INCENTIVE CONTRACTING AND THE FRANCHISE DECISION

FRANCINE LAFONTAINE

University of Michigan Business School
Ann Arbor, MI 48109, USA
and NBER

and

MARGARET E. SLADE

Department of Economics
The University of British Columbia
Vancouver, BC. V6T 1Z1, Canada
and GREQAM

September 1998

ABSTRACT

This paper examines theoretical predictions and econometric evidence concerning franchise contracting and sales-force compensation. We identify a number of factors that, according to theory, ought to influence the contracts that are written between principals and agents. For each factor, we construct the simplest theoretical model that is capable of capturing what we feel to be its essence. The comparative statics from the theoretical exercise are then used to organize our discussion of the empirical evidence, where the evidence is taken from published studies that have attempted to assess each factor's effect on the power of agent incentives. We also discuss theoretical issues and empirical results pertaining to a few topics that have been addressed in the literature but that do not fit easily into our simple modeling framework.

Key words: Agency costs, moral hazard, contract choice, franchising, retailing, organizational form, sales-force compensation, vertical integration, vertical separation

JEL Classifications: L2, L8, D3, D4

This paper will appear as chapter 2 in *Advances in Business Applications of Game Theory*, K. Chatterjee and W. Samuelson (eds.) Kluwer Academic Press. It unifies and expands our joint work, Lafontaine and Slade [1996 and 1997].

I: Introduction

The modern theory of the internal organization of firms -- the ownership, management, and structure of production -- has its roots in the writings of Knight [1921] and Coase [1937]. Knight emphasized the role of risk and uncertainty and the need to insure workers and consolidate managerial-decision making, whereas Coase focused on the costs of transacting in different organizational environments, particularly the costs of writing contracts. Over time, these notions have been expanded and formalized. Moreover, in the process, two distinct but related branches of literature have emerged. The first concentrates on the tradeoff that a principal must make between providing an agent with insurance against risk and giving that agent incentives to work efficiently (e.g., Williamson [1971], Alchian and Demsetz [1972], Mirlees [1976], and Holmstrom [1982]), whereas the second emphasizes the market failures that accompany relationship-specific assets and the associated need to assign property and residual-decision rights correctly (e.g., Klein, Crawford, and Alchian [1978], Williamson [1979, 1983], Grossman and Hart [1986], and Hart and Moore [1990]).

On the empirical side, efforts to test these theories have been channeled into areas that satisfy two criteria. First, the institutional regularities must correspond to the assumptions that underlie the theories, and second, sufficient data must be available. Three areas that satisfy these constraints have received a large fraction of the attention of applied contract theorists: executive compensation, sales-force and franchise contracting, and industrial procurement.

Executive-compensation packages provide a rich laboratory in which to test the insurance/incentive aspects of contract theory.[1] Incentive pay is a nontrivial fraction of top-management compensation, where it takes the form of, for example, performance-based bonuses, stock ownership, and options to purchase shares in the firm. Furthermore, the details of executive-compensation packages are often publicly available.

Incentive pay is less prevalent, however, for low-level managers and production workers inside the firm. Nevertheless, it surfaces at this level of the hierarchy in at least one area where it takes a somewhat different form.[2] Franchise contracting is an increasingly popular method of organization for retail markets. Rather than employ an agent to sell a product and give that agent high-powered incentives within the firm, companies often choose a less integrated form of organization that allows them to share their risks and profits with their local managers or agents in a flexible way. In particular, principals can control the incentive/insurance tradeoff and minimize

---

[1] See, for example, Murphy [1984], Jensen and Murphy [1990], Kaplan [1994], and Garen [1996].

[2] For other areas, see e.g. Lazear [1996] on the effect of piece rates on production-worker productivity. For a broader discussion of the effect of human-resource-management practices on production-worker productivity, see e.g. Ichniowski, Shaw and Prennushi, [1997].

transaction costs by proper choice of sales-force compensation and franchise contract terms. The principal's problem is thus whether to use internal or external salespeople and, in the latter case, how to structure the external contract.

Finally, the theory of relationship-specific investment and the associated need to assign property rights has been most extensively tested in the area of input procurement.[3] When firms require specialized inputs that have higher value inside the buyer/seller relationship than in a more general market, they must decide if they will produce those inputs themselves or purchase them from an independent supplier. In the latter case, they must also decide whether to interact in a spot market or enter into a long-term contract. Moreover, the tradeoff between productive efficiency and the severity of the holdup problem can be dealt with through the choice of the terms of the procurement contract, specifically its length and flexibility.

In this paper, we look at the second of the above areas of empirical research, franchise contracting and sales-force compensation, and we examine different aspects of the incentive/insurance tradeoff in that context.[4] We do this in two ways. First, we construct the simplest theoretical model that is capable of capturing the effect of our focus, and second, we examine the empirical evidence from published studies that have assessed this aspect of the problem.

The models that we construct are based on the standard principal/agent paradigm. We make no effort to be theoretically sophisticated. Instead, we choose convenient functional forms that lead to definite solutions to the contracting problem. Furthermore, we construct models that involve only a few parameters, and we examine the models' comparative statics with respect to those parameters. Finally, we use the comparative statics from the theoretical exercise to organize our discussion of the empirical evidence.

The object of our exercise is to determine how well the simple theories perform in predicting the empirical regularities. It turns out that the empirical evidence is very consistent. In other words, coefficients from different studies that focus on a particular aspect of the contracting problem are usually of the same sign. This means that there is a set of stylized facts that should be explained. Unfortunately, the agreement between theoretical predictions and empirical regularities is less satisfactory than the robustness of the empirical findings. For this reason, when we discover that theory and evidence do not agree, we attempt to modify the simple model by introducing neglected aspects of the problem that move the theory in the direction of the data.

---

[3] For example, see Monteverde and Teece [1982], Anderson and Schmittlein [1984], Masten [1984], Anderson [1985], Masten and Crocker [1985], Joskow [1988], Klein [1988], and Crocker and Reynolds [1992]. For surveys of this empirical literature, see Shelanski and Klein [1995], Crocker and Masten [1996], and Rindfleisch and Heide [1997].

[4] For surveys of the franchising literature with a different emphasis, see Dnes [1996] and Elango and Fried [1997]. For surveys with a broader contracting focus, see Lyons [1996] and Masten [1998].

The organization of the paper is as follows. In the next section, we develop some background material on the environment in which franchising operates and the constraints that franchising data impose on the analysis.

In section III, which is the heart of the paper, we decompose the contract-choice problem into components that are amenable to econometric investigation. We make use of a standard agency model to organize our discussion of nine aspects of the contracting problem and how each affects the choice of organizational form. These aspects are local-market risk, the importance of the agent's effort, the size of the outlet, the difficulty of monitoring the agent, the importance of the principal's effort, the nature of product substitutability, spillovers among units of the chain, strategic delegation of the pricing decision, and the division of the agent's effort among tasks. We model each of these factors with a different specification of the effort/sales relationship in an otherwise standard model, and then examine the relevant evidence. We conclude this section with a short overview of studies that assess the effects of these same factors but have focused on contract terms rather than contract choice.

In section IV, we turn to some loose ends that need tying. In particular, we touch upon the consequences of contract choice for the level of product prices, its effect on firm performance, the lack of contract fine tuning in most real-world markets, why royalties are based on sales rather than profits, and the relevance of asset specificity for retail-contract choice.

Finally, section V summarizes and concludes.

II: Background

Manufacturers of retail products must decide whether to sell their products to consumers themselves (vertical integration) or to sell via independent retailers (vertical separation). When manufacturers do not perform the sales function internally, but want exclusive retailers, they either choose some form of franchising or employ an independent sales force.

Within the realm of franchising, there are two commonly used modes. Traditional franchising, which involves an upstream producer and a downstream seller (e.g., gasoline), accounts for the larger fraction of sales revenues. Business-format franchising, however, is the faster growing of the two. With business-format franchising, the franchisor provides a trademark, a marketing strategy, and quality control to the franchisee in exchange for royalty payments and up-front fees. Production, however, usually takes place at the retail outlet (e.g., fast-food).[5]

[5] The distinction between these two types of franchising can be blurred sometimes because business-format franchisors can sell inputs to franchisees (e.g. Baskin-Robbins), and traditional franchisors offer training and ongoing business support to their dealers as well. See Dnes [1992, 1993] for more on this.

Not all selling agents that are separated from the parent firm are franchisees. Some industrial companies choose between an internal sales force, which is known as "direct" sales, and an external sales representative. A manufacturer's external sales representative is an independent business entity that offers selling services and receives commissions on realized sales. This agency often serves a number of noncompeting manufacturers whose products form a package or product line. Moreover, the agency is normally each principals' exclusive representative for a designated set of customers.

Both the use of franchising and independent sales forces normally involve profit and risk sharing. As a consequence, much of the agency-theoretic literature in the retail-contracting area focuses on explaining the size of the share parameter in a franchise or sales contract, where the share parameter determines the partition of residual-claimancy rights between principal and agent. In particular, the literature shows how this parameter should vary as a function of the specific characteristics of the agent, the principal, the outlet, and the market.[6]

In real-world markets, in contrast, instead of offering contracts tailored to the characteristics of each unit, location, and agent, most firms employ a limited set of contracts, often just two -- a separated and an integrated contract. In doing this, they reduce the problem of choosing the contract terms for any particular unit from a continuum of options to a simpler dichotomous choice.[7] Consequently, much of the empirical literature has analyzed the dichotomous choice between company operation or in-house sales force (vertical integration, which is associated with lower-powered incentives) and franchising or sales representatives (vertical separation, which is associated with higher-powered incentives) using arguments that were developed to explain how firms should choose the terms of their contracts. In what follows, we focus mostly on the findings from the literature that examines this dichotomy. However, we discuss the more limited literature on the determinants of the terms of franchise contracts at the end of Section 3. We also return later, in Section 4, to the reasons why firms employ a set of standard

---

[6] See for example Rubin [1978], Mathewson and Winter [1985], Lal [1990], and Bhattacharyya and Lafontaine [1995]. Also see Stiglitz [1974] for the earliest application of agency theory to explain the use and properties of another type of share contract, namely sharecropping.

[7] In business-format franchising, different franchisors choose different contract terms -- different royalty rates and franchise fees -- but a given franchisor offers the same terms to all potential franchisees at a given point in time. This makes the franchise versus company-operation dichotomy a meaningful one; if contracts were allowed to vary for each franchisee, then, assuming for simplicity that the company manager is paid a fixed salary, company ownership would be a limit case where the royalty rate is zero and the franchise fee negative. Of course, such a limit case would hardly ever be observed. In reality, the dichotomy involves more than just differences in the compensation scheme of the unit manager; it also involves differences in asset ownership and in the distribution of responsibilities between upstream and downstream parties. Similarly, in traditional franchising, while commission rates and fees can vary across a firm's agents, the distinction between integration and separation is well defined. This distinction again involves differences in the distribution of power between manufacturer and retailer (see, for example, Smith II [1982] and Slade [1998a]).

contracts, and discuss in some detail how the dichotomous choice between franchising and integration then relates to the issue of high and low powered incentives within contracts.

Our analysis of the empirical evidence concerning retail contracting makes use of two sorts of studies. Data for the first sort are at the level of the upstream firm (or sector) and describe the extent to which managers choose to integrate vertically (i.e., their proportion of company-owned units). These data are typically cross sections of either a large number of firms from a broad range of industries or from a number of narrowly defined retail sectors.[8] Data for the second type are either at the level of the downstream unit or the sales force in a district and refer to whether this unit is integrated with the upstream firm. These data are typically cross sections from a few upstream firms in a single industry.[9] In other words, with the first type of study, an observation is an upstream firm, whereas with the second, it is a contract. The two sets of studies also differ in that the first involves mostly business-format franchising, whereas the second includes many industries in which the principal is a manufacturer.

Tables 1 to 6 summarize the findings of studies that assess the choice between integration and separation. In all these tables, the signs in the final columns show the observed effect of a variable of interest on the tendency towards vertical separation. A minus sign thus indicates a negative correlation with the extent of franchising in a chain or with the use of "separated" sales representatives in the sales-force-integration problem. Moreover, in all tables, an asterisk next to a plus or minus sign indicates that the finding is statistically significant at the 0.05 level based on a two-tailed test.

In what follows, each table is discussed in the subsection that presents the corresponding theory. One should be aware that the authors of the empirical studies do not always interpret their results in the way that we do. However, since we try to organize the empirical evidence using the framework of our model, we make no attempt to reconcile their interpretations and ours.

III: Factors that Influence Contract Choice

IIIa: The Basic Model

We have identified nine factors that frequently surface in empirical investigations of the determinants of retail contracting. These factors are not necessarily the most important, since our list is constrained by considerations of measurability and data availability. To illustrate, the agent's degree of risk aversion plays an important role in the theoretical incentive-contracting literature.

---

[8] For example, Brickley and Dark [1987], John and Weitz [1988], Martin [1988], Norton [1988], Lafontaine [1992a], and Scott [1995].

[9] For example, Anderson and Schmittlein [1984], Barron and Umbeck [1984], Anderson [1985], Brickley and Dark [1987], Minkler [1990], Muris, Scheffman and Spiller [1992], Shepard [1993], Graddy [1995], Lafontaine [1995], and Slade [1996 and 1998a].

Unfortunately, from an empirical point of view, it is virtually impossible to measure this factor directly. For this reason, we do not include it on our list.

In performing our analysis of the factors, we use the following standard principal/agent model. An agent exerts an effort, *a*, that results in an outcome, *q*, according to the relationship

$$q = f(a,\varepsilon,\Theta), \qquad \varepsilon \sim N(0,\sigma^2). \tag{1}$$

In equation (1), ε is a random variable that determines risk, and Θ is a vector of parameters. We identify the outcome, *q*, with sales, which is indistinguishable from sales revenue since we normalize product price to one (with some exceptions, clearly noted). The functional form of $f(.)$ will vary, depending on the aspect of the incentive-contracting problem that we examine. Indeed, it is our principal method of distinguishing the various factors whose effects we analyze below. [10]

The agent bears a private cost of effort, $C(a) = a^2/2$, and receives utility from his income *y*, $u(y) = -\exp(-ry)$, where *r* is his coefficient of absolute risk aversion. It is well known that in this setup, the agent behaves as if he were maximizing his certainty-equivalent income, *CE*, which is $E(y) - r/2\ Var(y)$, where *E* is the expectation operator, and *Var* is the variance function.

