manager's reservation wage to 0. In this section, the reservation wage rate is not neutral and is denoted by w_0.

(A) We first turn to the question of whether ownership arrangements can ensure self enforcement of the high-powered contract $(\alpha^*; \beta^*)$. Under this contract, for the production period, the manager is promised a payment of $\alpha^* y(s^e; G^e) + (\beta^* - w_1)$ while the HQ is to get $(1 - \alpha^*) y(s^e; G^e) - (\beta^* - w_1)$.

(A1) Suppose that the manager owns the unit's physical assets.[25] In this case, the manager's outside option is either to provide generic goods and services without access to the company goodwill $(y(s^e; 0))$ or to work for another company and get her reservation wage (τw_0).[26] On the other hand, the HQ's outside option is $y(0; G^e) - \tau w_0$ (where the

[24] The payment has to be a fixed amount, as the revenue is not verifiable yet.

[25] It can be checked that the cost of asset ownership is unimportant to the following analysis and is thus assumed to be zero.

[26] For the clarity of argument, we assume that there is no uncertainty about the revenue, i.e., $\sigma = 0$, and restrict the discussion to linear incentive contracts. Note that in Section 3, uncertainty

HQ pays τw_0 to hire a new manager for the production stage) as it will lose access to the effort-embedded physical asset.

Lemma 6 Suppose that the manager owns the physical assets. The H contract is self-enforcing if conditions

$$\max_s \{y(s;0) - c(s)\} + w_1 \leq \alpha^* y(s^*;G^*) - c(s^*) + \beta^* = w_0; \quad (6)$$

$$\tau w_0 + w_1 < \alpha^* y(s^*;G^*) - c(s^*) + \beta^* = w_0; \quad (7)$$

$$\max_g \{py(T-g;pg) + (1-p)y(0;pg)\}$$
$$\leq py(T-g^*;G^*) + (1-p)[y(s^*;G^*) - c(s^*)] + (1-p)(w_1 + \tau w_0 - w_0). \quad (8)$$

are met.

(6) implies that the manager can not benefit from leaving to work for herself with her asset at t=2 even if she chooses s optimally for this purpose. (7) means that neither can the manager benefit from leaving to work for other employers at t=2. (8) means that the HQ can not benefit from kicking out H managers even if it directs the L managers to choose the optimal g for this purpose. Together, the three conditions ensure that the second-best s^* and G^* will be chosen and renegotiation will not occur.

(6), (7) and (8) are implied by strong complementarity between s and G, and high efficiency of the production technology. To parameterize complementarity and efficiency, we consider the following class of revenue functions:

$$y(s;G) = kz(s;G) + k_1\mu(s) + k_2\nu(G);$$

where $z(0;G) = z(s;0) = \mu(0) = \nu(0) = 0$, $\frac{\partial^2 z}{\partial s \partial G} > 0$, z, μ and ν are increasing and concave. $z(s;G)$ captures the complementarity between s and G, and thus k parameterizes the strength of the complementarity. k also parameterizes the efficiency of the production technology. We can show:

Lemma 7 Let $k_1 = k_2 = 0$. Then for sufficiently large k, there exists some w_1 such that (6)-(8) hold with strict inequality.

Lemma 8 Suppose for some k and w_1, (6)-(8) hold with strict inequality as $k_1 = k_2 = 0$. Then (6)-(8) also hold for sufficiently small k_1 and k_2.

Corollary 2 For sufficiently large k and sufficiently small k_1 and k_2, there exists some w_1 such that (6)-(8) are satisfied.

---

plays no essential role other than rationalizing linear incentive contracts if we assume that the HQ cannot take a negative share of any unit's revenue and thus the shares of any unit's revenue going to the managers must total no greater than one. All the results in Section 3 except Proposition 6 still hold. The discussion in Subsection 3.6 can also be modified so that it does not depend on uncertainty.

(A2) Suppose instead that the HQ owns the unit's physical assets. In this case, the HQ can deny the manager access to the company goodwill and the effort-embedded physical assets. The manager has the outside option of working for another company and getting her reservation wage (τw_0), while the HQ has the outside option of hiring a new manager and capturing all the revenue ($y(s^e; G^e) - \tau w_0$).

A necessary condition for the contract to be self-enforcing is that, when $(s^*; G^*)$ is chosen, neither party wants to renegotiate. By Proposition 8, this is equivalent to:

$$\alpha^* y(s^*; G^*) + (\beta^* - w_1) \geq \tau w_0;$$

$$(1 - \alpha^*) y(s^*; G^*) - (\beta^* - w_1) \geq y(s^*; G^*) - \tau w_0.$$

These inequalities imply that $w_1 = \alpha^* y(s^*; G^*) + \beta^* - \tau w_0$. Given such w_1, however, the manager will deviate from s^*. By choosing $s' = T$, the manager gets w_1 before the realization of the revenue and τw_0 after the renegotiation, and her total payoff is $w_1 + \tau w_0 - c(T) = \alpha^* y(s^*; G^*) + \beta^* = w_0 + c(s^*)$, which is higher than her payoff of w_0 by choosing s^*. Therefore, the H contract is not self-enforcing when the HQ owns the physical assets. Combining (A1) and (A2), we have:

Proposition 9 For sufficiently large k and sufficiently small k_1 and k_2, the high-powered incentive contract $(\alpha^*; \beta^*)$ is self-enforcing ex post if and only if the manager owns the unit's physical assets.

In other words, when s and G are strongly complementary and the production technology is very efficient, the necessary and sufficient condition for the high-powered contract $(\alpha^*; \beta^*)$ to be self-enforcing is manager ownership of the unit's physical assets.

(B) Next we turn to the question of whether ownership arrangements can ensure self enforcement of the low-powered contract ($w^* = w_0$). Under this contract, the production period payoffs are $w_0 - w_1$ for the manager and $y(s^e; G^e) - w_0 + w_1$ for the HQ.

(B1) Suppose that the HQ owns the physical assets. As in the case of (A2), the manager and the HQ have, respectively, outside options of τw_0 and $y(s^e; G^e) - \tau w_0$. By Proposition 8, the contract is not renegotiated if and only if

$$w_0 - w_1 \geq \tau w_0;$$

$$y(s^e; G^e) - w_0 + w_1 \geq y(s^e; G^e) - \tau w_0.$$

These inequalities imply that $w_1 = (1 - \tau) w_0$. Note that this condition does not depend on $(s^e; G^e)$. In particular, it ensures $\{s^*; G^*;$ no renegotiation$\}$ as the equilibrium. Thus, the L contract is self-enforcing if the HQ owns the physical assets.

(B2) Suppose instead that the manager owns the physical assets. As in the case of (A1), the manager has the outside option of $\max\{\tau w_0; y(s^e; 0)\}$ while the HQ has $y(0; G^e) -$

τw_0. For $\{s = T - g^*, g = g^*\text{, no renegotiation}\}$ to be a subgame perfect equilibrium, by Proposition 8, the following conditions have to be satisfied:

$$y(T - g^*; 0) \quad w_0 - w_1; \tag{9}$$

$$\tau w_0 \quad w_0 - w_1; \tag{10}$$

$$y(0; G^*) - \tau w_0 \quad y(T - g^*; G^*) - w_0 + w_1. \tag{11}$$

In addition, w_1 has to satisfy

$$w_1 + \max_s \{y(s; 0) - c(s)\} \quad w_0. \tag{12}$$

Otherwise, the manager could benefit from deviation by choosing the sales effort that maximizes $y(s; 0) - c(s)$ and getting a total payoff of $w_1 + \max_s\{y(s; 0) - c(s)\} > w_0$. It is easy to see that (12) implies (9). Therefore, (10)-(12) are necessary conditions for $\{T - g^*, g^*\text{, no renegotiation}\}$ to be a subgame perfect equilibrium. We can show that they are also sufficient.