The risk-neutral principal offers the agent a linear contract, $s(q) = \alpha q + W$, where α is a commission rate, and *W* is a fixed wage.[11] In other words, αq is the agent's incentive pay, whereas *W* is his insured income. One can write the contract in an alternate but equivalent form that corresponds more closely to a business-format franchise contract. In particular, as we have not restricted the signs of α and *W*, it is possible to express the agent's payment as $s(q) = (1 - \rho)q - F$, where *F* is the franchise fee, and ρ is the royalty rate. As we want our model to describe both types of franchising as well as industrial selling, in what follows, we choose to use the former notation. The agent's income is then $y = \alpha q + W - a^2/2$.

The parameter, α, plays a key role in the analysis as it determines the agent's share of residual claims. Two limit cases are of interest. When $\alpha = 0$, the agent is a salaried employee who is perfectly insured, whereas when $\alpha = 1$, the agent is the residual claimant who bears all of the risk. One expects that, in general, $0 \leq \alpha \leq 1$. We identify α with the power of the agent's incentives. Moreover, we assume that inside the firm these incentives are low, whereas the contracts that are written with non-employees are higher powered. In theory, this need not be the case.[12] In practice, however, it is a strong empirical regularity.[13]

---

[10] Note that, as we assume below that the error term enters all of our functional forms in some additive way, our assumption that $\varepsilon \sim N(0,\sigma^2)$ also implies that *q* is normally distributed.

[11] We use the word linear here as has traditionally been done in the share-contract literature. The contracts, however, typically include a fixed component and are thus affine.

[12] See e.g. Lutz [1995] for a discussion of this issue in the context of franchising.

[13] For a possible explanation, see Holmstrom and Milgrom [1994]

We also restrict attention to linear contracts. Clearly, linearity is associated with mathematical tractability, which is desirable from our point of view. Unfortunately, however, optimal contracts are rarely linear. Nevertheless, linearity is the rule, not the exception, when one examines the contracts that are written in real-world situations.[14]

We do not attempt to explain these two observed phenomena -- low-powered incentives inside firms and linear contracts. Instead, we take them as empirical regularities that can be used to simplify the model. Furthermore, as a way to focus the paper more specifically on the theories and factors of interest, we relegate most of the mathematical derivations to the appendix.

We now turn to the factors of interest, the first of which is risk.

IIIb: Risk

One can use the simplest possible form of the effort/sales relationship to capture the effect that risk has on the form of the agent's contract. Specifically, let

$$q = a + \varepsilon. \tag{2}$$

The random variable, ε, is a proxy for either demand or supply uncertainty. In other words, one can interpret (2) as a demand equation (with price suppressed) where the role of effort is to increase sales. On the other hand, one can view (2) as an effort/output production function.[15]

With this form of the effort/sales function, the agent's certainty-equivalent income is given by

$$CE = \alpha a + W - \frac{a^2}{2} - \frac{r}{2} \alpha^2\sigma^2, \tag{3}$$

where the last term, $- r/2\ Var(y)$, is the agent's risk premium. Given a contract (α,W), the agent will choose effort to maximize equation (3), which leads to the first-order condition

$$a = \alpha. \tag{4}$$

The principal is assumed to maximize the total surplus, which she can extract from the agent with the fixed payment, *W*. Alternatively, *W* can be used to divide the surplus between

14 For possible explanations, see Holmstrom and Milgrom [1987], Romano [1994], and Bhattacharyya and Lafontaine [1995].

15 In franchising applications, see Lal [1990] for an example of the first type of interpretation, and Bhattacharyya and Lafontaine [1995] for an example of the second.

principal and agent when some rent is left downstream.[16] We do not model the choice of *W*, which we leave intentionally vague. The principal's problem is then to

$$\max_{\alpha,a} \; a - \frac{a^2}{2} - \frac{r}{2}\alpha^2\sigma^2, \qquad (5)$$

subject to the agent's incentive constraint (4) and a participation constraint that we also do not model.[17]

When equation (4) is substituted into the first-order condition for the maximization of (5) with respect to α, one finds that, in the optimal contract,

$$\alpha^* = \frac{1}{1 + r\sigma^2} \quad . \qquad (6)$$

Equation (6) implies that when either risk or the agent's degree of risk aversion increases, α falls.

The standard agency model of retail contracting therefore suggests that, as the level of uncertainty increases, so does the cost of agent insurance and thus the desirability of vertical integration. In other words, the firm will choose to integrate its retail activities more when facing more uncertainty because the higher-powered incentives used outside the firm expose the agent to the vagaries of the market, and the risk premium that the firm must pay consequently rises.

The notion of uncertainty or risk that is relevant in this context is the risk that is borne by the agent, not by the manufacturer. In other words, it is risk at the outlet or downstream level. Unfortunately, data that measure outlet risk are virtually nonexistent. For this reason, imperfect proxies are employed. The two most common are some measure of variation in detrended sales per outlet, and some measure of the fraction of outlets that were discontinued in a particular period of time.[18] Furthermore, data are more often available at the level of the sector rather than at the level of the franchisor or upstream firm. While this is an advantage from the point of view of resolving endogeneity issues, it can be a disadvantage if firm and sector risk are likely to be very different.

Table 1 gives details of five studies that assess the role of risk in determining the tendency towards franchising (i.e., vertical separation). In all but one of these studies, contrary to prediction, increased risk leads to more franchising (increased separation). Moreover, this positive

---

[16] See Kaufmann and Lafontaine [1994] for evidence that there are rents left downstream at McDonald's. The authors argue that they serve an incentive role similar to that of efficiency wages. Michael and Moore [1995] find evidence that such rents are present in other franchised systems as well.

[17] The participation constraint is normally used to determine *W*, not α.

[18] On the relative merits of these two measures, see Lafontaine and Bhattacharyya [1995].

association does not depend on the measure of risk that is used. These results suggest a robust pattern that is unsupportive of the standard agency model.[19]

The finding that risk is positively associated with vertical separation in the data is indeed a puzzle. Moreover, allowing effort to interact with risk in the model only makes matters worse: with such specifications, increased incentives can cause effort to fall, making high-powered incentives particularly costly to the principal, and thus especially undesirable.

Table 1: The Effect of Risk on the Propensity to Contract Out

Author	Year	Data	Measure	% Contracted
Anderson & Schmittlein	1984	Electronics Components by Product Line and Territory	% Forecast Error of Product-Line Sales by Territory	+
John & Weitz	1988	Industrial Firms with Sales above $50 million	Index capturing environmental uncertainty	-
Martin	1988	Sectoral Panel	Coefficient of Variation of Detrended Sectoral Sales	+*
Norton	1988	Restaurants and Motels by State & Sector	Variance of Detrended % Change in Sectoral Sales by State	+*
Lafontaine	1992	Bus. Format Franchising Firms from All Sectors	Fraction of Outlets Discontinued in Sector	+*

Note: * indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.

Some authors have concluded from this that franchisors shed risk onto franchisees (e.g. Martin [1988]). This could be optimal if franchisors were more risk averse than franchisees. However, if franchisors were indeed more risk averse, there would be less need to balance franchisee incentive and insurance needs, and hence less need to use a share contract to start with. At the extreme, franchising would involve franchisees paying only lump-sum fees to franchisors, a situation that is rarely observed in practice.

An alternative, and we believe more satisfactory, explanation for the observed risk/franchising phenomenon surfaces when one considers that market uncertainty can be endogenous and that the power of incentives can influence sales variability. Indeed, franchisees often have superior information concerning local-market conditions (separate from ε). Moreover,

19 See also Allen and Lueck [1992, 1995] and Leffler and Rucker [1991] for evidence that risk-sharing does not explain contract terms well in sharecropping and in timber harvesting respectively.

since franchising gives retailers greater incentives to react to these conditions, one is likely to find more sales variability in franchised than in company-owned units. In that sense, the positive relationship between risk and franchising can be understood as support for incentive-based arguments for franchising.[20]

IIIc: Agent Effort

Not all agents are equally important in determining the success or failure of a retail outlet. For example, consider the case of gasoline retailing. Some station operators are merely cashiers who sit in kiosks and collect payment from customers. Others, in contrast, offer a range of services that can include pumping gas, washing windows, checking oil, selling tires, batteries, and other automobile-related items, and repairing cars. Still others manage affiliated convenience stores.

To capture the notion that there are varying degrees of agent importance, we amend the effort/sales function as follows,

$$q = \eta a + \varepsilon, \tag{7}$$

while keeping the rest of the model intact. In equation (7), the parameter η, which is positive by assumption, is a proxy for the importance of the agent's effort.

After performing the same set of calculations as in the previous subsection, one finds that, with the new effort/sales function,

$$\alpha^* = \frac{\eta^2}{\eta^2 + r\sigma^2} \quad . \tag{8}$$

Moreover, differentiation of (8) with respect to η shows that $d\alpha^*/d\eta > 0$.

The theory thus predicts that increases in the importance of the retailer's input should be associated with more separation and higher-powered contracts. In other words, when the agent's job is more entrepreneurial in nature, his payment should reflect this fact.

From a practical point of view, the measures that have been used to capture this effect have been determined both by data availability and by the industry being studied. Proxies for the importance of the agent's effort (or its inverse) have included measures of labor intensity (either employees/sales or capital/labor ratios) as the agent is the one who must oversee the provision of

---

[20] See e.g. Lafontaine and Bhattacharyya, [1995]. Note that the positive relationship between incentives and output variability that they find depends on the form of the function that maps effort and the random variables into output.

labor. Researchers have also used a measure of the agent's value added, or discretion over input choices, and a variable that captures whether previous experience in the business is required. Finally, two studies of gasoline retailing rely on a dummy variable that distinguishes full from self service.

Table 2 summarizes the results from seven studies that assess agent importance. In every case where the coefficient of the agent-importance variable is statistically significant, its relationship with separation from the parent company is positive, as predicted by standard agency considerations and other incentive-based arguments. In other words, when the agent's effort plays a more significant role in determining sales, franchising is more likely.

Table 2: The Effect of the Importance of the Agent's Effort on the Propensity to Contract Out

Author	Year	Data	Measure	% Contracted
Caves and Murphy	1976	Sectoral Data	"Personalized Service" Dummy	+*
Norton	1988	Restaurants and Motels by State & Sector	Employees/Sales	+*
Lafontaine	1992	Bus. Format Franchising Firms from All Sectors	1- (Sales - Franchisor Inputs) / Sales for Sector 2- Previous Experience Required	+ -
Shepard	1993	Gasoline Service Stations in Massachusetts	Full Service	+
Scott	1995	Bus. Format Franchising Firms from All Sectors	Capital/labor ratio	(-*)
Slade	1996	Gasoline Service Stations in Vancouver	Full Service	+*
Bercovitz	1998b	Individual outlets from 20 Fast-Food and Retailing Chains	Discretionary Inputs = (Costs - $ Value of Franchisor Inputs) / Costs	+*

Notes: Parentheses in the last column indicate that the relevant variable is an inverse measure of agent effort and is therefore expected to have a sign opposite to the others.
* indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.

IIId: Outlet Size

Modeling the effect of outlet size is less straightforward than for the previous two factors, and model predictions are more sensitive to specification as a consequence. We confess that the particular specification that we adopt was chosen so that results are consistent with the empirical regularity that we present below. Indeed, it is necessary that we model size as interacting with risk in order to obtain our prediction.[21] With this caveat, we specify the effort/sales relationship as a production function whose arguments are franchisee effort, *a*, and outlet size or capital, *k*,

$$q = \eta\, a + (\gamma + \varepsilon)\, k. \tag{9}$$

All other assumptions are as before.

There are two things to note about equation (9). The parameter γ measures the direct effect of capital in the production function, whereas *k* is a proxy for the amount of capital invested. Furthermore, our specification assumes that a larger outlet is associated with increased agent risk. This does not mean that the market is riskier *per se*; it simply means that more capital is subject to the same degree of risk.

After the standard set of manipulations, we obtain

$$\alpha^* = \frac{\eta^2}{\eta^2 + r\sigma^2 k^2} \quad . \tag{10}$$

Note that γ does not appear in this solution: thus outlet size, if it enters the production function in an additive way, has no effect on optimal contract terms. However, when interacted with risk, *k* does matter. In other words, the amount of capital invested in the outlet rather than its importance in determining sales directly is what matters here.

Furthermore, differentiating α^* with respect to *k* yields a negative relationship, which implies that the agent should be given lower-powered incentives when the size of the capital outlay increases. This presumes that it is the agent's capital, not the principal's, that is at risk. In other words, the larger the outlet, the more capital the franchisee has at stake and the more insurance he requires. Thus the solution implies a lower share for the agent, or more vertical integration.

[21] As shown below, if one assumes that *k* enters (9) only in an additive way, then changes in *k* have no effect on the optimal share parameter, α. If one assumes that *k* multiplies *a*, then its effect is the same as that of η in the previous subsection, and increases in *k* lead to higher values of α, the reverse of what we obtain with our formulation. With a combination of interactive terms with risk and franchisee effort, we would get two opposing effects, and the sign of the net effect would depend on the specific parameters of the problem.

Furthermore, vertical integration in this context has the added advantage that it substitutes the principal's capital for the agent's.[22]

Unlike the factors discussed above, the empirical measurement of size is fairly straightforward. Common measures are average sales per outlet and the initial investment required. Table 3 shows that, with one exception, greater size leads to less separation or increased company ownership. In other words, as the model above predicts, people responsible for large outlets tend to be company employees who receive low-powered incentives.[23]

Table 3: The Effect of Outlet Size on the Propensity to Contract Out

Author	Year	Data	Measure	% Contracted
Brickley & Dark	1987	Selected Franchising Firms	Initial Investment	-*
Norton	1988	State Level Sectoral Data for Restaurants and Motels	Sales/Outlet	+*
Martin	1988	Sectoral Panel	Sales/Outlet	-*
Brickley, Dark and Weisbach	1991	1-State Level Sectoral Data 2- Outlet Data from 36 Chains	Initial Investment Initial Investment	-* -*
Lafontaine	1992	Bus. Format Franchising Firms from All Sectors	1- Initial Investment 2- Sales/Outlet for Sector	-* -*
Thompson	1994	Bus. Format Franchising Firms from All Sectors	Initial Investment	-*
Lafontaine	1995	Individual Fast-Food Restaurants in the Pittsburgh and Detroit Metropolitan Areas	Number of Seats in an Outlet	-*
Scott	1995	Bus. Format Franchising Firms from All Sectors	Initial Investment	-
Kehoe	1996	Individual Hotels from 11 Major Chains	Number of Rooms	-*

Note: * indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.

---

22 See Brickley and Dark [1987] for more on this argument, which they refer to as the "inefficient risk-bearing" argument against franchising.

23 Consistent with the above evidence, on a sectoral basis, company units have higher sales (are larger) than franchised units (US Dept. of Commerce, 1988). Moreover, Muris, Scheffman and Spiller [1992] argue that the increase in the efficient size of bottling operations led soft-drink manufacturers to buy back several of their independent bottlers and enter into joint-venture agreements with many others.

It is comforting to see that theory and evidence agree. Nevertheless, as noted above, it is possible to argue for the opposite relationship in an equally convincing manner. Indeed, when an outlet is large, the agent has more responsibility. For this reason, outlet size has been used in the empirical literature as a measure of the importance of the agent's input. Not surprisingly then, it is often claimed that an agency model should predict that an increase in size will be associated with more separation and higher-powered incentives (see footnote 21). Furthermore in a model with spillovers across units of the same chain, smaller outlets have a greater tendency to free ride since outlets with larger market shares can internalize more of the externality (Gal-Or [1995]). In this type of model, small units would be more likely to be vertically integrated.[24] The data, however, contradict this prediction.

IIIe: Costly Monitoring[25]

The idea that monitoring the agent's effort can be costly or difficult for the principal is central to the incentive-based-contracting literature. In fact, if monitoring were costless and effort contractible, there would be no need for incentive pay. The agent's effort level would be known to the principal with certainty, and a contract of the following form could be offered: If the agent worked at least as hard as the first-best effort level,[26] he would receive a salary that compensated him for his effort, whereas if his effort fell short of this level, he would receive nothing. In equilibrium, the agent would be fully insured, and the first-best outcome would be achieved.

Given the centrality of the notion of costly monitoring, it is somewhat surprising that there exists confusion in the literature concerning the effect of an increase in monitoring cost on the tendency towards company operation. For example, consider the following statements from the empirical literature:

> *The likelihood of integration should increase with the difficulty of monitoring performance.* (Anderson and Schmittlein [1984 p. 388]).
>
> *Franchised units* (as opposed to vertical integration) *will be observed where the cost of monitoring is high.* (Brickley and Dark [1987 p. 408], text in parentheses added).

These contradictory statements imply that monitoring difficulties should both encourage and discourage vertical integration.

---

24 This result also depends on the assumption that information flows are superior within the firm.

25 This subsection is based on Lafontaine and Slade [1996].

26 The first-best effort level is defined as the level that the principal would choose if she were not constrained by incentive considerations in maximizing the total surplus.

To reconcile these discrepancies, we modify the standard agency model to include the possibility that the principal can use not only outcome (i.e., sales) information to infer something about the agent's effort, but also a direct signal of effort.[27] Furthermore, the principal is allowed to base the agent's compensation on both signals.[28]

We consider two types of signals because, in most real-world manufacturer-retailer relationships, it is possible to supervise the actions of a retailer directly by, for example, testing food quality, assessing the cleanliness of the unit, and determining work hours. This direct supervision provides the manufacturer with information on retailer effort that supplements the information contained in sales data. In general, the informativeness principle (Holmstrom [1979] Milgrom and Roberts [1992, p. 219]) suggests that compensation should be based on both sales data and signals of effort obtained via direct monitoring.

To model this situation, we replace the effort/sales relationship (1) with two functions to denote the fact that the principal receives two noisy signals of the agent's effort.[29] First, the principal observes retail sales of the product, q, and second, the principal receives a direct signal of effort, e,

$$\begin{aligned} q &= a + \varepsilon_1, \\ & \qquad\qquad \varepsilon \sim N(0,\Sigma), \qquad\qquad (11) \\ e &= a + \varepsilon_2, \end{aligned}$$

where $\varepsilon = (\varepsilon_1,\varepsilon_2)^T$, $\Sigma = (\sigma_{ij})$, $\sigma_{ij} = \sigma_{ji}$, and $\sigma_{ii} > \sigma_{ij}$, $i = 1,2$, $j \neq i$.

The contract that the principal offers the agent is amended to include, in addition to the fixed wage W, not only an outcome-based or sales commission rate, α_1, but also a behavior-based commission rate, α_2, that relates to the direct signal of effort. The agent's certainty-equivalent income is then $(\alpha_1 + \alpha_2)\, a + W - a^2/2 - (r/2)\, \alpha^T\Sigma\alpha$, where α is the vector of commissions, $\alpha = (\alpha_1,\alpha_2)^T$. The agent’s incentive constraint for this problem is $a = \alpha_1 + \alpha_2$.

As before, the risk-neutral principal chooses the agent's effort and the commission vector to maximize the total surplus subject to the agent’s incentive constraint. When the two first-order conditions for this problem are solved they yield

---

27 One alternative source of information that we do not consider arises when uncertainty is correlated across agents in a multi-agent setting. In that case, the optimal contract for agent *i* includes some measure of other agents' performance in addition to his own, as in Holmstrom [1982]. Empirically, such relative-performance contracts are not used in franchising.

28 The type of mechanism that we have in mind is sometimes called “behavior-based” compensation, as opposed to “outcome-based” compensation. See Anderson and Oliver [1987].

29 The model is similar to Holmstrom and Milgrom [1991], who model multiple tasks and signals.

$$\alpha_1{*} = \frac{\sigma_{22} - \sigma_{12}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12} + r(\sigma_{11}\sigma_{22} - \sigma_{12}{}^2)}, \qquad (12a)$$

and

$$\alpha_2{*} = \frac{\sigma_{11} - \sigma_{12}}{\sigma_{11} + \sigma_{22} - 2\sigma_{12} + r(\sigma_{11}\sigma_{22} - \sigma_{12}{}^2)}. \qquad (12b)$$

When the noisy signals are uncorrelated, so that $\sigma_{ij} = 0$, equation (12) takes the simpler form

$$\alpha_1{*} = \frac{1}{(1 + r\sigma_{11} + \sigma_{11})/\sigma_{22}}, \qquad (13a)$$

and

$$\alpha_2{*} = \frac{1}{(1 + r\sigma_{22} + \sigma_{22})/\sigma_{11}}, \qquad (13a)$$

which shows that the optimal contract described in equation (6) must now be modified to account for the relative precisions of the two signals. In other words, the compensation package places relatively more weight on the signal with the smaller variance. Equation (6) is a special case of (13) in which σ_{22} is infinite (direct monitoring contains no information).

We are interested in the effect of increases in the two sorts of uncertainty on the size of α_1^* since this is the incentive-based pay that appears in the data. Differentiating equation (12) with respect to the two variances shows that $d\alpha_1^* / d\sigma_{11} < 0$ and $d\alpha_1^* / d\sigma_{22} > 0$. Increases in the precision of sales data ($1/\sigma_{11}$) thus lead to a higher reliance on outcome-based compensation (higher α_1), which corresponds to less vertical integration. However, increases in the precision of the direct signal of effort ($1/\sigma_{22}$) lead to less outcome-based compensation (lower α_1) or more vertical integration.

While the above model does not explicitly include monitoring costs, it should be clear that if the upstream firm can choose some action that reduces σ_{11} (increases the precision of sales as a signal of effort) at some cost, it will do so to a greater extent the lower this cost is. The resulting decrease in σ_{11} will in turn lead to a greater reliance on sales data in the compensation scheme. In other words, when the cost of increasing the precision of sales data as an indicator of effort is low, we should observe more reliance on sales data in the compensation scheme, which means less vertical integration. On the other hand, when the cost of behavior monitoring, or of reducing σ_{22}, is low, the firm will perform more of this type of monitoring. A low σ_{22} will then lead the firm to choose a lower α_1, which amounts to more vertical integration.[30]

To summarize, our comparative statics show that the effect of monitoring on the degree of vertical integration depends on the type of information garnered by the firm in the process. If this information gives a better direct signal of effort, it reduces the need to use sales-based incentive contracting. If, on the other hand, monitoring increases the value of sales data by increasing its precision, it makes incentive contracting more attractive.

Turning to the empirical evidence, we separate the studies in two parts in table 4 based on their interpretation of monitoring costs. The first part of the table shows results obtained in the sales-force compensation literature, where the focus has been on the usefulness of observed sales data as an indicator of agent effort. The second part of table 4 contains empirical results from the franchising literature, where authors have focused on the cost of behavior monitoring.

Table 4, Part I:
The Effect of Monitoring Difficulty on the Propensity to Contract Out

Author	Year	Data	Measure	% Contracted
Anderson & Schmittlein	1984	Electronics Components by Product Line and Territory	Index indicating that it is difficult to measure results of individuals	-*
Anderson	1985	Electronics Components by Product Line and Territory	Index indicating that 1) team sales are common, 2) records are inaccurate and 3) sales and cost figures are insufficient for a fair evaluation	-*
			Importance of non-selling activities	-*
John & Weitz	1988	Industrial Firms with Sales above $50 million	Length of Selling Cycle	-*

Note: * indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.

[30] In mapping our results from more or less sales-based compensation to more or less vertical integration, we are implicitly assuming that behavior monitoring takes place, and behavior-based compensation is used, inside the firm, but that sales commissions are not or are little used inside the firm. With complete separation, in contrast, the agent is the residual claimant, and there is no (or very little) behavior monitoring or behavior-based compensation. See Holmstrom and Milgrom [1991] for a discussion of these issues. See Bradach [1997] for descriptions of business practices in five franchised restaurant chains that suggest that these assumptions are realistic.

Table 4, Part II:
The Effect of Monitoring Difficulty on the Propensity to Contract Out

Author	Year	Data	Measure	% Contracted
Brickley & Dark	1987	Selected Franchising Firms	Distance From Monitoring Headquarters	+*
Norton	1988	Restaurants and Motels by State & Sector	Fraction of State Population Rural	+*
Minkler	1990	Taco Bell Restaurants in Northern California and Western Nevada	1- Distance From Monitoring Headquarters 2- Outlet Density = Number of Outlets within a 5 Mile Radius	+* (+)
Brickley, Dark and Weisbach	1991	1- State Level Sectoral Data 2- Outlet Data from 36 Chains	Density: Units per Square Mile Density: Company's Units in County	(-*) (-*)
Carney and Gedajlovic	1991	Canadian Bus. Format Franchising Firms from all Sectors	Density: Proportion of Outlets in Quebec	(-*)
Lafontaine	1992	Bus. Format Franchising Firms from All Sectors	Number of States in which the Chain has Established Outlets	+*
Lafontaine	1995	Fast Food in Pittsburgh and Detroit Metropolitan Areas	Outlet Density = Number of Outlets from the Same Chain in same Zip Code	(-*)
Scott	1995	Bus. Format Franchising Firms from All Sectors	Number of States in which the chain has established outlets	+*
Kehoe	1996	Individual Hotels from 11 Major Chains	Density: Number of Hotels from the Same Chain in Same City	(-*)
Bercovitz	1998b	Individual Outlets from 20 Fast-Food and Retailing Chains	1- Miles to Monitoring HQ 2- Density: Inverse of the Average Distance of the Four Closest Units from the Same Chain	+* (-)

Notes: Parentheses in the last column indicate that the relevant variable is an inverse measure of monitoring cost and is therefore expected to have a sign opposite to the others.
* indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.

In the first part of the table, in the first two studies, researchers asked managers to respond to various statements: In Anderson and Schmittlein [1984], they responded to "it is very difficult to measure equitably the results of individual salespeople" while in Anderson [1985], the measure was tabulated from responses to "(1) team sales are common, (2) sales and cost records tend to be inaccurate at the individual level, and (3) mere sales volumes and cost figures are not enough to make a fair evaluation." In John and Weitz [1988], the length of the selling cycle was used on the basis that a long lag between actions and market responses makes it difficult to attribute output to effort. In addition, these authors included a measure of environmental uncertainty, which captures the extent to which agents "control" sales outcomes. Using scores thus obtained as measures of the cost of monitoring sales and inferring effort from it, researchers found that higher monitoring costs lead to more vertical integration, as predicted by our model.

The second part of table 4 includes studies in which authors have used a variety of measures of behavior-monitoring costs, including some notion of geographical dispersion (captured in one case by whether the unit is more likely to be in a mostly urban or rural area) or of distance from monitoring headquarters. These measures are proxies for the cost of sending a company representative to visit the unit to obtain data on cleanliness, product quality, etc. Outlet density has also been used, but as an inverse measure of behavior-monitoring cost. One can see that when behavior-monitoring costs are measured either directly by dispersion or distance, or inversely by density, in all cases where coefficients are significant, higher monitoring costs lead to more vertical separation. This reflects the fact that when behavior monitoring is costly, firms rely on it less, and rely more on residual claims to compensate their agents. Again the evidence is consistent with the model.

It should be clear then that the two types of measures used in the empirical literature have captured different types of monitoring costs: the fit of sales data to individual effort versus direct monitoring that is a substitute for sales data. Taking this difference into account, the seemingly contradictory results obtained and claims made by these researchers are in fact consistent with each other as well as with standard downstream-incentives arguments for retail contracting.

IIIf: Franchisor Effort

The standard agency model assumes, as we have, that only one party, the agent, provides effort in the production (or sales-generation) process. In reality, success at the retail level often depends importantly on the behavior of the upstream firm or principal. For example, franchisees expect their franchisors to exert effort towards maintaining the value of the trade name under which they operate, via advertising and promotions, as well as screening and policing other franchisees in the chain. If this behavior is not easily assessed by the franchisee, there is moral hazard on both sides -- the franchisee's and the franchisor's -- and the franchisor, like the franchisee, must be given incentives to perform.[31]

To capture the effect of franchisor effort on the optimal contract, we amend the effort/sales relationship to include not only franchisee effort, *a*, but also franchisor effort, *b*,

$$q = \eta a + \theta b + \varepsilon, \qquad (14)$$

where the parameter $\theta > 0$ is a proxy for the importance of the franchisor's effort. Assume also that the private cost of effort for the franchisor is $C(b) = b^2/2$, the same as for the franchisee. The franchisor still chooses the share parameter, α, in the first stage, but now the contract must satisfy incentive compatibility for both parties. As before, the first-order condition for the franchisee's effort gives $a = \alpha \eta$. The first-order condition for the franchisor's choice of effort is $b = (1-\alpha)\theta$. Substituting these into the total-surplus function, one obtains the optimal-share parameter

$$\alpha^* = \frac{\eta^2}{\eta^2 + \theta^2 + r\sigma^2} \quad . \qquad (15)$$

Differentiating α^* with respect to η again shows that the optimal share, or the extent of vertical separation, goes up as the franchisee's input becomes more important. However, differentiating α^* with respect to θ yields the opposite effect; not surprisingly, when the input of the franchisor increases in importance, it is the share of output that she receives, $(1-\alpha^*)$, or the extent of vertical integration, that must rise.