Lemma 9 $\{T - g^*, g^*\text{, no renegotiation}\}$ is a subgame perfect equilibrium if and only if there exists some w_1 satisfying (10)-(12).

However, $\{T - g^*, g^*\text{, no renegotiation}\}$ is not the only subgame perfect equilibrium. There are other equilibria that are more likely to be selected. We know the second-best g^* is chosen to maximize $\pi_1 = p[y(T - g; pg) - w] + (1 - p)[(1 - \alpha)y(s; pg) - \beta]$. Let g^0 be the goodwill effort that maximizes $y(T - g; pg)$. It can be shown that $g^0 < g^*$ and $y(T - g^0; pg^0) > y(T - g^*; pg^*)$. These inequalities imply that (10)-(12) still hold when g^* is replaced by g^0. Similar to Lemma 9, we have:

Lemma 10 If there exists some w_1 satisfying (10)-(12), then $\{T - g^0, g^0\text{, no renegotiation}\}$ is a subgame perfect equilibrium.

It remains to be argued that $E^0 = \{T - g^0, g^0\text{, no renegotiation}\}$ is more likely to be selected than $E^* = \{T - g^*, g^*\text{, no renegotiation}\}$. E^0 and E^* have the same equilibrium payoffs. However, compared to E^*, E^0 gives the manager a higher outside option payoff, yields higher surplus for the unit, and gives the HQ the same outside option payoff. Therefore, E^0 gives the manager higher off-equilibrium payoffs than E^*. Note that in the game between the L manager and the HQ, the manager moves first by choosing s and g before any possible renegotiation. If the manager is even slightly concerned about possible uncertainty in the payoffs, the game form, or the execution of the strategies, she will choose E^0 rather than E^*. In this sense, E^0 is more likely to be the outcome of the game. In E^0, however, the second-best effort levels are not implemented.

Combining (B1) and (B2), we have:

Proposition 10 The low-powered incentive contract is self-enforcing ex post if and only if the HQ owns the unit's physical assets.

Propositions 9 and 10 are established under the assumption that the HQ owns the brand name in which goodwill effort is embedded. This is in fact an optimal arrangement. Following the logic of the above analysis, we can show that managers with the low-powered incentive contract should not own any assets, including the brand name. If managers with the high-powered incentive contract have some claim over the ownership of the brand name, they will want to renegotiate the low-powered incentive contract, because they are not given any of L units' revenue under the contract. Therefore, the HQ should be the sole owner of the brand name.

It is useful to compare self enforcement of high-powered and low-powered incentive contracts with that of performance pay and efficiency wage contracts (MacLeod and Malcomson (1996)). In their model, as revenue is unverifiable, performance contracts may not be self-enforcing. In our model, the high-powered contracts may not be self-enforcing, because contracts are incomplete about goodwill. Low-powered contracts, on the other hand, are similar to efficiency wage contracts, as neither is contingent on revenue or profit. As shown by MacLeod and Malcomson (1996), either performance pay contracts or efficiency wage contracts are not self-enforcing depending on market conditions, and their relative optimality follows accordingly. In this paper, contract mix is optimal and multiple ownership arrangements are required to ensure self enforcement of both high-powered and low-powered contracts.

4.4. Contractual remedies versus ownership arrangement

One may well ask the following question: can contractual remedies mimic the ownership arrangement to ensure self enforcement of the high-powered contract? Specifically, while the HQ owns both the unit's physical assets and company goodwill, it could write an ex ante contract that stipulates a payment (denoted by M) from the HQ to the manager once the high-powered contract is not enforced.[27] Note that the actual sales revenue is no longer verifiable when the outside option is taken. This is because the HQ would hire another manager on a fixed-wage contract to finalize the production, and it is impossible for the court to verify the actual revenue without the help from the HQ or the new manager. Thus, the HQ's payment to the manager can only be a fixed payment depending on the expected sales revenue.

There are two conflicting concerns when choosing M. First, the manager should be given incentive to put in sales effort. Second, the HQ should ex post prefer enforcement of the high-powered contract. Specifically, the manager has the option of not putting in any sales effort but still receiving a fixed payment M from the HQ. To provide the manager the effort incentive, $M^* < \alpha^* y(s^*; G^*) - c(s^*)$ is required.[28] That, however, implies that the HQ

[27] The low-powered contract is self-enforcing so long as there is no contractual remedy at all.

[28] The timing for the lump-sum payment β^* is unimportant to the main results of this subsection. For simplicity of notations, it is assumed that all of β^* is paid before the renegotiation stage. The value of the reservation wage does not affect the argument either and thus is again assumed to be 0.

gets more from the outside option than from the high-powered contract ($y(s^*;G^*) - M^* > (1-\alpha^*)y(s^*;G^*)$), and the high-powered contract would be renegotiated. Underlying the above argument is an important feature of the contractual remedies, namely, there is no automatic loss of surplus when the outside option is taken. In contrast, under the optimal ownership arrangement for the high-powered contract, there is a loss of surplus upon the outside option (namely, $y(s^*;0) + y(0;G^*) < y(s^*;G^*)$), which is necessary to ensure self enforcement of the high-powered contract (Proposition 9).

One may further propose the following contractual remedies when the outside option is taken: the HQ pays M^* to the manager and also makes a large enough donation (D^*) to a third party such that the HQ prefers enforcement of the high-powered contract (i.e., $y(s^*;G^*) - M^* - D^* < (1-\alpha^*)y(s^*;G^*)$). However, under some reasonable circumstances, the proposed contractual remedies are inferior to the optimal ownership arrangement.

Without losing generality, suppose that, after the manager makes the sales and goodwill efforts, she may have to quit the business for some benign (for example, family) reasons with certain probability. Interestingly, this uncertainty does not present any problem for the ownership arrangement of the high-powered contract. In case of friendly separation, the manager could sell her asset to a third party and get her expected payoff $\alpha^* y(s^*;G^*)$.[29] In case of unfriendly separation, the manager gets $y(s^*;0)$ and the HQ has $y(0;G^*)$, which prevents renegotiation of the high-powered contract and ensures optimal ex ante incentive for both parties.

Now, we consider contractual remedies in the presence of possible friendly separation. Again, payments in the contract have to be fixed. One reason is that revenue also becomes unverifiable under friendly separation, for a new manager will be hired by the company and it is impossible for the court to verify the actual revenue without the help from the HQ or the new manager. A more important reason is that the incumbent manager will not have as much control over the hiring of the new manager when she does not own any assets as the control she has over the choice of the buyer of her assets when she owns then.[30] In particular, she will not be given the rights of choosing the new manager and dictating the contract with the new manager. Only with these rights can she get $\alpha^* y(s^*;G^*)$ through side payment by the new manager to her.