Table 5 shows results obtained in five studies that have considered how the importance of the franchisor's input affects the optimal contract choice. The importance of these inputs is measured by the value of the trade name (proxied by the number of outlets in the chain or the difference between the market and the book value of equity), the amount of training or advertising provided by the franchisor, or the number of years spent developing the business format prior to franchising. The table shows that, in all cases, when franchisor inputs are more important, less vertical separation is observed, as predicted.

One proxy for the importance of the franchisor's input that has been used in the literature but is not included in table 5 is the chain's number of years of franchising (or business experience). The idea is that more years in franchising (or business) lead to a better known, and

[31] See e.g. Rubin [1978], Mathewson and Winter [1985], Lal [1990] and Bhattacharyya and Lafontaine [1995] for more on this. Consistent with the argument that the franchisor must be given incentives in these cases, in the one case of a franchise agreement that does not involve any ongoing royalties or company ownership on the part of the franchisor, Dnes [1993] notes that "Franchisees (in this system) do complain of insufficient effort by the franchisor in supporting the development of their businesses." (p. 386; text in parentheses added)

thus more valuable, trade name. However, this variable is also a proxy for the extent to which franchisors have access to capital as well as for learning and reputation effects. Furthermore, the empirical results that pertain to this variable are mixed. Using panel data at the franchisor level, Lafontaine and Shaw [1998b] find that, after the first few years in franchising, the proportion of corporate units within chains levels off and remains quite stable. They conclude that a firm's years in franchising is not a major determinant of the extent of vertical integration in franchised chains.[32]

Table 5: The Effect of the Importance of the Franchisor's Effort on the Propensity to Contract Out

Author	Year	Data	Measure	% Contracted
Lafontaine	1992	Bus. Format Franchising Firms from All Sectors	1- Weeks of Training 2- Lagged No. of Outlets 3- % Time Not Franchising	-* -* -*
Minkler and Park	1994	Panel of Publicly Traded Bus. Format Franchising Firms from All Sectors	Market Minus Book Value of Equity	-*
Thompson	1994	Bus. Format Franchising Firms from All Sectors	Number of Years in Business Prior to Franchising	-*
Scott	1995	Bus. Format Franchising Firms from All Sectors	Days of Training	-
Bercovitz	1998b	Individual Outlets from 20 Fast-Food and Retailing Chains	Franchisor Advertising	-*

Note: * indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.

IIIg: Spillovers Within the Chain

The standard incentive-cum-insurance model of retail contracting does not usually consider the competitive environment in which the principal/agent relationship operates. Instead, this relationship is modeled as if the market were perfectly competitive and price were exogenous to the firm. Alternatively, the franchisor is modeled as a monopolist, an assumption that also eliminates the importance of rivals. Most markets in which franchising is prevalent, however, are better characterized as monopolistically competitive. Usually, there are several firms that produce similar

[32] For a review of the empirical literature on the "ownership redirection hypothesis", according to which franchising is just a transitory phase for firms that face capital constraints, see Dant, Kaufmann and Paswan [1992]. For recent contributions, see also Lafontaine and Kaufmann [1994], Thompson [1994] and Scott [1995].

but not identical products, and firms as well as units within firms face downward-sloping demand. In this and the next two subsections, we consider the consequences of endogenous prices.

One reason for the prevalence of chains rather than independent sales outlets is that there are externalities that are associated with the brand or chain name. Although spillovers can be beneficial, they can also create problems for both franchisees and franchisors. For example, one form that a spillover can take is a demand externality. With this sort of spillover, a low price at one outlet in a chain increases demand, not only at that outlet but also for other franchisees in the same chain. Conversely, a high price can cause customers to switch their business to another chain rather than merely seek a different unit of the same chain.

In order to investigate the effect of demand spillovers, we amend the effort/sales relationship to include own price, p, and $\bar{p}$, the price charged by another outlet in the same chain,

$$q = 1 - p - \mu \bar{p} + a + \varepsilon. \tag{16}$$

Equation (16) is a standard linear demand equation, with a parameter, μ, that represents the extent of demand spillovers. Thus we assume that $\mu > 0$, which means that a high price at a given unit causes an erosion of the sales of all members of the chain. We also assume that the franchisor chooses downstream prices in addition to the share parameters in this version of the model.[33] All other model assumptions are as before.

None of the modifications of the model change the agent's incentive constraint, which gives $a = \alpha$. Using this to eliminate a, one finds that, in a symmetric equilibrium,

$$\alpha^* = \frac{1}{(1 + r\sigma^2)2(1 + \mu) - 1}, \tag{17}$$

and $d\alpha^*/d\mu < 0$. In other words, when there are demand externalities of the type one normally associates with branding, integration becomes more desirable. This is because the chain internalizes the spillover that is external to the individual unit.

There are other sorts of spillovers, such as franchisee free riding. Indeed, once an agent is given high-powered incentives via a franchise contract, the franchisee can shirk and free ride on the trade name (see e.g. Klein, [1980], Brickley and Dark, [1987], and Blair and Kaserman [1994]). The problem is that the cost of the agent's effort to maintain the quality of the trademark is private,

[33] For our current purposes, it is simpler to assume that the franchisor chooses price. There is some evidence that franchisors try to control franchisee prices (see e.g. Ozanne and Hunt [1971]), but rules against resale price maintenance have made this difficult up until recently, when the Supreme Court decided in *State* Oil *v. Khan* that maximum resale price maintenance would no longer be a *per se* violation of Antitrust Law. See Lafontaine (1998) for more on price controls in franchising. Note that the spillover problem is exacerbated when the franchisee chooses price. This situation can be modeled by changing the sign of δ in the demand equation in subsection IIIi.

whereas the benefits of his activities accrue, at least partially, to all members of the chain. In this case, the spillover works through effort, not price.

Whether the externalities work through price and/or effort, spillover problems are exacerbated in situations where consumers do not impose sufficient discipline on retailers, namely in cases of non-repeat businesses. The franchisor may therefore decide to operate directly those units in transient-customer locations, such as those around freeway exits, or to operate more outlets directly if involved in businesses subject to significant non-repeat business.

Table 6: The Effect of Non-Repeat Business on the Propensity to Contract Out

Author	Year	Data	Measure	% Contracted
Brickley & Dark	1987	1- Franchising Firms from All Sectors	Dummy Variable for Non Repeat Sectors	-*
		2- Outlets from 36 Franchising Firms in Various Sectors	Highway Dummy Variable	+*
Norton	1988	Restaurants and Motels by State & Sector	Tourism: Household Trips in the State	+* (in hotels)
Brickley, Dark and Weisbach	1991	1- State Level Sectoral Data	Non-Repeat Industry Dummy	-*
		2- Outlet Data from 36 Chains	Non-Repeat Industry Dummy	+ (at means)
Minkler	1994	Taco Bell Restaurants in Northern California and Western Nevada	Highway Dummy Variable	-
Lafontaine	1995	Fast Food in Pittsburgh and Detroit Metropolitan Areas	Highway Dummy Variable	+

Note: * indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.

Table 6 summarizes the evidence from those studies that have examined the effect of non-repeat business on the propensity to franchise. This table shows that the evidence on non-repeat is mixed. One explanation for this may be that franchisors find other ways to control franchisee free-riding, for example, by using approved-supplier requirements or self-enforcing contracts. If so, the role of the franchisor in maintaining service quality and trademark reputation should be particularly important in sectors where most business is transient, which in turn brings us back to the issue of franchisor incentives in a double-sided moral-hazard model of franchise contracting. In fact, measures of the "value of the trade name" have been used in the literature to test both the notion that franchisors must be given more incentives to perform when the trade name is very valuable (see table 5) and the notion that franchisee free-riding opportunities are greater under those circumstances. Furthermore, both sides of this coin lead to the same prediction -- that chains will rely more on vertical integration when the trade name is very valuable -- and are thus empirically indistinguishable. The results in table 5 are consistent with this prediction, whereas the results in table 6 overall do not support the non-repeat component of the free-riding model.

IIIh: Product Substitutability

In some franchising industries, products are easily distinguishable from one another. For example, most customers have definite preferences between McDonald's hamburgers and KFC's chicken. There are, however, other industries in which the services that the agents provide are perhaps the only things that distinguish the output of one firm from that of another. Real-estate franchises, for example, fall in the latter group. Given that, across industries, there are varying degrees of differentiation among products that are provided within the industry, one can ask how these differences affect contract choice.

The situation just described is the converse of the spillover case. Specifically, one can rewrite the demand equation as

$$q = 1 - p + \delta\bar{p} + a + \varepsilon. \tag{18}$$

There are two differences between equations (16) and (18). First, $\bar{p}$ in (18) is the price charged by an outlet from a rival chain, whereas it was the price charged in another unit of the same chain in (16). Second, δ here represents the degree of product substitutability between the two chains. We assume that δ is positive, but less than 1 so that the own-price effect is greater than the cross-price effect. The principal now chooses price, p, and the share parameter, α, given rival choices, $\bar{p}$ and $\bar{\alpha}$.[34] With these modifications, the corresponding equation for the optimal contract is

$$\alpha^* = \frac{1}{(1 + r\sigma^2)(2 - \delta) - 1} \quad , \tag{19}$$

and $d\alpha^*/d\delta > 0$. In other words, as products become closer substitutes, the power of the agents' incentives should be increased. This is true because it becomes more important to induce the agent to promote the product so that sales will not be eroded by customers switching to rival brands. Indeed, one can interpret the substitution effect as yet another measure of the importance of the agent's effort. The higher the degree of substitutability, the harder is the agent's task of preventing

[34] We continue to assume that the franchisor chooses price. In the next section, we relax this assumption.

the erosion of its sales. As in subsection IIIc, therefore, the principal has an additional motive for emphasizing high-powered incentives relative to other objectives.

Note that in modeling competition, we have implicitly assumed that the random variables that are associated with own and rival demand are uncorrelated. If, however, these variables are correlated, and if the agent has private information about his own demand realization, the tendency towards separation is strengthened when competition increases.[35] Indeed, demand correlation is information that the principal can use to reduce the agent's informational rent and thus the need to integrate (Gal-Or [1995]).

Given that most agency-theoretic models neglect the demand side of the market, it is not surprising that most empirical studies rely solely on attributes of the upstream firm and its outlets and ignore the firm's competitors. To our knowledge, Coughlan [1985] and Slade [1998a] are the only studies that have looked at contract choice as a function of the demand characteristics that agents face. Coughlan finds that firms are more likely to use a middleman (separation) to enter a foreign market if they sell highly substitutable products, and to sell directly (integration) if their product is more unique. Similarly, Slade relates outlet-level own and cross-price elasticities of demand to the contracts under which outlets operate. As the model predicts, she finds that higher cross-price elasticities are associated with higher-powered incentives for the agent.[36]

IIIi: Strategic Delegation of the Pricing Decision

We have assumed thus far that, when prices are endogenous, the principal chooses the retail price herself. In reality, however, with franchising, whether traditional or business-format, the principal usually delegates the pricing decision to the agent.[37] We now examine the principal's incentive to delegate in a strategic setting.

When price is exogenous, it is possible to normalize and make no distinction between rewarding the agent on the basis of revenues or sales. With endogenous prices, in contrast, particularly when the agent chooses price, it is important to be more specific. We therefore adopt an alternative notation that conforms more closely with actual compensation schemes in franchise chains. We maintain the demand assumption of the previous subsection (i.e., $q = 1 - p + \delta\overline{p} + a +$

35 Here the increase in the cross-price elasticity is due to an increase in the number of competitors.

36 When our evidence is from very few studies, we do not construct a table.

37 US Antitrust laws prevent franchisors from enforcing specific prices in franchised units as these are independent businesses under the law. Of course, this does not prevent franchisors from trying to affect franchisees' choice of prices indirectly, through advertising (Caves and Murphy, 1976) or other means. Moreover, as noted above, a recent Supreme Court decision *(State Oil v. Khan)* has transformed the *per se* status of maximum resale price maintenance to a rule of reason status, which opens the possibility that franchisors will control franchisee prices more in the U.S. in the future. See Blair and Lafontaine [1998] and Lafontaine [1998] for more on this.

ε), and assume that the business-format franchisee now pays the franchisor a royalty, ρ, per unit sold as well as a fixed franchise fee, F.[38] The retailer's surplus is then

$$(p - \rho)(1 - p + \delta\bar{p} + a) - F - \frac{a^2}{2} - \frac{r}{2}(p - \rho)^2\sigma^2. \tag{20}$$

The agent now chooses effort, a, and price, p, to maximize this surplus, given rival choices, $\bar{p}$ and $\bar{a}$, where the rival is again a franchisee from another chain in the same industry.

The two first-order conditions for the maximization of (20) can be solved to yield the retail reaction functions,

$$p = \frac{1 + r\sigma^2\rho + \delta\bar{p}}{1 + r\sigma^2}, \tag{21}$$

which are clearly upward sloping. Furthermore, in a symmetric equilibrium, the retail price is

$$p_D^* = \frac{1 + r\sigma^2\rho}{1 + r\sigma^2 - \delta}, \tag{22}$$

where D stands for delegation.

Comparative statics, with ρ exogenous to the retailer, yield $dp/dr < 0$, $dp/d\sigma^2 < 0$, $dp/d\delta > 0$, and $dp/d\rho > 0$. Finally, if the retailer is risk neutral or there is no risk, the equilibrium retail price is[39]

$$p_D^* = \frac{1}{1 - \delta}. \tag{23}$$

We compare the delegated situation to the integrated, in which the retailer is a salaried employee, whose wage is F, and ρ is equal to 0. In this case, the manufacturer (who is, as always, assumed risk neutral) chooses the retail price, p, given rival price, $\bar{p}$, which is chosen by the rival manufacturer. In a symmetric equilibrium of the integrated game, the retail price is

---

[38] With traditional franchising, ρ can be interpreted as the wholesale price that the retailer pays to the manufacturer for the product, and F as the fixed rent that he pays for the use of the retail outlet, which we assume is owned by the upstream firm. If there were no rent, or equivalent fixed payment, dealings between principal and agent would be arms length, and the principal would maximize the wholesale, not the total, surplus.