The contract should include the following payments: under unfriendly separation, M^* to the manager and D^* to a third party; under friendly separation, M^{**} to the manager. These payments should be chosen so that (1) neither party would initiate the unfriendly separation, and (2) friendly separation is allowed to occur when there are good reasons for the manager to initiate it. (1) implies that payments $(M^*; D^*)$ would not be invoked. However, (2) implies that the fixed payment M^{**} would be invoked with a positive probability. Note that,

[29] Note that for the sale of the assets to go through, revenue does not need to be verifiable. However, for a contract contingent on revenue to be enforceable, revenue needs to be verifiable. See Kaufmann and Lafontaine (1994) for empirical evidence on the franchisees' sales of their assets. In general, the partner's right to sell his/her assets is well protected by law, though the sale often needs to be first offered to other partners (see for example Lynch (1989)).

[30] Hiring employees is a residual right of control that can not be contracted on.

whenever the manager is expected to get ¯xed payo® (in the event of friendly separation) with positive probability, the ex ante e®ort incentive will be adversely a®ected. Thus, there is a tension between adequate protection in case of friendly separation and proper ex ante incentive. In contrast, the ownership arrangement of the high-powered contract provides the manager with payo®s that are contingent upon both the manager's e®ort and the nature of the separation.

In summary, given there is severe contractual incompleteness in franchising, the incentive contracts of various power can only be voluntarily enforced in the presence of the corresponding ownership arrangements of the units' physical assets.

5. Conclusion

The contract mix and multiple ownership structure in a franchise both challenge the recent developments in the theory of the ¯rm. With the observation that system-wide goodwill and unit-speci¯c sales activity are crucial to a franchise company, we construct a multi-task model in which one task has the feature of the public good and the other has that of the private good. We show that, when the two tasks are complementary, the principal should o®er a ¯xed-wage contract to some agents and a revenue-sharing contract to the remaining agents. In addition, by incorporating the stylized feature of contractual incompleteness in a franchise and possible ex post unenforceability of ex ante incentive contracts, we show that the ex ante incentive contracts of various power can be made self-enforcing only by the corresponding ownership arrangements.

This paper thus provides the ¯rst theory that explains both contract mix and multiple ownership structure in the franchise. More importantly, it adopts and extends important features from both the multi-task theory and the incomplete-contract theory of the ¯rm. On the one hand, by incorporating the task of the public-good nature, it makes it possible for the multi-task model to explain the optimality of contract mix for ex ante homogeneous agents. On the other hand, it considers ex ante revenue-sharing contracts in settings of contractual incompleteness and explores other roles of ownership structures. In particular, the optimality of the contract mix in the presence of a multi-task framework implies multiple ownership arrangements of complementary assets.

References

[1] Anderlini, Luca. and Felli, L. \Incomplete Written Contracts: Undescribable States of Nature." Quarterly Journal of Economics 1994, 1085-1124.

[2] Berge, C. Topological Spaces. Macmillan, 1963.

[3] Brickley, James A. and Dark, F. H. \The Choice of Organizational Form: The Case of Franchising." Journal of Financial Economics 1987, 401-420.

[4] Dicke, Thomas S. Franchising in America. The University of North Carolina Press: Chapel Hill & London, 1992.

[5] Dnes, Antony W. Franchising: A Case-study Approach. Avebury, 1992.

[6] Gallini, Nancy T., and Lutz, Nancy A. \Dual Distribution and Royalty Fees in Franchising." Journal of Law, Economics and Organization 1992, 471-501.

[7] Grossman, Sanford J. and Hart, Oliver D. \The Costs and Bene¯ts of Ownership: A Theory of Vertical and Lateral Integration." Journal of Political Economy August 1986, 691-719.

[8] Had¯eld, Gillian K. \Problematic Relations: Franchising and Law of Incomplete Contracts." Stanford Law Review 1990, 927-992.

[9] Hart, O. Firms, Contracts, and Financial Structure. Oxford University Press, 1995.

[10] Hart, Oliver, and John Moore. \Property Rights and the Nature of the Firm," Journal of Political Economy 1990, 1119-1158.

[11] Hart, Oliver, Shleifer, Andrei, and Vishny, Robert W. \The Proper Scope of Government: Theory and an Application to Prisons." Mimeo, Harvard, 1996.

[12] Holmstrom, Bengt. \Moral Hazard in Teams." Bell Journal of Economics 1982, 324-40.

[13] Holmstrom, Bengt and Milgrom, Paul. \Aggregation and Linearity in the Provision of Intertemporal Incentives." Econometrica March 1987, 55(2), 303-28.

[14] ——. \Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design." Journal of Law, Economics and Organization 1991, 7, Special Issue, 24-52.

[15] ——. \The Firm as an Incentive System." American Economic Review September 1994, 84(4), 972-91.

[16] Kaufmann, Patrick J. and Lafontaine, F. \Cost of Control: The Source of Economic Rents for McDonald's Franchisees." Journal of Law and Economics 1994, 417-453.

[17] Kostecka, Andrew. Franchising in the Economy, 1985-1987. Washington, D.C.: U.S. Department of Commerce, 1987.

[18] Lafontaine, Francine. \Agency Theory and Franchising: Some Empirical Results." Rand Journal of Economics 1992, v23 n2, 263-283.

[19] _____. \Contractual Arrangements as Signaling Devices: Evidence from Franchising.." Journal of Law, Economics and Organization 1993, 256-289.

[20] _____. and Shaw, Kathryn L. \The Dynamics of Franchise Contracting: Evidence from Panel Data," NBER Working Paper 5585, May 1996.

[21] Lewin, Shira. \Innovation and Authority in Franchise Systems: Toward a Grounded Theory of the Plural Form." mimeo, Harvard University, 1996.

[22] Love, John F. McDonald's: Behind the Arches. Bantam Book: New York, 1986.

[23] Lutz, Nancy A. \Ownership Rights and Incentives in Franchising." Journal of Corporate Finance 1995, 103-131.

[24] Lynch, Robert. The Practical Guide to Joint Ventures and Corporate Alliances: How to Form, How to Organize, How to Operate. New York: Wiley, 1989.

[25] MacLeod, W. Bentley. \Decision, contract, and emotion: some economics for a complex and confusing world." Canadian Journal of Economics November, 1996, XXIX, No. 4.

[26] MacLeod, W. Bentley. \Complexity, Contract and the Employment Relationship." mimeo, Boston College, 1996.

[27] MacLeod, W. Bentley and Malcomsom, James M. \Investment, Holdup, and the Form of Market Contracts." American Economic Review September 1993, 811-837.

[28] Maskin, Eric and J. Tirole. \Unforeseen Contingencies, Property Rights, and Incomplete Contracts." mimeo, Harvard, 1996.

[29] McAfee, R.P. and M. Schwartz (1994): \Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity." American Economic Review 210-230.

[30] Segal, Ilya. \Complexity and Renegotiation: A Foundation for Incomplete Contracts." Mimeo, Harvard, 1996.

[31] Shelton, John P. \Allocative E±ciency vs. `X-E±ciency': Comment." American Economic Review 1967, 1252-58.

Appendix

Proof of Proposition 1: The HQ should request the same level of goodwill effort from all managers of L units.

Let $\{g(j)\}_{j\in L}$ be a profile of goodwill effort levels and $g \equiv \int_L g(j)dj/p$, the average of $\{g(j)\}_{j\in L}$. Since $y(s;G)$ is concave in s, $y(T-g;G)$ is concave is g. Then by Jensen's Inequality,

$$\int_L y(T-g(j);G)dj \le \int_L y(T-g;G)dj.$$

Therefore, the solution to program $(OP-G)$ is to choose $g(j)$ to be a constant. That is, The HQ should request the same level of goodwill effort from all managers of L units. ■

Proof of Lemma 1: Program $(OP-HQ')$ is equivalent to

$$\begin{array}{lll}
\max & py(T-g;G)+\int_{I-L}[y(s(i);G)-c(s(i))-\frac{1}{2}r\sigma^2\alpha(i)^2]di & (OP-HQ'') \\
s.t.: & py_G(T-\frac{G}{p};G)-y_s(T-\frac{G}{p};G)+\int_{I-L}(1-\alpha(i))y_G(s(i);G)di \ge 0 & (IC-G) \\
 & \alpha(i)y_s(s(i);G)=c'(s(i)) & (IC-s)
\end{array}$$

Let us consider program $(OP-HQ')$. Assumptions 1 and 3 say that $y(s;G)$ is concave in s and $\lim_{s\to 0} y_s(s;G)=\infty$. Therefore, incentive compatibility constraint $(OP-s)$ can be replaced by

$$\alpha(i)y_s(s(i);G)=c'(s(i)).$$

By Proposition 1, incentive compatibility constraint $(OP-G)$ becomes

$$g=\arg\max_g py(T-\frac{G}{p};G)+\int_{I-L}[(1-\alpha(i))y(s(i);G)-\beta(i)]di;$$

where, $G=pg$. It is easy to show that, since $y(s;G)$ is concave in $(s;G)$,

$$py(T-\frac{G}{p};G)+\int_{I-L}[(1-\alpha(i))y(s(i);G)-\beta(i)]di$$

is concave in G. Its derivative with respect to G is

$$py_G(T-\frac{G}{p};G)-y_s(T-\frac{G}{p};G)+\int_{I-L}(1-\alpha(i))y_G(s(i);G)di;$$

which decreases with G, by Assumption 3, goes to ∞ as $G\to 0$, and goes to $-\infty$ as $G\to pT$. Therefore, $(OP-G)$ can be replaced by

$$py_G(T-\frac{G}{p};G)-y_s(T-\frac{G}{p};G)+\int_{I-L}(1-\alpha(i))y_G(s(i);G)di=0. \qquad (FOC-G)$$

The objective function of program $(OP - HQ')$ is now

$$py(T-g;G)+\int_{I-L}\left[y(s(i);G)-c(s(i))-\frac{1}{2}r\sigma^2\alpha(i)^2\right]di,$$

which is also concave in G and the derivative of which with respect to G is positive for G satisfying constraint $(FOC - G)$. Therefore, program $(OP - HQ')$ can be rewritten as

$$\begin{array}{lll}
\max & py(T-g;G)+\int_{I-L}[y(s(i);G)-c(s(i))-\frac{1}{2}r\sigma^2\alpha(i)^2]di & (OP-HQ'') \\
s.t. & py_G(T-\frac{G}{p};G)-y_s(T-\frac{G}{p};G)+\int_{I-L}(1-\alpha(i))y_G(s(i);G)di \geq 0 & (IC-G) \\
 & \alpha(i)y_s(s(i);G)=c'(s(i)) & (IC-s)
\end{array}$$

When we change the equality sign in $(FOC - G)$ to $\geq$ in $(IC - G)$, we expand the feasible region of program $(OP - HQ')$ to the left along the G-direction, as the left hand side of $(FOC - G)$ decreases with G. This does not change the optimum because the objective function of $(OP - HQ')$ increases with G in the expanded feasible region. ■

Proof of Lemma 2: Let $\alpha(i) = c'(s)/y_s(s;G)$. Then, given G, the integrand in the objective function of program $(OP - HQ'')$,

$$y(s(i);G)-c(s(i))-\frac{1}{2}r\sigma^2\alpha(i)^2,$$

and the integrand in constraint $(IC - G)$,

$$(1-\alpha(i))y_G(s(i);G),$$

are both concave functions of $s(i)$.

Let $\phi(s;G) \equiv c'(s)/y_s(s;G)$. Then $(IC - s)$ implies $\alpha(i) = \phi(s(i);G)$, which by Assumption 4 is convex in $s(i)$. Substitute $\alpha(i) = \phi(s(i);G)$ into the objective function and constraint $(IC - G)$. Then the integrand in the objective function,

$$y(s(i);G)-c(s(i))-\frac{1}{2}r\sigma^2\alpha(i)^2,$$

becomes a concave function of $s(i)$. The integrand in constraint $(IC - G)$,

$$(1-\alpha(i))y_G(s(i);G),$$

is also concave in $s(i)$ in the convex range $\{s(i) : c'(s(i))/y_s(s(i);G) \leq 1\}$, because $(1-\alpha(i))$ is non-negative, concave and decreasing in $s(i)$, and $y_G(s(i);G)$ is, by Assumption 5, positive, concave and increasing in $s(i)$; the product of two non-negative concave functions is concave if one of them is increasing and the other decreasing. ■

Proof of Proposition 2: The HQ should offer the same high-powered contract to all managers of H units.

Given a profile of sales effort levels, $\{s(i)\}_{i\in I-L}$, let $s \equiv \int_{I-L} s(i)di/(1-p)$, the average of $\{s(i)\}_{i\in I-L}$. Then, for any given G, Jensen's inequality implies

$$\int_{I-L}[y(s(i),G) - c(s(i)) - \frac{1}{2}r\sigma^2\alpha(i)^2]di \le \int_{I-L} y(s,G) - c(s) - \frac{1}{2}r\sigma^2\alpha^2 di;$$

and

$$\begin{aligned} &py_G(T - \tfrac{G}{p},G) - y_s(T - \tfrac{G}{p},G) + \textstyle\int_{I-L}(1-\alpha(i))y_G(s(i),G)di \\ \le\; &py_G(T - \tfrac{G}{p},G) - y_s(T - \tfrac{G}{p},G) + \textstyle\int_{I-L}(1-\alpha)y_G(s,G)di; \end{aligned}$$

where $\alpha = \phi(s,G)$. Therefore, choosing the same α and s for all managers of H units is better than choosing different ones. That is, the HQ should offer the same high-powered contract to all managers of H units. ■

Proof of Lemma 4: s and G are continuous functions of (α,p) on $[0,1]^2$.

By Lemma 3, we only need to show continuity (1) at $\alpha = 0$ and (2) at $p = 0$.
(1) We first prove that $\lim_{(\alpha,p)\to(0,p')} s(\alpha,p) = T$, for all p'. By (FOC $-$ s),

$$\alpha y_s(s(\alpha,p),G) = c'(s(\alpha,p)).$$

Therefore,

$$0 \le c'(s(\alpha,p)) - c'(T) = \alpha y_s(s(\alpha,p),G) \le \alpha y_s(T,T) \to 0$$

as $\alpha \to 0$. The last inequality holds because $s(\alpha,p) \ge T$, $G \le T$, and $y_s(s,G)$ increases with G but decreases with s. Since $c'' > 0$ as $t > T$, $(c')^{-1}$ is continuous in $[0,\infty)$ with $(c')^{-1}(0) = T$. Therefore, $c'(s(\alpha,p)) - c'(T) \to 0$ implies $\lim_{(\alpha,p)\to(0,p')} s(\alpha,p) = T$. The definition of $s(0,p')$ implies that $s(0,p') = T$. Therefore, s is continuous at $\alpha = 0$.