[39] Most of the theoretical papers on this subject assume that there is no uncertainty and thus no moral hazard (e.g., McGuire and Staelin [1983], Vickers [1985], Bonanno and Vickers [1988], and Rey and Stiglitz 1995]).

$$p_I^* = \frac{1}{2 - \delta} \quad , \qquad (24)$$

where *I* stands for integrated. Clearly, if the retailers are risk neutral, principals prefer the delegated situation. Indeed, since reaction functions slope up, when a principal increases the royalty rate to her franchisee, not only does her retailer raise price but also the rival retailer responds with a price increase. In equilibrium, prices and profits are higher as a consequence.[40]

Under agent risk neutrality then, delegation is a dominant strategy. However, as $r\sigma^2$ increases, the advantages of delegation fall. This occurs because the higher retail price is accompanied by an increase in the proportion of the franchisee's income that is variable, thereby increasing the risk that the retailer must bear, and the risk premium he therefore requires. At some level of risk and/or risk aversion, the retailer's need for compensation for bearing increased risk makes vertical separation unattractive, and the firm chooses to vertically integrate instead. On the other hand, the more substitutable the products of the competing chains (the higher is δ), the more firms benefit from delegation (franchising) and thus the more likely it will be chosen. Overall then, this model predicts that vertical separation will be preferred when products are highly substitutable and there is little risk or risk aversion.

One can test these hypotheses individually, and we discussed the relevant literature and results in subsections IIIb and IIIh. Alternatively, a joint test can be constructed from the observation that delegation is more apt to occur when reaction functions are steep, since the slope of the reaction functions is $\delta/(1 + r\sigma^2)$. As with the product-substitutability model, however, these tests require information about each unit's competitors. Slade [1998a], who has such data, finds that delegation is more likely when rival reaction functions are steep, as predicted.[41]

It is interesting to note that once again we come face to face with the prediction that franchising should be discouraged by local-market risk. As we have already discussed, however, the data are inconsistent with this prediction.

IIIj: Multiple Tasks

In many retailing situations the agent performs more than one task. For example, a service-station operator might repair cars as well as sell gasoline, a publican might offer food services as well as beer, and a real-estate agent might rent houses as well as sell properties. Generally, when this is the case, the optimal contract for one task depends on the characteristics of the others [Holmstrom and Milgrom 1991 and 1994].

---

[40] In the terminology of Fudenberg and Tirole [1984], this is a fat-cat game.
[41] In her model, however, there is no risk and therefore no agency cost.

There are many possible variants of multi-task models. We develop a very simple version that illustrates our point. Suppose that there are n tasks and that the agent exerts effort, a_i, on the ith task. Effort results in output according to the linear relationship

$$q = a + \varepsilon, \qquad \varepsilon \sim N(0,\Sigma), \tag{25}$$

where q, a, and ε are vectors of outputs, efforts, and shocks, respectively, and Σ is the variance/covariance matrix of ε. The agent's cost of effort is given by $(a^T a)/2$, and the risk premium is $-(r/2)\,\alpha^T\Sigma\alpha$. First-order conditions for the maximization of the agent's certainty-equivalent income with respect to the vector of effort levels yield $a_i = \alpha_i$, $i = 1,...,n$.

The principal chooses the vector of commissions, α, to maximize the total surplus, which after substitution of the incentive constraint is

$$\alpha^T j - \frac{\alpha^T\alpha}{2} - \frac{r}{2}\,\alpha^T\Sigma\alpha, \tag{26}$$

where j is a vector of ones. First-order conditions for this maximization can be manipulated to yield:

$$\alpha^* = (I + r\Sigma)^{-1} j. \tag{27}$$

In the special case where $n = 2$ and $\sigma_{11} = \sigma_{22} = \sigma^2$, equation (27) simplifies to

$$\alpha_i^* = \frac{1}{1 + r(\sigma^2 + \sigma_{12})}, \quad i = 1,2. \tag{28}$$

If one compares equations (6) and (28) it is clear that, when a second task is added, the power of the agent's incentives in the optimal contract falls (rises) if the associated risks are positively (negatively) correlated. This occurs for pure insurance reasons. In other words, positive correlation means higher risk, whereas negative correlation is a source of risk diversification for the agent.

In this simple model, tasks are linked only through covariation in uncertainty. There are, however, many other possible linkages. For example, the level of effort devoted to one task can affect the marginal cost of performing the other, and, when prices are endogenous, nonzero cross-price elasticities of demand for the outputs can link the returns to effort.

Slade [1996] develops a model that incorporates these three effects and shows that, if an agent has full residual-claimancy rights on outcomes for a second task, the power of incentives for a first task (here gasoline sales) should be lower when the tasks are more complementary. Her empirical application of the model to retail gasoline supports the model's prediction. Specifically, she finds that when the second activity is repairing cars, which is less complementary with selling gasoline than managing a convenience store, agent gasoline-sales incentives are higher powered.

IIIk: Franchise Contract Terms

As noted in the introduction, much of the empirical literature on retail contracting has focused on the dichotomous choice between integration and separation rather than on the terms of the franchise contract. Some authors, however, have examined factors that affect the share parameter, α, directly.[42] Three principal conclusions arise from this set of studies. First, the effects of factors such as risk, the importance of the agent's or the principal's inputs, outlet size, and monitoring difficulty are consistent with those that we have discussed. In other words, factors that tend to increase the degree of separation also tend to increase the agent's share of residual claims. Second, these factors explain a much larger proportion of the variation in the extent of vertical integration than of the variation in share parameters.[43] Thus it appears that firms, in responding to risk, incentive, and monitoring-cost issues, adjust by changing how much they use franchising rather than by altering the terms of their franchise contracts. In that sense, the theoretical models seem to be missing some important aspects of the upstream/downstream relationship. Third, and finally, franchise fees are in general not negatively correlated with royalty rates, despite the fact that the standard principal-agent model suggests that they should be.[44] Instead, fixed fees tend to be set at levels that compensate the franchisor for expenses incurred in setting up a franchised unit.[45]

Other contributions to this literature provide some additional insights. In particular, Lafontaine and Shaw [1996] show that not only are contract terms the same for all franchisees that join a chain at a point in time, as established in the earlier literature, but also they are quite persistent over time. In fact, they find that contract terms are changed very infrequently, that they do not follow any obvious pattern up or down when they are adjusted, and that they do not vary in an obvious way as firms age or grow. Moreover, firm fixed effects account for about 85% of the variation in royalty rates and franchise fees and a very small proportion of this firm-level heterogeneity is related to sectoral differences. They conclude that royalty rates are principally

42 For example, Lafontaine [1992a and 1993], Sen [1993], Rao and Srinivasan [1995], Wimmer and Garen [1996], Gagné et al. [1997], and Lafontaine and Shaw [1996].

43 See Lafontaine [1992a] specifically on this.

44 This prediction results from the fact that, in most theoretical models, the principal is assumed to extract all rent from the agent, an assumption that we have not exploited.

45 See Lafontaine [1992a], Dnes [1993] and Lafontaine and Shaw (1996) on this issue.

determined by differences across firms, differences that likely arise from unobserved heterogeneity in production and monitoring technologies, as well as quality differences.

Finally, several studies examine the use of various franchise contract terms other than royalty rates and franchise fees.[46] For example, Dnes [1993] focuses on franchisor control of leases, and on non-compete covenants, tie-in clauses, and clauses governing the transfer of franchisee assets upon termination. He argues that these clauses act together to protect each party from the potentially opportunistic behavior of the other. Brickley [1997] finds evidence that franchisors impose restrictions on passive ownership, rely on area-development plans, and require mandatory advertising contributions more often when the potential for franchisee free riding is high. He also finds that these contract clauses are complementary. Finally, Mathewson and Winter [1994] show that certain contract clauses, especially exclusive territories and various forms of quantity forcing, occur together in franchise contracts.[47]

IV: Further Comments

Having completed our survey of the factors that determine contract choice and contract terms in franchise markets, we are left with a number of loose ends. In this section, we address issues that we believe are important but that do not lend themselves to being integrated into the framework of section III. In particular, we first discuss one of the most important consequences of franchising -- its effect on the level of retail prices. We then consider the effect of franchising on firm performance, the reasons why franchisors choose to employ a standard set of contracts rather than fine tune each contract to the characteristics of the agent and the market, and the reasons why royalties are typically calculated as a percentage of sales rather than profits. Finally, we address the issue of asset specificity that we touched upon briefly in the introduction.

IVa: Prices at Delegated Outlets

In addition to considering when firms might want to use delegation or integration, empirical research on retail contracting has also been concerned with some consequences of this decision. One area that has received relatively more attention is the effect of contractual form on the final prices that consumers pay.

There are several reasons why prices might be higher at separated outlets. First, some transactions are more costly in a market than inside a firm. For example, contracts written with franchisees are often more complex and thus costlier to write and enforce than those written with employees. Second, because separation involves two firms rather than one, it can introduce an

46 In addition to those specifically mentioned, see Bercovitz [1998a].

47 See Athey and Stern [1997] for theoretical arguments as to why one might expect such complementarities.

additional administrative layer. Third, when retailers have market power, double-marginalization (i.e., successive output restrictions) can arise. Fourth, the existence of spillovers such as those described in section IIIh can lead franchisees to choose prices above those that maximize the chain's profits. Finally, as we showed above, in a strategic model of contracting, separation lowers retailers' perceived elasticities of demand and thus increases retail markups (see also Rey and Stiglitz, [1995]).

Table 7 summarizes results from six studies that are relevant to this issue. Three deal with retail prices of gasoline in the U.S., another deals with prices charged by retailers of separated and integrated soft-drink bottlers, still another involves beer sold in public houses in the U.K., and the last two are concerned with fast-food franchising in certain U.S. submarkets.

Barron and Umbeck [1984] and Slade [1998b] look at legally mandated changes in contractual arrangements (i.e., before and after studies). Muris, Scheffman, and Spiller [1992] also do a before-and-after study in that they focus on the temporal effect on retail prices of soft-drink manufacturers' decisions to buy back some of their bottlers. The other studies look at the effect of contract type on prices in a cross section of contracts, though Lafontaine [1998] looks at both longitudinal and cross-sectional patterns in her data. All seven studies find that, as predicted by theory, increases in the degree of vertical separation, whether voluntary or mandated, result in higher retail prices.

Table 7: The Effect of Vertical Separation on Price

Author	Year	Data	Price Effect
Barron and Umbeck,	1984	Gasoline Service Stations in Maryland	+*
Muris, Scheffman and Spiller	1992	Prices of Retailers Served by Integrated or Separated Soft-Drink Bottlers	+*
Shepard	1993	Gasoline Service Stations in Mass.	+ (and sign. for one product)
Slade	1998b	Beer in the UK	+*
Lafontaine	1998	Fast-Food in Pittsburgh and Detroit Metropolitan Areas	+*
Graddy	1997	Selected Fast -Food Chains in New Jersey and Western Pennsylvania	+*

Note: * indicates a result that is significant in the original study at the 0.05 level, based on a two-tail test.

IVb: Franchising and Firm Performance

Another area that has received attention in the literature is the effect of franchising, or of franchise-contract terms, on firm performance, where firm performance can refer to profitability, service quality, or survival. Shelton's [1967] analysis is a classic in this respect. He uses data on costs, revenues, and profits for outlets in a single chain to examine the effect of switching from franchising to company ownership and from company ownership back to franchising and finds no tendency for revenues to differ according to regime. However, he finds that, under company ownership, costs are higher and thus profits are lower. The major advantage of his study is that its design holds most things constant as the mode of organization changes. Its main drawback is that units were operated under company ownership only when there was no franchisee available or during a transition period. In other words, franchising was the preferred mode, and company ownership was a transitory phase.

Shelton's findings suggests that franchising was indeed more efficient for all units of the firm that he studied. Consequently one might expect company-owned units to under perform in other settings. In a context where firms prefer to own and operate some of their units, Krueger [1991] finds that company employees are paid slightly more and face somewhat steeper earnings profiles than employees in franchised units, which is evidence of higher labor costs in units that are operated by the chain. He argues that the lower powered incentives given to managers of company restaurants make it necessary to offer greater incentives to employees, in the form of efficiency wages and steeper earnings profiles. Thus, consistent with Shelton [1964], Krueger [1991] finds that costs are higher in company units.

As for service quality, Beheler [1991] assesses the effect of company ownership on the health-inspection scores of a sample of 100 fast-food restaurants from 14 chains operating in the St.-Louis metropolitan area. He finds that these scores are poorer for company-owned units.[48]

Turning to the effect of franchise contract terms on performance, Agrawal and Lal [1995] assess how royalty rates affect the level of services provided by franchisees, where these are measured by hours of work per dollar of sales. They find that higher royalty rates lead to lower franchisee services. At the same time, and consistent with a double-sided moral hazard model of franchising, they find that higher royalties lead to greater brand-name investment by franchisors. They measure this investment as a combination (the sum in this case) of four standardized variables, namely advertising expenditures per dollar of sales, the number of franchises in the chain, the number of full-time franchisor staff per dollar of sales, and the number of ongoing

[48] Barron and Umbeck [1984] examine the effect of divorcement, or "forced franchising," of gasoline stations on hours of operation. They find that franchising leads to a reduction in hours, which corresponds to lower quality. This finding, however, as those related to pricing, mostly reflects the ease of setting and controlling hours of operation in company units. In other words, this result occurs because there is no agency problem with respect to hours of operation (or pricing) under vertical integration, but there is one under separation.

services provided by the franchisor. Lafontaine and Shaw [1998a] examine the effect of initial contract terms on franchisor survival five years later and find a positive relationship with both royalty rates and franchise fees. Only the latter, however is significant, suggesting a limited role for royalty rates in affecting future performance.[49]

To summarize, the limited evidence concerning the effect of franchising on performance suggests that lower-powered downstream incentives, in the form of company ownership or of higher royalty rates, tend to lower (raise) franchisee (franchisor) performance. However, much more work is needed in this area before one can draw more definitive conclusions.