When $p > 0$,

$$\frac{\partial^2 \Pi_1}{\partial G^2} = py^L_{GG} - 2y^L_{sG} + \frac{1}{p}y^L_{ss} + (1-p)(1-\alpha)y^H_{GG} < 0$$

by (FOC $-$ G) and the concavity of y, where a function with a superscript H (L, resp.) means that it is evaluated at (s,G) ($(T - \frac{G}{p},G)$, resp.). Therefore (FOC $-$ G) implies G is a differentiable function of (s,α,p) when $p > 0$, by the Implicit Function Theorem. Since we have shown that s is continuous at $\alpha = 0$, so is G.

When $p = 0$,

$$G(\alpha,p) - G(0,0) = pg \to 0.$$

Note that the above argument also showed the continuity of s and G at $(\alpha,p) = (0,0)$. (2) First, G is continuous at $p = 0$ because $G(\alpha,0) = 0$ for all α and

$$G(\alpha,p) = pg(\alpha,p) \le pT \to 0 \text{ as } p \to 0, \text{ for all } \alpha.$$

$s(\alpha,p)$ is defined by

$$\alpha y_s(s(\alpha,p),G) - c'(s(\alpha,p)) = 0.$$

At $(\alpha' \neq 0; p = 0)$,

$$\alpha' y_s(s(\alpha'; 0); 0) - c'(s(\alpha'; 0)) = 0.$$

If $y_s(T; 0) \neq 0$, then $s(\alpha'; 0) > T$ and thus $\alpha' y_{ss}(s(\alpha'; 0); 0) - c''(s(\alpha'; 0)) < 0$. By the Implicit Function Theorem, s is a continuous function of $(\alpha; G)$ at $(\alpha' \neq 0; p = 0)$. Since we have shown that G is continuous at $p = 0$, s is also continuous at $p = 0$.

If $y_s(T; 0) = 0$, then $s(\alpha'; 0) = T$ for all α'.

$$|c'(s(\alpha; p)) - c'(T)| = \alpha y_s(s(\alpha; p); G(\alpha; p)) \quad \alpha y_s(T; G(\alpha; p)) \to 0 \text{ as } p \to 0.$$

Since $(c')^{-1}$ is continuous in $[0; \infty)$ with $(c')^{-1}(0) = T$, $s(\alpha; p) \to T$ as $(\alpha; p) \to (\alpha'; 0)$, i.e., s is continuous at $p = 0$. ■

Proof of Proposition 3: For any $p \in [0; 1)$, the optimal $\alpha \in (0; 1)$.

When $p = 0$, α is chosen to maximize

$$\begin{array}{ll} \max & y(s; 0) - c(s) - \frac{1}{2} r\sigma^2\alpha^2 \\ s.t. & \alpha y_s(s; 0) - c'(s) = 0. \end{array}$$

Apply the implicit function theorem to the constraint. We have,

$$\frac{ds}{d\alpha} = \frac{y_s}{\alpha(c''(s) - y_{ss})} > 0 \text{ for } \alpha > 0.$$

Then,

$$\begin{aligned} \frac{d\pi}{d\alpha} &= (y_s - c')\frac{ds}{d\alpha} - r\sigma^2\alpha \\ &= (1 - \alpha)y_s\frac{ds}{d\alpha} - r\sigma^2\alpha \\ &= \frac{1-\alpha}{\alpha}\frac{y_s^2}{c'' - y_{ss}} - r\sigma^2\alpha. \end{aligned}$$

Therefore, as $\alpha \to 0$, $\frac{d\pi}{d\alpha} \to \infty$, and at $\alpha = 1$, $\frac{d\pi}{d\alpha} = -r\sigma^2\alpha < 0$. Consequently, the optimal $\alpha \in (0; 1)$.

Now consider $p \in (0; 1)$. At $\alpha = 1$, (2) becomes

$$\frac{d\Pi_{-0}}{d\alpha} = -(1 - p)r\sigma^2 + (1 - p)y_G^H\frac{dG}{d\alpha}.$$

Apply the implicit function theorem to (FOC $-$ s) and (FOC $-$ G). We have, at $\alpha = 1$,

$$\frac{dG}{d\alpha} = \frac{1}{|J|}(1 - p)y_G^H(y_{ss}^H - c'');$$

where, the Jacobian matrix

$$J = \begin{pmatrix} -c''(s) & 0 \\ 0 & py_{GG}^L - 2y_{sG}^L + \frac{1}{p}y_{ss}^L \end{pmatrix} + \begin{pmatrix} \alpha y_{ss}^H & \alpha y_{sG}^H \\ (1 - p)(1 - \alpha)y_{sG}^H & (1 - p)(1 - \alpha)y_{GG}^H \end{pmatrix}$$

has positive determinant. Therefore, at $\alpha = 1$, $\frac{dG}{d\alpha} < 0$, which implies that $\frac{d\Pi_0}{d\alpha} < 0$. Thus the optimal α is not 1.

As $\alpha \to 0$,

$$\begin{aligned} \frac{ds}{d\alpha} &\to -\frac{1}{|J|} y_s^H [p y_{GG}^L - 2 y_{sG}^L + \frac{1}{p} y_{ss}^L + (1-p) y_{GG}^H] > 0; \\ \frac{dG}{d\alpha} &\to \frac{1}{|J|}(1-p) y_{sG}^H y_s^H > 0. \end{aligned}$$

These imply that $\lim_{\alpha \to 0} \frac{d\Pi_0}{d\alpha} > 0$. Therefore, the optimal α is not 0. ■

Proof of Proposition 4: For any given $p \in (0,1)$, $\pi^L < \pi^H$ at the corresponding HQ's optimal choice of α, denoted by $\alpha^*(p)$.

By the definition of π^L,

$$\pi^L(\alpha = \alpha^*(p)) = y(T - g(\alpha^*), pg(\alpha^*)) \le \max_g y(T - g, pg).$$

Since $y(s, G)$ increases with s and $p < 1$,

$$\max_g y(T - g, pg) < \max_g p y(T - g, pg) + (1-p) y(T, pg) = \pi(\alpha = 0).$$

By the definition of $\alpha^*(p)$,

$$\pi(\alpha = 0) \le \pi(\alpha = \alpha^*(p)) = p\pi^L(\alpha = \alpha^*(p)) + (1-p)\pi^H(\alpha = \alpha^*(p)).$$

Combining the above three inequalities, we have:

$$\pi^L(\alpha = \alpha^*(p)) < p\pi^L(\alpha = \alpha^*(p)) + (1-p)\pi^H(\alpha = \alpha^*(p)),$$

which implies that $\pi^L(\alpha = \alpha^*(p)) < \pi^H(\alpha = \alpha^*(p))$. ■

Proof of Lemma 5: If $y(s, 0) \not\equiv 0$, then $\lim_{p \to 0} \frac{\partial \Pi_0}{\partial p} > 0$ for $\alpha < 1$.

If $y(s, 0)$ is not always zero, then the concavity, the monotonicity, and the non- negativity of y implies that $y(s, 0) > 0$ for all $s > 0$. Let $\alpha < 1$. By (3),

$$\frac{d\Pi_0}{dp} = (\pi^L - \pi^H) + g y_s^L + (1-p)\alpha y_G^H \frac{\partial G}{\partial p} + (1-p)(1-\alpha) y_s^H \frac{\partial s}{\partial p}.$$

(FOC − G) implies that,

$$y_s(T - g, G) \ge (1-p)(1-\alpha) y_G(s, G).$$

Since $s \ge T$ and $y_{sG} > 0$,

$$(1-p)(1-\alpha) y_G(s, G) \ge (1-p)(1-\alpha) y_G(T, G);$$

the right hand side of which $\to \infty$ because $G = pg \leq pT$, and Assumption 3 says that $\lim_{G \to 0} y_G = \infty$. Therefore, $y_s(T - g; G) \to \infty$, which implies $g \to T$. Then, by Assumption 6, the substitution effect in (3), $gy_s^L \to \infty$,

$$\lim_{p \to 0} gy_s^L = T \lim_{p \to 0} y_s(T - g; G) = T \lim_{(s;G) \to (0;0)} y_s(s; G) = \infty.$$

In (3), $\pi^L - \pi^H$ is bounded. Then to determine the sign of $\frac{d\Pi_0}{dp}$ as $p \to 0$, it is sufficient to show that $\frac{\partial G}{\partial p} > 0$ and $\frac{\partial s}{\partial p} > 0$ as $p \to 0$. By Lemma 3, it suffices to show that $\frac{\partial^2 \Pi_1}{\partial p \partial G} > 0$. Substitute (FOC $-$ G) into $\frac{\partial^2 \Pi_1}{\partial p \partial G}$ and rearrange. Then

$$(1 - p)\frac{\partial^2 \Pi_1}{\partial p \partial G} = y_G^L + (1 - p)gy_{sG}^L - \frac{(1 - p)}{p}gy_{ss}^L - y_s^L, \tag{A1}$$

in which only the last term is negative. By Assumption 6, y_s is weakly convex. Then

$$y_s(T; G) - y_s(T - g; G) \geq gy_{ss}(T - g; G),$$

in which $y_s(T - g; G) \to \infty$. Therefore,

$$-gy_{ss}^L = -gy_{ss}(T - g; G) \to \infty.$$

Rearranging (A1) yields

$$\begin{aligned}(1 - p)\tfrac{\partial^2 \Pi_1}{\partial p \partial G} &= y_G^L + (1 - p)gy_{sG}^L - \tfrac{(1-2p)}{p}gy_{ss}^L - gy_{ss}(T - g; G) - y_s^L \\ &\geq y_G^L + (1 - p)gy_{sG}^L - \tfrac{(1-2p)}{p}gy_{ss}^L - y_s(T; G) \to \infty.\end{aligned}$$

In summary, we have shown $\lim_{p \to 0} \frac{\partial \Pi_0}{\partial p} > 0$ for $\alpha < 1$. ■

Proof of Proposition 5: It is optimal for the company to have some L units.

Corollary 1 says that the optimal $(\alpha; p)$ exists. Let it be $(\alpha^*; p^*)$. We wish to show that $p^* > 0$. Suppose instead that $p^* = 0$. By Proposition 4, $\alpha^* < 1$. By Lemma 5, $\lim_{p \to 0} \frac{\partial \Pi_0}{\partial p}(\alpha^*; p) > 0$. Thus there exists some positive p such that $\Pi_0(\alpha^*; p) > \Pi_0(\alpha^*; p^*)$. This is a contradiction. This completes the proof of the Proposition. ■

Proof of Proposition 6: It is more likely for the HQ to have H units when the managers are less risk averse, or when the uncertainty about the revenue is smaller, or when marginal cost of effort is smaller.

We first consider the optimal choice of α and s when only an infinitesimal proportion of units are given H contracts, i.e., $p = 1$. In this case, (FOC $-$ G) becomes

$$y_G(T - G; G) - y_s(T - G; G) = 0,$$

which implies that G does not depend on α. Therefore, α is chosen to

$$\begin{aligned} \pi^H &= \max \ y(s;G) - Kc(s) - \tfrac{1}{2} r\sigma^2\alpha^2 \\ & \quad s.t.: \ \alpha y_s(s;G) - Kc'(s) = 0 \qquad (IC) \end{aligned}$$

where K is a cost parameter.

The Lagrangian of the program that chooses the optimal α and s is

$$L = y(s;G) - Kc(s) - \frac{1}{2} r\sigma^2\alpha^2 + \lambda[\alpha y_s(s;G) - Kc'(s)].$$

Differentiation yields

$$\begin{aligned} \tfrac{\partial L}{\partial \alpha} &= -r\sigma^2\alpha + \lambda y_s(s;G); \\ \tfrac{\partial L}{\partial s} &= (1-\alpha) y_s(s;G) + \lambda[\alpha y_{ss}(s;G) - Kc''(s)]. \end{aligned}$$

By the incentive compatibility constraint, $s \geq T > 0$. Therefore, $\frac{\partial L}{\partial s} = 0$. If $\alpha = 1$, then $\frac{\partial L}{\partial s} = 0$ implies $\lambda = 0$, which in turn implies $\frac{\partial L}{\partial \alpha} < 0$. This is a contradiction. If $\alpha = 0$, the incentive compatibility constraint implies that $s = T$. Therefore, $\frac{\partial L}{\partial s} = y_s(T;G) > 0$. This is again a contradiction. Therefore, the optimal $\alpha \in (0;1)$ and $\frac{\partial L}{\partial \alpha} = 0$, which implies that $\lambda > 0$ and the optimal $s > T$.

By the envelope theorem, we have, $\frac{d\pi^H}{dr} = -\frac{1}{2}\sigma^2\alpha^2 < 0$, $\frac{d\pi^H}{d\sigma} = -r\sigma\alpha^2 < 0$ and $\frac{d\pi^H}{dK} = -c(s) - \lambda c'(s) < 0$.

Since at $p = 1$, $\pi^L = y(T - G;G)$ and $gy_s^L = Gy_s(T - G;G)$ do not depend on σ or K, the proposition follows from equation (3) and the above results. ■

Proof of Proposition 7: It is more likely for the HQ to have H units when T is smaller.

This proposition needs additional assumptions:

$$y(s;0) > 0 \ \text{ for } \ s > 0 \ \text{ and } \lim_{(s;G) \to (0;0)} Gy_G(s;G) = 0. \qquad (A2)$$

Suppose the cost function is now $c = c(s + g - T)$ with c being strictly increasing and convex. It suffices to show that $\frac{d\pi}{dp}|_{p=1} < 0$ for sufficiently small T. As $T \to 0$, $\pi^L = y(T - g;T) \to y(0;0)$. When $p = 1$, $gy_s^L = Gy_s(T - G;G)$. By (FOC $-$ G), This is $Gy_G(T - G;G)$, which, by Assumption (A2), approaches 0 as both G and T go to 0.

$$\begin{aligned} \pi^H &= \max_\alpha \ y(s;G) - c(s - T) - \tfrac{1}{2} r\sigma^2\alpha^2 \\ & \quad s.t.: \ \alpha y_s(s;G) - c'(s - T) = 0. \end{aligned}$$