IVc: Within-Firm Contract Uniformity

Though our model in Section III did not highlight this, most theoretical contracting models imply that the principal should tailor the terms of the contract to suit the characteristics of the agent, the outlet, and the market. In other words, equation (1) is the output/effort relationship for a particular franchisee and franchisor pair, and for a particular local market. It is clear then that the optimal share parameter, α^*, should differ by outlet within a chain as well as across chains. Contracts that are observed in practice, in contrast, are remarkably insensitive to variations in individual, outlet, and market conditions. Indeed, most firms use a standard business-format franchising contract -- a single royalty-rate and franchise-fee combination -- for all of their franchised operations that join the chain at a point in time. The same lack of variation is observed in traditional franchising, where a manufacturer often charges the same wholesale price to all of her leased operations.[50] When this is true, the only choice that the principal makes in the end is whether to franchise or to self operate. In other words, when the characteristics of individual units differ, the upstream firm chooses to vertically integrate those units with characteristics that require less high-powered incentives, and to franchise those that require more, which explains the focus in empirical work on the choice between integration and separation rather than on the terms of the contract.

Models that emphasize incentive issues for both parties -- double-sided moral-hazard models -- provide one possible explanation for this lack of contract fine tuning. These models recognize that, with most franchising arrangements, not only does the agent have to provide effort,

[49] See also Shane [1997] on the effect of franchise contract terms on instantaneous survival, and Shane and Azoulay [1998] on the effect of exclusive territories on survival. For assessments of franchisor survival rates, see Price [1996], Shane [1996] and Stanworth [1996]. For the effect of franchising on small-business survival, see Williams (1998), Bates [1998], and the references therein.

[50] In the U.S., the Robinson-Patman Act requires wholesale-price uniformity, at least locally. This is not true, however, in Canada. Nevertheless, price uniformity across buyers is common there as well (e.g., in gasoline markets; see Slade [1996 and 1998a] on this). Also, the Robinson-Patman act does not explain contract uniformity in business-format franchising, as the Act applies to the sale of commodities, which do not include franchising rights. See McAfee and Schwartz [1994] and Bhattacharyya and Lafontaine [1995] for further arguments against legal constraints as the main source of contract uniformity in business-format franchising.

but also the principal must maintain the value of the trademark or company logo. With moral hazard on the part of both parties, even when both are risk neutral, an optimal contract involves revenue sharing.[51]

In such a double-sided moral-hazard context, Bhattacharyya and Lafontaine [1995] show that, under specific assumptions concerning functional forms, the benefits of customizing contracts can be quite limited, if not zero. This implies that the optimal contract is insensitive to many relationship-specific circumstances. In addition, their model might at least partially explain the persistence of uniform contract terms over time found by Lafontaine and Shaw [1996]. Indeed, in the Bhattacharyya and Lafontaine model, the terms of the optimal contract remain unchanged as the franchise chain grows.[52]

Other reasons that have been advanced in the literature to explain the lack of customization involve the high costs of customizing, either the direct cost of designing and administering many different contracts (Holmström and Milgrom [1987] and Lafontaine [1992b]) or the high potential for franchisor opportunism that arises when contracts can vary (McAfee and Schwartz [1994]).

Whatever the reason for the lack of customization in franchise contracting, it remains that most of the empirical research has focused either on the discrete choice to operate a unit as a franchise or not (when the data consist of individual contracts) or on the fraction of a franchisor's units that are franchised (when the data consist of firm proportions). One might therefore ask if the same factors that lead to granting higher-powered incentives in the fine-tuning case also lead to a higher fraction of franchised outlets in the uniform-contract case. We construct a formal model in which this is the case.

Suppose that each outlet or unit is associated with some characteristic x that affects its profitability, and let the expected profitability of that unit depend on the power of the agent's incentives as well as on this characteristic. One can express this relationship as $E\pi(\alpha,x)$. We assume that A: the expected profit function is concave, and B: $E\pi_{\alpha x} > 0$. In other words, as x increases, the marginal profitability of higher-powered incentives also increases.[53]

With the fine-tuning model in which contracts are outlet specific, the principal's problem is to choose α_i to maximize $E\pi(\alpha_i, x_i)$ for each unit i, subject to the agent's incentive constraint. The

---

[51] See e.g. Rubin [1978], Lal [1990], and Bhattacharyya and Lafontaine [1995]. Carmichael [1983] has shown that with two agents or more, and moral hazard on the principal's side as well as the agents', the first best can be achieved with a contract based on relative outputs. However, we do not observe this type of contract in franchising. Why this is the case is beyond the scope of the present paper.

[52] More specifically, Bhattacharyya and Lafontaine [1995] show that, when the production function is Cobb-Douglas and the cost-of-effort function is exponential, the optimal share parameter is independent of the scale of operation, and, as a result, of the level of demand and the degree of competition in the market. The share parameter is also independent of both parties' cost-of-effort parameters.

[53] For example, x might be the importance of the agent's effort or the negative of the agent's degree of risk aversion.

first-order condition for this maximization can be solved to yield the optimal contract, $\alpha^*(x)$. Moreover, assumption B guarantees that $d\alpha^*/dx > 0$.

Now suppose that fine tuning is sufficiently expensive so that the principal offers only two contracts, a franchise contract with $\alpha > 0$ and a vertical integration contract with $\alpha = 0$. Moreover, the power of incentives (α) is the same for all franchisees. If the principal has n units, one can order those units such that $x_1 \leq x_2 \leq ... x_n$. Now the principal's problem is to

$$\max_{\alpha, i^*} \Sigma_{i \geq i^*} E\pi(\alpha, x_i) + \Sigma_{i < i^*} E\pi(0, x_i). \qquad (29)$$

Given i^*, the optimal contract $\alpha^*(i^*)$ can be obtained from the first-order condition $\Sigma_{i \geq i^*} E\pi_\alpha = 0$, and given α, the optimal i^* is chosen such that (i) $E\pi(\alpha, x_{i^*}) - E\pi(0, x_{i^*}) \geq 0$, and (ii) $E\pi(\alpha, x_{i^*-1}) - E\pi(0, x_{i^*-1}) < 0$.[54] In this uniform-contract situation, an exogenous increase in x at some of a company's units leads to both higher powered incentives (higher α^*) and to a larger fraction of outlets franchised (lower i^*).

IVd: Why Royalties on Sales?

With most variants of the model of section III, price is normalized to one and there are no input costs other than agent effort. As a result, there is no operational difference between royalties on sales, input markups, and royalties on profits. Indeed, most models of retail contracting make no distinction among these possibilities. In reality, however, business-format contracts usually involve royalties on sales.[55] The puzzling issue then is why business-format-franchise contracts systematically emphasize “sales sharing” rather than profit sharing. For example, Lafontaine [1992b] conducted a survey and found that, of the 127 franchisors who responded to this question, 123 charged some form of royalties. Of these, 112 asked for a percentage of sales or revenues. Only two franchisors requested a proportion of profits, while another four were paid a proportion of gross margins.[56]

The traditional explanation for the use of sales rather than profit-based royalties is that the latter are too difficult to measure. For example, franchisees can pad their costs by including personal cars and salaries for family members, and cost padding can be difficult to observe or to

[54] We are assuming an interior solution with $1 < i^* < n$. Assumption B guarantees that the left-hand side of (i) > the left-hand side of (ii) for any i.

[55] In traditional franchise agreements, the franchisor sells a manufactured product to the franchisee who then resells it. Assuming that the franchisee has little leeway on prices, input markups are equivalent to royalties on sales. See Dnes [1993] and Lafontaine [1993] and the references therein for more on this and on tying in business-format franchising.

[56] Of the remaining 5 firms, 4 charged a fixed amount per time period, and one did not answer this part of the question.

contract upon. However, this measurement argument does not explain why franchisors do not collect a proportion of gross margins more often.

Rubin [1978] proposes a more substantive explanation for sales sharing: he argues that franchisee effort controls costs as well as stimulating demand. Franchisor effort, in contrast, only affects demand. Consequently, franchisees should be given full residual claimancy on cost reductions, whereas franchisors should be paid some proportion of sales so that they have incentives to maintain the value of the trade name. Maness [1996] formalizes this argument by assuming that costs are noncontractible, and as such must be borne by the owner of the outlet. Thus the decision to franchise (to have the franchisee own the outlet) or operate directly (to have the company own the unit) hinges on which party is better at controlling unit costs. Furthermore, the sharing rule must allow the owner of the unit to cover the costs of operation and thus satisfy his or her individual-rationality constraint. Therefore, in contrast to say sharecropping, where the 50/50 rule for output often applies, royalty rates in franchise agreements are low, typically between 5 and 10%.[57]

IVe: Asset Specificity

Asset specificity is an important area of the theoretical literature that we have, up to now, had little to say about. We made this choice because we believe that it is far less important for retail contracting than for the purchase and sale of intermediate inputs. As a result, we don't think it sheds much new light on the empirical regularities highlighted herein. Nevertheless, as this issue regularly surfaces in the literature, we discuss how we arrived at this conclusion.

The positive effect of unit size on company ownership has been interpreted by some (e.g. Brickley and Dark [1987] and Scott [1995]) as evidence that franchisors find it more costly to rely on franchising when franchisees are required to make large relationship-specific investments. We, however, find no evidence that total investment relates positively to asset specificity in retail contracting. For example, the largest gasoline stations are high-volume self-service stations that are the least specialized. The owner of such a station, if terminated by one refiner, could easily obtain a contract with another. The value of his assets should therefore not be significantly lower outside of the relationship. The same is true in business-format franchising. Within this group, the hotel industry requires the largest absolute level of investment. This investment, however, is again not specific; hotel banners are routinely changed with little effect on property values. Our point is that overall investment is not a good measure of asset specificity.[58]

57 For another argument on this issue, see also Lewin (1997).

58 See Dnes [1993] and Wimmer and Garen [1996] for attempts to capture the part of total investment that is specific.

Klein [1995] notes that, from an incentive perspective, what matters is not the level of specific investment by franchisees, as these are sunk and should not affect behavior, but rather the rents or quasi rents that the franchisee can expect to lose if he is terminated.[59] Moreover, Dnes [1993] finds that franchisees' specific investments are protected by the terms of franchise agreements. More specifically, he argues that the contractual clauses that govern the transfer of franchisee assets upon termination are set such that "*if the franchisor withdraws from a contract and offers to buy assets (even if this follows the franchisee offering assets for sale), then the prices are governed by something other than just the franchisor's wishes,*" be it arbitration or some notion of fair-market value. (p. 390)[60] Presumably, units of franchisees who are terminated for disciplinary reasons are viable, and franchisors will want to buy them or to allow other franchisees to do so. Consequently, upon termination, the current franchisee does not forego the rents that are attached to specific assets, and in that sense, these rents cannot play a self-enforcement role. On the other hand, other rents are lost by franchisees upon termination. In particular, the non-compete clauses that are found in most franchise contracts make it difficult for franchisees to put the human capital they have accumulated within the chain to good use upon termination. Similarly, given that franchisees are often allowed to expand their business by owning additional outlets in a chain, whatever rents are associated with the right to purchase these extra units are foregone upon termination from a franchised system.[61] However, the value of such rents is not well captured by measures of specific investments.

V: Final Remarks

Our survey of retail contracting under exclusive marks has highlighted the existence of many stylized facts and the robust nature of the evidence. Indeed, in almost every case where a factor is statistically significant, its effect on the power of agent incentives in real-world contracts is the same across studies. In other words, in spite of the fact that researchers assess different industries over different time periods using a number of proxies for a given factor, their empirical findings are usually consistent with one another.

---

[59] While specific assets can generate rents, they are not necessary; downstream rents can also arise because franchisors choose to leave them with franchisees (see Kaufmann and Lafontaine [1994] and Michael and Moore [1995] for evidence that some franchisors choose to do this).

[60] Dnes argues that franchisees sustain a loss if they fail and their franchisor decides not to buy back their unit because it is not viable, and that this loss is larger the more specific assets are involved. He argues that such potential losses give franchisees incentives to get involved only if they truly are able to perform as they say they are. Thus he concludes that specific investments can serve a franchisee-screening function.

[61] See for example Kaufmann and Dant [1996] and Kalnins and Lafontaine [1998] on multi-unit ownership in franchised chains. Also see Bradach (1997) on the importance of additional units for franchisee growth and statements that refusing to grant extra units to franchisees serves a disciplinary role.

The theories, on the other hand, are much more fragile. In fact, in order to obtain a tractable model, it is important to use simple specifications for agent utility and risk preference. The results of the model then can depend non-trivially on these assumptions. Furthermore, the way in which the unobservable risk factors interact with the tangible variables is also crucial, as we have demonstrated in our discussion of outlet size. Nevertheless, we hope that our attempt to organize the evidence in a unified framework will be helpful to theorists in that it gives them a set of stylized facts to explain, and to applied researchers by providing a framework and a sense of where more work would be most beneficial.

One theoretical prediction, however, is not very fragile; it surfaces over and over again. We refer to the effect of risk on agent incentives. Whether one considers the simplest incentive/insurance model or imbeds this model in one with endogenous prices and strategic delegation or one with multiple tasks and linked efforts, the theory predicts that more risky units should tend to be operated by the parent company. The evidence, however, strongly rejects this predicted tendency. We have suggested one possible explanation for the discrepancy between theory and evidence -- endogenous output variability in a situation where agents have private information about local-market conditions. However, a similar finding surfaces in the sharecropping literature (see Allen and Lueck [1995] for a survey), an area where exogenous output fluctuations are apt to dominate endogenous fluctuations. Given the central role that agent risk plays in the incentive-contracting literature, and given the strength of the empirical evidence, we believe that this puzzle in particular deserves further attention.

References Cited:

Agrawal, D. and R. Lal (1995), "Contractual Arrangements in Franchising: An Empirical Investigation," *Journal of Marketing Research*, 32, 213-221.

Alchian, A.A. and H. Demsetz (1972), "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777-795.

Allen, D.W. and D. Lueck (1992), "Contract Choice in Modern Agriculture: Cropshare versus Cash Rent," *Journal of Law and Economics*, 35, 397-426.

Allen, D.W. and D. Lueck (1995), "Risk Preferences and the Economics of Contracts," *American Economic Review*, 85, 447-451.

Anderson, E. (1985), "The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis," *Marketing Science*, 4, 234-254.

Anderson, E. and R.L. Oliver (1987), "Perspectives on Behavior-Based Versus Outcome-Based Salesforce-Control Systems, *Journal of Marketing*, 51, 76-88.

Anderson, E. and D. Schmittlein (1984), "Integration of the Sales Force: An Empirical Examination," *Rand Journal of Economics, 15*, 385-395.

Athey, S. and S. Stern (1997), "An Empirical Framework for Testing Theories about Complementarities in Organizational Design," MIT, mimeo.