As $T \to 0$, π^H approaches,

$$\begin{aligned} \pi^H(T = 0) &= \max_\alpha \ y(s;0) - c(s) - \tfrac{1}{2} r\sigma^2\alpha^2 \\ & \quad s.t.: \ \alpha y_s(s;0) - c'(s) = 0; \end{aligned}$$

which is independent of T and greater than $y(0;0)$. Therefore,

$$\lim_{T \to 0} \frac{d\pi}{dp}\Big|_{p=1} = y(0;0) - \pi^H(T = 0) < 0;$$

that is, $\frac{d\pi}{dp}\big|_{p=1} < 0$ for sufficiently small T. ■

Proof of Lemma 6: Suppose that the manager owns the physical asset. The H contract is self-enforcing if conditions

$$\max_s \{y(s;0) - c(s)\} + w_1 \geq \alpha^* y(s^*;G^*) - c(s^*) + \beta^* = w_0; \tag{6}$$

$$\tau w_0 + w_1 < \alpha^* y(s^*;G^*) - c(s^*) + \beta^* = w_0; \tag{7}$$

$$\max_g \{py(T - g;pg) + (1-p)y(0;pg)\}$$

$$\geq py(T - g^*;G^*) + (1-p)[y(s^*;G^*) - c(s^*)] + (1-p)(w_1 + \tau w_0 - w_0): \tag{8}$$

are met.

To make the ex ante contract $(\alpha^*;\beta^*)$ self-enforcing in the incomplete-contracting framework, we need to prove: (1) given that $(s^*;G^*)$ is chosen the contract is not renegotiated at $t = 2$, and (2) $(s^*;G^*)$ is an Nash equilibrium when the managers choose effort levels at $t = 1$ in anticipation of possible renegotiation of the contract. Note that part (1) is met under the following conditions:

$$\alpha^* y(s^*;G^*) + (\beta^* - w_1) \geq y(s^*;0); \tag{A3}$$

$$\alpha^* y(s^*;G^*) + (\beta^* - w_1) \geq \tau w_0; \tag{A4}$$

$$(1-\alpha^*)y(s^*;G^*) - (\beta^* - w_1) \geq y(0;G^*) - \tau w_0: \tag{A5}$$

It is easy to show that (A3), (A4) and (A5) are implied by (6), (7) and (8), respectively.

Now we consider the investment stage to show that $(s^*;G^*)$ is an Nash equilibrium. We first consider the manager's sales effort given the goodwill effort G^*.

1. If the manager chooses s^*, then renegotiation does not occur under (A3) - (A5). Her total payoff is $\alpha^* y(s^*;G^*) - c(s^*) + \beta^* = w_0$ [31] due to the individual rationality constraint.

2. If the manager deviates by choosing s', then her total payoff is

   (a) $\alpha^* y(s';G^*) - c(s') + \beta^*$ if s' does not lead to renegotiation.

[31] As there is no contract renegotiation, the manager's total payoff is equal to w_1 from the HQ before the realization of the revenue plus $\alpha^* y(s^*;G^*) + (\beta^* - w_1)$ from the HQ after the realization minus her effort cost $c(s^*)$.

(b) $y(s'; 0) - c(s') + w_1$ if $y(s'; 0) > \max\{\alpha^* y(s'; G^*) + (\beta^* - w_1), \tau w_0\}$ (the manager initiates renegotiation by threatening to work for herself).

(c) $\tau w_0 - c(s') + w_1$ if $\tau w_0 > \max\{\alpha^* y(s'; G^*) + (\beta^* - w_1), y(s'; 0)\}$ (the manager initiates renegotiation by threatening to work for others).

(d) $y(s'; G^*) - y(0; G^*) + \tau w_0 - c(s') + w_1$ if $y(0; G^*) - \tau w_0 > (1 - \alpha^*) y(s'; G^*) - (\beta^* - w_1)$ (the HQ initiates renegotiation).

To induce the manager to make the second-best sales effort s^*, we need to ensure that, given the goodwill effort G^*, her payoff when choosing s^* is higher than that when choosing s'. Note that, by the definition of s^* in Section 3, $\alpha^* y(s'; G^*) - c(s') + \beta^* \le \alpha^* y(s^*; G^*) - c(s^*) + \beta^*$, which implies that the manager does not benefit from deviation 2(a). Condition (6) ensures that deviation 2(b) is not worthwhile. The manager will not take deviation 2(c) if $\tau w_0 - c(s') + w_1 \le \alpha^* y(s^*; G^*) - c(s^*) + \beta^*$; which is guaranteed by (7). Finally, deviation 2(d) is not profitable as $y(s'; G^*) - y(0; G^*) + \tau w_0 - c(s') + w_1 < \alpha^* y(s'; G^*) - c(s') + \beta^*$; and the same logic for the inferiority of 2(a) applies.

We next consider the HQ's effort choice given the sales effort s^*:

1. If the HQ chooses G^*, then renegotiation does not occur under (A3) - (A5). Its total payoff is $p[y(T - g^*; G^*) - w_0] + (1 - p)[(1 - \alpha^*) y(s^*; G^*) - \beta^*]$.

2. If the HQ deviates by choosing $G' = pg'$, then its total payoff is

   (a) $p[y(T - g'; G') - w_0] + (1 - p)[(1 - \alpha^*) y(s^*; G') - \beta^*]$ if G' does not lead to renegotiation.

   (b) $p[y(T - g'; G') - w_0] + (1 - p)[y(0; G') - \tau w_0 - w_1]$ if $y(0; G') - \tau w_0 > (1 - \alpha^*) y(s^*; G') - (\beta^* - w_1)$ (the HQ initiates renegotiation).

   (c) $p[y(T - g'; G') - w_0] + (1 - p)[y(s^*; G') - \max\{y(s^*; 0), \tau w_0\} - w_1]$
   if $\max\{y(s^*; 0), \tau w_0\} > \alpha^* y(s^*; G') + (\beta^* - w_1)$ (the manager initiates renegotiation).

To induce the HQ to make the second-best goodwill effort G^*, we need to ensure that, given the sales effort s^*, its payoff from choosing G^* is higher than that from choosing G'. Note that, by the definition of G^* in Section 3, it is not profitable for the HQ to have deviation 2(a). The HQ's payoff under deviation 2(c) is even lower than that under deviation 2(a), which implies that deviation 2(c) is unprofitable either. Finally, deviation 2(b) is not worthwhile if

$$\begin{aligned} \max_g \{py(T - g; pg) + (1 - p)[y(0; pg) - \tau w_0 - w_1]\} \\ \le py(T - g^*; G^*) + (1 - p)[(1 - \alpha^*) y(s^*; G^*) - \beta^*] \\ = py(T - g^*; G^*) + (1 - p)[y(s^*; G^*) - c(s^*)] - (1 - p) w_0; \end{aligned}$$

which is guaranteed by (8). ■

Proof of Lemma 7: Let $k_1 = k_2 = 0$. Then for sufficiently large k, there exists some w_1 such that (6)-(8) hold with strict inequality.

With $k_1 = k_2 = 0$, (6)-(8) become:

$$w_1 \geq w_0;$$

$$\delta w_0 + w_1 \geq w_0;$$

$$\max_g pkz(T - g; G) \geq pkz(T - g^*; G^*) + (1 - p)[kz(s^*; G^*) - c(s^*)] + (1 - p)(w_1 + \delta w_0 - w_0).$$

Choose $w_1 = (1 - \delta)w_0 - \epsilon$ for some $\epsilon > 0$. Then the first two inequalities hold strictly. The third inequality, divided by k, becomes:

$$\max_g pz(T - g; G) \geq pz(T - g^*; G^*) + (1 - p)\left(z(s^*; G^*) - \frac{c(s^*)}{k}\right) - \frac{1}{k}(1 - p)\epsilon; \qquad (8').$$

The left-hand-side of $(8')$ is bounded from above by $z(T; T)$. If we can show $pz(T - g^*; G^*) + (1 - p)\left(z(s^*; G^*) - \frac{c(s^*)}{k}\right) \to \infty$ as $k \to \infty$, then $(8')$ holds for sufficiently large k.

Consider the HQ's optimization problem:

$$\begin{array}{lll} \max_{p;\alpha} & \pi_0 = pz\left(T - \frac{G}{p}; G\right) + (1 - p)\left(z(s; G) - \frac{c(s)}{k}\right) & \\ s.t. & \alpha k z_s(s; G) - c'(s) = 0 & (FOC - s) \\ & pz_G(T - \frac{G}{p}; G) - z_s(T - \frac{G}{p}; G) + (1 - p)(1 - \alpha)z_G(s; G) = 0 & (FOC - G) \end{array}$$

Let us first fix $p \in (0; 1)$ and $\alpha \in (0; 1)$. From $(FOC - s)$ and $(FOC - G)$, it is a standard exercise to show that $\frac{ds}{dk} > 0$ and $\frac{dG}{dk} > 0$. (The computation is similar to that of $\frac{dG}{d\alpha}$ in the proof of Proposition 3.) Then $\alpha z(s; G) - \frac{c(s)}{k}$ increases with k, because its derivative with respect to k is, by the envelope theorem, $\alpha z_G \frac{dG}{dk} + \frac{c(s)}{k^2} > 0$. Furthermore, $\lim_{k\to\infty} s = \infty$. Otherwise, $\lim_{k\to\infty} s = s_\infty < \infty$, and $\alpha z_s(s_\infty; G_\infty) = \lim_{k\to\infty} \frac{c(s)}{k} = 0$ by $(FOC - s)$. This is a contradiction because $z_{sG} > 0$. Therefore, $z(s; G) - \frac{c(s)}{k} = (1 - \alpha)z(s; G) + \left(\alpha z(s; G) - \frac{c(s)}{k}\right) \to \infty$ as $k \to \infty$. In summary, we have shown that for fixed p and α, the HQ's objective function $\pi_0 \to \infty$ as $k \to \infty$. It follows immediately that when p and α are optimally chosen, π_0 is even larger. ■

Proof of Lemma 8: Suppose for some k and w_1, (6)-(8) hold with strict inequality as $k_1 = k_2 = 0$. Then (6)-(8) also hold for sufficiently small k_1 and k_2.

The conclusion about (6) and (7) is very easy to see. The left-hand-side of (8) is $l(p; k_1; k_2) \equiv \max_g \{py(T - g; pg) + (1 - p)y(0; pg)\}$. By the "theorem of the maximum" (Berge (1963)), the optimal g is an upper hemi-continuous function of $(p; k_1; k_2)$ and thus $l(p; k_1; k_2)$ is continuous in $(p; k_1; k_2)$. On the right-hand-side of (8),

$$\begin{array}{lll} & p(T - g^*; G^*) + (1 - p)[y(s^*; G^*) - c(s^*)] = & \\ \max & py(T - \frac{G}{p}; G) + (1 - p)[y(s; G) - c(s)] & (OP - HQ3) \\ s.t. & \alpha y_s(s; G) - c'(s) = 0 & (FOC - s) \\ & py_G(T - \frac{G}{p}; G) - y_s(T - \frac{G}{p}; G) + (1 - p)(1 - \alpha)y_G(s; G) = 0 & (FOC - G) \end{array}$$

To prove the inequality, we first perform an exercise similar to Lemmas 1 and 2. Let $\phi(s;G) \equiv c'(s)/y_s(s;G)$. Then $(FOC-s)$ becomes $\alpha = \phi(s;G)$. By Assumption 4, ϕ is convex in s. Substitute $\alpha = \phi(s;G)$ into the objective function and constraint $(FOC-G)$ in $(OP-HQ3)$. Then $(OP-HQ3)$ becomes

$$\begin{aligned} \max \quad & py(T - \tfrac{G}{p};G) + (1-p)[y(s;G) - c(s) - \tfrac{1}{2}r\sigma^2\phi(s;G)^2] & (OP-HQ4) \\ s.t.: \quad & py_G(T - \tfrac{G}{p};G) - y_s(T - \tfrac{G}{p};G) + (1-p)(1-\phi(s;G))y_G(s;G) \geq 0 & (FOC-G) \end{aligned}$$

The reason why we can change the equality in $(FOC-G)$ to inequality is the same as that offered in Lemma 1. Now, given $(p;G)$, $(OP-HQ4)$ is a concave program that chooses the optimal s. The solution $s = s(p;G;k_1;k_2)$ is differentiable. Substitute the solution into the objective function. We have an unconstrained optimization problem[32]

$$\max_{p;G} f(p;G;k_1;k_2);$$

where f is differentiable. Again, by the "theorem of the maximum", $(p;G)$ are upper hemi-continuous functions of $(k_1;k_2)$; that is

$$S \equiv \{(p;G;k_1;k_2) : (p;G) = \arg\max_{p;G} f(p;G;k_1;k_2)\}$$

is a closed set.

Now, we are ready to prove (8) by contradiction. Denote the left (right)-hand side of (8) by LHS (RHS). Suppose there exists a sequence $(p_n;G_n;k_{1n};k_{2n}) \in S$ such that $\lim_{n\to\infty}(p_n;G_n;k_{1n};k_{2n}) = (p_0;G_0;0;0) \in S$, and $\lim_{n\to\infty}(LHS - RHS) \geq 0$. Then,

$$\lim_{n\to\infty} LHS = l(p_0;0;0);$$

$$\lim_{n\to\infty} RHS = p_0 y(T - g_0;G_0) + (1-p)[y(s_0;G_0) - c(s_0)] + (1-p)(w_1 + \tau w_0 - w_0):$$

Lemma 7 says that $\lim_{n\to\infty} LHS < \lim_{n\to\infty} RHS$; i.e., (8) holds with strict inequality as $k_1 = k_2 = 0$. This is a contradiction. ■

Proof of Lemma 9: $\{T - g^*; g^*;$ no renegotiation$\}$ is a subgame perfect equilibrium if and only if there exists some w_1 satisfying (10)-(12).

By (12), the manager will not choose $s > T$ because such a costly action does not improve her payoff. (12) also implies that $y(T;0) \leq w_0 - w_1$. Therefore, if the manager chooses $s \leq T$, her outside option payoff will not be greater than her payoff under the contract and hence she will not be in a position to initiate renegotiation. Given other L manager's choice of $g = g^*$, (11) implies that the HQ will not want to renegotiate the contract either for $s \geq T - g^*$. Therefore, the manager's payoffs are the same for all $s \in [T - g^*; T]$. For $s < T - g^*$, the total surplus $y(s;G^*)$ is smaller than that for $s = T - g^*$, and thus the manager cannot be better-off than choosing $s = T - g^*$. ■

[32] The constraint that $p \in [0;1]$ does not affect the argument and is thus omitted.