Bates, T., (1998), "Survival Patterns Among Newcomers to Franchising," *Journal of Business Venturing*, 13, 113-130.

Barron, J.M. and J.R. Umbeck (1984), "The Effects of Different Contractual Arrangements: The Case of Retail Gasoline," *Journal of Law and Economics*, 27, 313-328.

Beheler, R.L. (1991), "Control in Various Organizational Forms: An Empirical Study of Company-Owned and Franchisee-Owned Units' Health Inspections," in J.R. Nevin (Ed.), *Franchising: Embracing the Future*, Fifth Annual Proceedings of the Society of Franchising, University of St-Thomas.

Bercovitz, J.E.L. (1998a), "An Analysis of Contractual Provisions in Business-Format Franchise Agreements," mimeo, Fuqua School of Business, Duke University.

Bercovitz, J.E.L. (1998b), "Franchising vs. Company Ownership," mimeo, Fuqua School of Business, Duke University.

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

Blair, R.D. and D. L. Kaserman, (1994), "A Note on Incentive Incompatibility under Franchising," *Review of Industrial Organization*, 9, 323-330.

Blair, R.D. and F. Lafontaine (1998), "Will *Khan* Foster or Hinder Franchising? An Economic Analysis of Maximum Resale Price Maintenance", mimeo, University of Michigan Business School.

Bonanno, G. and J. Vickers (1988), "Vertical Separation," *Journal of Industrial Economics*, 36, 257-266.

Bradach, J.L. (1997), "Using the Plural Form in the Management of Restaurant Chains," *Administrative Science Quarterly*, 42, 276-303.

Brickley, J.A., (1997), "Incentive Conflicts and Contracting: Evidence from Franchising," mimeo, Simon Graduate School of Business, University of Rochester.

Brickley, J. and F. Dark (1987), "The Choice of Organizational Form: The Case of Franchising," *Journal of Financial Economics*, 18, 401-420.

Brickley, J.A., F.H. Dark, and M.S. Weisbach (1991), "An Agency Perspective on Franchising," *Financial Management*, 20, 27-35.

Carmichael, H.L. (1983), "The Agent-Agents Problem: Payment by Relative Output," *Journal of Labor Economics*, 1, 50-65.

Carney, M. and E. Gedajlovic (1991), "Vertical Integration in Franchise Systems: Agency Theory and Resource Explanations," *Strategic Management Journal*, 12, 607-629.

Caves, R.E. and W.F. Murphy (1976), "Franchising: Firms, Markets, and Intangible Assets," *Southern Economic Journal*, 42, 572-586.

Coase, R. (1937), "The Nature of the Firm," *Economica*, 4, 386-405.

Coughlan, A.T. (1985), "Competition and Cooperation in Marketing Channel Choice: Theory and Application," *Marketing Science*, 4, 110-129.

Crocker, K.J. and K.J. Reynolds (1992), "The Efficiency of Incomplete Contracts: An Empirical Analysis of Air Force Engine Procurement," *RAND Journal of Economics*, 24, 126-146.

Crocker, K. J. and S. E. Masten (1996), "Regulation and Administered Contracts Revisited: Lessons from Transaction-Cost Economics for Public Utility Regulation," *Journal of Regulatory Economics*, 9 (1), 5-39.

Dant, R.P., P.J. Kaufmann and A.K. Paswan (1992), "Ownership Redirection in Franchised Channels," *Journal of Public Policy and Marketing*, 11, 33-44.

Dnes, A. W. (1992), *Franchising: A Case-study Approach,* Ashgate Publishing Ltd., Aldershot, England.

Dnes, A.W. (1993), "A Case Study Analysis of Franchise Contracts", *Journal of Legal Studies,* 22, 367-393.

Dnes, A. W. (1996), "The Economic Analysis of Franchise Contracts," *Journal of Institutional and Theoretical Economics*, 152, 297-324.

Elango, B. and V. H. Fried (1997), "Franchising Research: A Literature Review and Synthesis," *Journal of Small Business Management*, 35, 68-82.

Fudenberg, D. and J. Tirole (1984), "The Fat-Car Effect, The Puppy-Dog Ploy, and the Lean-and-Hungry Look," *American Economic Review*, 74, 361-366.

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

Gagné, R., S.P. Sigué, and G. Zaccour (1998), "Droit d'entrée et taux de redevance dans les franchises d'exploitation au Québec," *Actualité Economique,* forthcoming.

Gallini, Nancy T., and Nancy A. Lutz. (1992), "Dual Distribution and Royalty Fees in Franchising," *Journal of Law, Economics, & Organization*, 8, 471-501.

Garen, J.E. (1994), "Executive Compensation and Principal-Agent Theory," *Journal of Political Economy*, 102, 1175-1199.

Graddy, K. (1997), "Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area," *Journal of Business and Economic Statistics*, 15, 391-401.

Grossman, S.J. and O.D. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691-719.

Hart, O. and J. Moore (1988), "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755-786.

Holmstrom, B. (1979), "Moral Hazard and Observability," *Bell Journal of Economics,* 10, 74-91.

Holmstrom, B. (1982), "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324-340.

Holmstrom, B. and P. Milgrom (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303-328.

Holmstrom, B. and P. Milgrom (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7, 24-51.

Holmstrom, B. and P. Milgrom (1994), "The Firm as an Incentive System," *American Economic Review*, 84, 972-991.

Ichniowski, C., K.L Shaw and G. Prennushi (1997), "The Effects of Human Resource Management Practices on Productivity: Evidence from the Steel Industry," *American Economic Review*, 87, 291-313.

Jensen, M. and K. Murphy (1990), "Performance Pay and Top-Management Incentives," *Journal of Political Economy*, 98, 225-264.

John, G. and B.A. Weitz (1988), "Forward Integration into Distribution: An Empirical Test of Transaction Cost Analysis," *Journal of Law, Economics, and Organization*, 4, 337-355.

Joskow, P. (1988), "Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence," *Journal of Law, Economics, and Organization*, 4, 95-117.

Kalnins, A. and F. Lafontaine (1998), "Incentive and Strategic Motives for Vertical Separation: Evidence from Location Patterns in the Texan Fast-Food Industry," mimeo, University of Michigan Business School.

Kaplan, S.N. (1994), "Top Executive Rewards and Firm Performance: A Comparison of Japan and the U.S.," *Journal of Political Economy*, 102, 510-546.

Kaufmann, P. J. and R. P. Dant (1996), "Multi-Unit Franchising: Growth and Management Issues," *Journal of Business Venturing*, 11, 343-358.

Kaufmann, P.J. and Lafontaine, F. (1994), "Costs of Control: The Source of Economic Rents for McDonald's Franchisees," *Journal of Law and Economics*, 37, 417-543.

Kehoe, M. R. (1996), "Franchising, agency problems, and the cost of capital," *Applied Economics*, 28, 1485-1493.

Klein, B. (1980), "Transaction Cost Determinants of 'Unfair' Contractual Arrangements," *American Economic Review*, 70, 356-362.

Klein, B. (1988), "Vertical Integration as Organizational Ownership: The Fisher Body - General Motors Relationship Revisited," *Journal of Law, Economics, and Organization*, 4, 199-213.

Klein, B. (1995), "The Economics of Franchise Contracts," *Journal of Corporate Finance*, 2, 9-37.

Klein, B., R. Crawford, and A. Alchian (1978), "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21, 297-326.

Knight, F. (1921), *Risk, Uncertainty, and Profit*, Houghton Mifflin Pub., Chicago, USA.

Krueger, A. B., (1991), "Ownership, Agency and Wages: An Examination of the Fast Food Industry," *Quarterly Journal of Economics*, 106, 75-101.

Lafontaine, F. (1992a), "Agency Theory and Franchising: Some Empirical Results," *RAND Journal of Economics*, 23, 263-283.

Lafontaine, F. (1992b), "How and Why do Franchisors do What They do: A Survey Report", in P.J. Kaufmann (Ed.), *Franchising: Passport for Growth & World of Opportunity*, Sixth Annual Proceedings of the Society of Franchising, University of St-Thomas.

Lafontaine, F. (1993), "Contractual Arrangements as Signaling Devices: Evidence from Franchising," *Journal of Law, Economics, and Organization*, 9, 256-289.

Lafontaine, F. (1995), "Pricing Decisions in Franchised Chains: A Look at the Fast-Food Industry," NBER Working paper #5247.

Lafontaine, F. (1998), "Retail Pricing, Organizational Form, and the New Rule of Reason Approach to Maximum Resale Prices," mimeo, University of Michigan Business School.

Lafontaine, F. and Bhattacharyya, S., 1995, "The Role of Risk in Franchising", *Journal of Corporate Finance*, 2, 39-74.

Lafontaine, F. and P.J. Kaufmann (1994), "The Evolution of Ownership Patterns in Franchise Systems," *Journal of Retailing*, 70, 97-113.

Lafontaine, F. and K.L. Shaw, (1998a), "Franchising Growth and Franchisor Entry and Exit in the U.S. Market: Myth and Reality," *Journal of Business Venturing*, 13,95-112.

Lafontaine, F. and K.L. Shaw, (1998b), "Targeting Managerial Control: Evidence from Franchising," mimeo, University of Michigan Business School.

Lafontaine, F. and Shaw, K.L., (1999) "The Dynamics of Franchise Contracting: Evidence from Panel Data", *Journal of Political Economy*, forthcoming.

Lafontaine, F. and Slade, M.E. (1996), "Retail Contracting and Costly Monitoring: Theory and Evidence", *European Economic Review Papers and Proceedings*, 40, 923-932.

Lafontaine, F. and Slade, M.E. (1997), "Retail Contracting: Theory and Practice," *Journal of Industrial Economics*, 45, 1-25.

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

Lazear, E. P. (1996), "Performance Pay and Productivity," NBER Working Paper #5672.

Leffler, K.B. and R.R. Rucker (1991), “Transaction Costs and the Efficient Organization of Production: A Study of Timber-Harvesting Contracts,” *Journal of Political Economy*, 99, 1060-1087.

Lewin, S. (1998), “Autonomy, Contractibility, and the Franchise Relationship,” mimeo, Iowa State University.

Lutz, N.A. (1995), “Ownership Rights and Incentives in Franchising,” *Journal of Corporate Finance*, 2, 103-130.

Lyons, B.R. (1996), “Empirical Relevance of Efficient Contract Theory: Inter-Firm Contracts,” *Oxford Review of Economic Policy*,” 12, 27-52.

Maness, R. (1996), “Incomplete Contracts and the Choice Between Vertical Integration and Franchising,” *Journal of Economic Behavior and Organization*, 31, 101-115.

Martin, R.E. (1988), “Franchising and Risk Management,” *American Economic Review*, 78, 954-968.

Masten, S.E. (1984), “The Organization of Production: Evidence from the Aerospace Industry,” *Journal of Law and Economics*, 27, 403-417.

Masten, S.E. (1985), “Efficient Adaptation in Long-Term Contracts: Take-or-Pay Provisions for Natural Gas,” *American Economic Review*, 75, 1083-93.

Masten, S.E., [1998], “Contractual Choice,” *Encyclopaedia of Law & Economics*, B. Boukaert and G. De Geest ed., Edward Elgar Publishing, forthcoming.

Mathewson, F. and R. Winter (1985), “The Economics of Franchise Contracts,” *Journal of Law and Economics*, 28, 503–526.

McAfee, R.P. and M. Schwartz (1994), “Multilateral Vertical Contracting: Opportunism, Nondiscrimination, and Exclusivity”, *American Economic Review*, 84, 210–230.

McGuire T.W. and R. Staelin, (1983), “An Industry Equilibrium Analysis of Downstream Vertical Integration,” *Marketing Science*, 2, 161-192.

Michael, S.C. and H.J. Moore (1995), “Returns to Franchising, “ *Journal of Corporate Finance*, 2, 133-156.

Milgrom, P. and J. Roberts (1992), *Economics, Organization and Management*, Englewoods Cliffs: Prentice-Hall Inc.

Minkler, A. (1990), “An Empirical Analysis of a Firm's Decision to Franchise,” *Economics Letters*, 34, 77-82.

Minkler, A. and T. A. Park (1994), “Asset Specificity and Vertical Integration,” *Review of Industrial Organization*, 9, 409-423.

Mirrlees, J.A. (1976), “The Optimal Structure of Incentives and Authority Within an Organization,” *Bell Journal of Economics*, 7, 105-131.

Monteverde, K. and D. Teece (1982), “Supplier Switching Costs and Vertical Integration in the Automobile Industry,” *Bell Journal of Economics*, 13, 206-213.

Muris, T.J., D.T. Scheffman, and P.T. Spiller (1992), “Strategy and Transaction Costs: The Organization of Distribution in the Carbonated Soft Drink Industry,” *Journal of Economics & Management Strategy*, 1, 83-128.

Murphy, K. (1984), “Incentives, Learning, and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts,” *RAND Journal of Economics*, 17, 59-76.

Norton, S.W. (1988), “An Empirical Look at Franchising as an Organizational Form” *Journal of Business*, 61: 197-217.

Ozanne, U.B. and S.D. Hunt, (1971), *The Economic Effect of Franchising*, U.S. Senate, Select Committee on Small Business, Washington, Gov. Printing Office.

Price, S. (1996), “Behind the veneer of success: Propensities for UK franchisor failure,” *Small Business Research Trust Report*, May.

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

Rey, P. and J. Stiglitz (1995), “The Role of Exclusive Territories in Producers' Competition,” *RAND Journal of Economics*, 26, 431-451.

Rindfleisch, a and J.B. Heide (1997), “Transaction Cost Analysis: Past, Present and Future Applications,” *Journal of Marketing*, 61, 30-54.

Romano, R. E. (1994), “Double Moral Hazard and Resale Price Maintenance,” *RAND Journal of Economics*, 25, 455-466.

Rubin, P. (1978), “The Theory of the Firm and the Structure of the Franchise Contract,” *Journal of Law and Economics*, 21, 223-233.

Scott, F.A. (1995), “Franchising vs. Company Ownership as a Decision Variable of the Firm,” *Review of Industrial Organization*, 10, 69-81.

Sen, K. C. (1993), “The Use of Initial Fees and Royalties in Business Format Franchising,” *Managerial and Decision Economics*, 14, 175–190.

Shane, S.A. (1996), "Hybrid Organizational Arrangements and their Implications for Firm Growth and Survival: A Study of New Franchisors," *Academy of Management Journal*, 39, 216-234.

Shane, S.A. (1997), "Organizational Incentives and Organizational Mortality," mimeo, Sloan School of Management, MIT.

Shane, S.A. and P. Azoulay (1998), "Contract Design and Firm Survival," mimeo, Sloan School of Management, MIT.

Shelanski, H.A. and Klein, P. G. (1995), "Empirical Research in Transaction Cost Economics: A Review and Assessment," *Journal of Law, Economics, and Organization*, 11, 335-361.

Shelton, J. (1967), "Allocative Efficiency vs. 'X-Efficiency': Comment," *American Economic Review*, 57, 1252-1258.

Shepard, A. (1993), "Contractual Form, Retail Price, and Asset Characteristics," *Rand Journal of Economics*, 24, 58-77.

Slade, M.E. (1996), "Multitask Agency and Contract Choice: An Empirical Assessment," *International Economic Review*, 37, 465-486.

Slade, M.E. (1998a), "Beer and the Tie: Did Divestiture of Brewer-Owned Public Houses Lead to Higher Beer Prices?", *Economic Journal*, 108, 1-38.

Slade, M.E. (1998b), "Strategic Motives for Vertical Separation: Evidence from Retail Gasoline," *Journal of Law, Economics, & Organization*, 14, 84-113.

Smith II, R. L. (1982), "Franchise Regulation: An Economic Analysis of State Restrictions on Automobile Distribution," *Journal of Law and Economics*, 25, 125-157.

Stanworth, J., 1996, "Dispelling the myths surrounding franchise failure rates - some recent evidence from Britain," *Franchising Research: An International Journal*, 1, 25-28.

State Oil Co. v. Khan, US Sup. CT, No. 96-871.

Stiglitz, J.E. (1974), "Incentives and Risk-Sharing in Sharecropping," *Review of Economic Studies*, 41, 219-255.

Thompson, R.S. (1994), "The Franchise Life Cycle and the Penrose Effect," *Journal of Economic Behavior and Organization*, 24, 207-218.

U.S. Department of Commerce, (1998)*Franchising in the Economy, 1986-1988*, prepared by Andrew Kostecka.

Vickers, J. (1985), "Delegation and the Theory of the Firm," *Economic Journal*, 95, 138-147.

Williams, D. (1998), “Why do Entrepreneurs Become Franchisees”, *Journal of Business Venturing*, forthcoming.

Williamson, O.E. (1971), “The Vertical Integration of Production: Market Failure Considerations,” *American Economic Review*, 61, 112-123.

Williamson, O.E. (1979), “Transaction Cost Economics: The Governance of Contractual Relations,” *Journal of Law and Economics*, 22, 233-262.

Williamson, O.E. (1983), “Credible Commitments: Using Hostages to Support Exchange,” *American Economic Review*, 73, 519-540.

Williamson, O.E.(1985), *The Economic Institutions of Capitalism*, Free Press, New-York, NY.

Wimmer, B.S. and J.E. Garen (1997), “Moral Hazard, Asset Specificity, Implicit Bonding, and Compensation: The Case of Franchising,” *Economic Inquiry*, 35, 544-554

Appendix

Algebraic Derivations

In each case below, the agent (A) maximizes his certainty-equivalent income, $E(y) - r/2\ Var(y)$, whereas the principal (P) maximizes the expected total surplus -- expected output minus the agent's cost of effort minus the agent's risk premium -- $E(q) - a^2/2 - r/2\ Var(y)$. With one exception, noted below, the agent's compensation in each case is given by $S(q) = \alpha q + W$. The cases differ according to the specification of the function that maps effort into output, $q = f(a,\varepsilon,\Theta)$.

i) Risk: $q = a + \varepsilon.$

A: $$\max_a \ \alpha a + W - \frac{a^2}{2} - \frac{r}{2}\ \alpha^2\sigma^2, \quad \text{foc: } a = \alpha.$$

Substituting the agent's effort choice into the principal's problem yields:

P: $$\max_\alpha \ \alpha - \frac{\alpha^2}{2} - \frac{r}{2}\ \alpha^2\sigma^2, \quad \text{foc: } 1 - \alpha - r\,\sigma^2\alpha = 0.$$

$$\alpha^* = \frac{1}{1 + r\sigma^2} \ .$$

$$\frac{d\alpha^*}{dr} = -\frac{\sigma^2}{\Gamma^2} < 0, \quad \frac{d\alpha^*}{d\sigma^2} = -\frac{r}{\Gamma^2} < 0, \quad \Gamma = 1 + r\,\sigma^2.$$

ii) Agent Effort: $q = \eta a + \varepsilon.$

A: $$\max_a \ \alpha\eta a + W - \frac{a^2}{2} - \frac{r}{2}\ \alpha^2\sigma^2, \quad \text{foc: } a = \alpha\eta.$$

P: $$\max_\alpha \ \eta^2\alpha - \frac{\alpha^2\eta^2}{2} - \frac{r}{2}\ \alpha^2\sigma^2, \qquad \text{foc: } \eta^2 - \alpha\eta^2 - r\alpha\sigma^2 = 0.$$

$$\alpha^* = \frac{\eta^2}{\eta^2 + r\sigma^2} \ .$$

$$\frac{d\alpha^*}{d\eta} = \frac{(\eta^2 + r\sigma^2)2\eta - \eta^2 2\eta}{\Gamma^2} = \frac{2\eta r\sigma^2}{\Gamma^2} > 0, \quad \Gamma = (\eta^2 + r\sigma^2).$$

iii) Outlet Size: $q = \eta a + (\gamma + \varepsilon)k.$

A: $$\max_a \quad \alpha(\eta a + \gamma k) + W - \frac{a^2}{2} - \frac{r}{2} \alpha^2 \sigma^2 k^2, \qquad \text{foc: } a = \alpha\eta.$$

P: $$\max_\alpha \quad \eta^2\alpha - \frac{\eta^2\alpha^2}{2} - \frac{r}{2} \alpha^2 \sigma^2 k^2, \qquad \text{foc: } \eta^2 - \eta^2\alpha - r\alpha\sigma^2 k^2 = 0.$$

$$\alpha^* = \frac{\eta^2}{\eta^2 + r\sigma^2 k^2} .$$

$$\frac{d\alpha^*}{dk} = -\frac{2\eta^2 r\sigma^2 k}{\Gamma^2} < 0, \qquad \Gamma = \eta^2 + r\sigma^2 k^2.$$

iv) Costly Monitoring: This result is derived in Lafontaine and Slade (1996).

v) Franchisor Effort: $q = \eta a + \theta b + \varepsilon.$

This problem has two incentive constraints:

A: $$\max_a \quad \alpha(\eta a + \theta b) + W - \frac{a^2}{2} - \frac{r}{2} \alpha^2\sigma^2, \qquad \text{foc: } a = \alpha\eta.$$

P: $$\max_b \ (1 - \alpha)(\eta a + \theta b) - \frac{b^2}{2} , \qquad \text{foc: } b = (1 - \alpha)\theta.$$

α is chosen to maximize the total surplus, $\eta a + \theta b - \frac{a^2}{2} - \frac{b^2}{2} - \frac{r}{2} \alpha^2\sigma^2$, subject to the two incentive constraints. Substituting, we have:

$$\max_\alpha \quad \alpha\eta^2 + (1 - \alpha)\theta^2 - \frac{\alpha^2\eta^2}{2} - \frac{(1 - \alpha)^2\theta^2}{2} - \frac{r}{2} \alpha^2\sigma^2,$$

foc: $\eta^2 - \theta^2 - \alpha\eta^2 + (1 - \alpha)\theta^2 - r\alpha\sigma^2 = 0,$

$$\alpha^* = \frac{\eta^2}{\eta^2 + \theta^2 + r\sigma^2} \quad .$$

$$\frac{d\alpha^*}{d\theta} = -\frac{2\eta^2\theta}{\Gamma^2} < 0, \frac{d\alpha^*}{d\eta} = \frac{2\eta(\theta^2 + r\sigma^2)}{\Gamma^2} > 0, \Gamma = \eta^2 + \theta^2 + r\sigma^2.$$

vi) Spillovers Within the Chain:

$q = 1 - p - \mu\bar{p} + a + \varepsilon$, $\bar{p}$ = price at another outlet in the same chain.

A: $\max_a \; \alpha(1 - p - \mu\bar{p} + a) + W - \frac{a^2}{2} - \frac{r}{2}\alpha^2\sigma^2$, foc: $a = \alpha$.

The principal chooses $p = \bar{p}$ and α to

P: $\max \; [1 - (1 + \mu)p + \alpha]p - \frac{\alpha^2}{2} - \frac{r}{2}\alpha^2\sigma^2.$

$$\text{foc}_p : 1 - 2(1 + \mu)p + \alpha = 0, \quad p^* = \frac{1 + \alpha}{2(1 + \mu)} \quad .$$

$$\text{foc}_\alpha : p - \alpha - r\alpha\sigma^2 = 0.$$

Substituting for p yields:

$$\alpha^* = \frac{1}{2(1 + r\sigma^2)(1 + \mu) - 1} \quad .$$

$$\frac{d\alpha^*}{d\mu} = -\frac{2(1 + r\sigma^2)}{\Gamma^2} < 0, \qquad \Gamma = 2(1 + r\sigma^2)(1 + \mu) - 1.$$

vii) Product Substitutability: $q = 1 - p + \delta\bar{p} + a + \varepsilon$, $\bar{p}$ = price at a rival chain.

A: $\max_a \; \alpha(1 - p + \delta\bar{p} + a) + W - \frac{a^2}{2} - \frac{r}{2}\alpha^2\sigma^2$, foc: $a = \alpha$.

P: $$\max_{p \text{ and } \alpha} \ (1 - p + \delta \bar{p} + \alpha)\, p - \frac{\alpha^2}{2} - \frac{r}{2}\alpha^2\sigma^2,$$

$$\text{foc}_p : 1 - 2p + \delta \bar{p} + \alpha = 0.$$

Using symmetry to set $p = \bar{p}$ yields:

$$p^* = \frac{(1 + \alpha)}{2 - \delta} .$$

$$\text{foc}_\alpha : p - \alpha - r\alpha\sigma^2 = 0.$$

Substituting for p yields:

$$\alpha^* = \frac{1}{(1 + r\sigma^2)(2 - \delta) - 1} .$$

$$\frac{d\alpha^*}{d\delta} = \frac{(1 + r\sigma^2)}{\Gamma^2} > 0, \qquad \Gamma = (1 + r\sigma^2)(2 - \delta) - 1.$$

vii) <u>Strategic Delegation of the Pricing Decision:</u>

$q = 1 - p + \delta\bar{p} + a + \varepsilon$, $\bar{p}$ = price at a rival chain.

In this case, the agent is compensated by residual claims after he pays a royalty ρ per unit to the franchisor, as well as a franchise fee F, so we have:

A: $$\max_{a \text{ and } p} \ (p - \rho)(1 - p + \delta\bar{p} + a) - F - \frac{a^2}{2} - \frac{r}{2}(p - \rho)^2\sigma^2$$

$$\text{foc}_a : a = p - \rho,$$

$$\text{foc}_p : 1 - 2p + \delta\bar{p} + a + \rho - r(p - \rho)\sigma^2 = 0.$$

Substituting for *a* yields:

$$p = \frac{1 + \delta\bar{p} + r\rho\sigma^2}{1 + r\sigma^2} \quad .$$

$$\frac{dp}{d\bar{p}} = \frac{\delta}{\Gamma} > 0, \quad \frac{dp}{d\delta} = \frac{\bar{p}}{\Gamma} > 0, \quad \frac{dp}{dr} = \frac{\sigma^2(\rho - 1 - \delta\bar{p})}{\Gamma^2} < 0,$$

$$\frac{dp}{d\sigma^2} = \frac{r(\rho - 1 - \delta\bar{p})}{\Gamma^2} < 0, \quad \frac{dp}{d\rho} = \frac{r\sigma^2}{\Gamma} > 0,$$

where $\Gamma = 1 + r\sigma.^2$ Using symmetry to set $p = \bar{p}$ yields:

$$p_D^* = \frac{1 + r\sigma^2\rho}{1 - \delta + r\sigma^2}$$

$$= \frac{1}{1 - \delta} \quad , \quad \text{when } r\sigma^2 = 0.$$

<u>Under Vertical Integration</u>: Assuming that $a = 0$ and $\rho = 0,$

P: $\max_p \, (1 - p + \delta\bar{p}) \quad p - F,$ foc: $1 - 2p + \delta\bar{p} = 0$

Setting $p = \bar{p}$ yields:

$$p_I^* = \frac{1}{2 - \delta}$$

$$p_I^* = \frac{1}{2 - \delta} < 1 < \frac{1}{1 - \delta} \quad .$$

Therefore $p_I^* < p_D^*$ when $r\sigma^2$ is small.

ix) <u>Multiple tasks</u>: $q = a + \varepsilon,$ $\varepsilon \sim N(0,\Sigma),$ $C(a) = a^Ta/2,$

where $q, a, \alpha,$ and ε are now vectors, but W is still a scalar.

A: $\max_{a} \quad \alpha^T a + W - \frac{a^T a}{2} \quad - \frac{r}{2} \quad \alpha^T \Sigma \alpha,$ foc: $a = \alpha$.

After substituting, we have:

P: $\max_{\alpha} \alpha^T j \quad - \frac{\alpha^T \alpha}{2} \quad - \frac{r}{2} \quad \alpha^T \Sigma \alpha,$ foc: $j - \alpha - r\Sigma\alpha = 0,$

$$\alpha^* = (I + r\Sigma)^{-1} j,$$

where j is a vector of ones. When $n = 2,$ this becomes

$$\alpha_i^* = \frac{1}{D}(1 + r\sigma_{jj} - r\,\sigma_{12}) \quad , \quad D = (1 + r\sigma_{11})(1 + r\sigma_{22}) - r^2\,\sigma_{12}{}^2.$$

Setting $\sigma_{11} = \sigma_{22} = \sigma^2$ yields:

$$\alpha_i^* = \frac{1 + r\sigma^2 - r\sigma_{12}}{(1 + r\sigma^2)^2 - r^2\sigma_{12^2}} = \frac{1}{1 + r(\sigma^2 + \sigma_{12})}$$

and

$$\frac{d\alpha_i^*}{d\sigma_{12}} = -\frac{r}{(1 + r(\sigma^2 + \sigma_{12}))^2} < 0.$$