RETAIL CONTRACTING: THEORY AND PRACTICE*

FRANCINE LAFONTAINE AND MARGARET E. SLADE

We summarize a number of regularities that arise in the empirical literature on contractual relationships between manufacturers and their exclusive resellers. We do this using studies of traditional and business-format-franchise relationships, as well as studies of sales-force-integration decisions. Some of the patterns that we uncover are consistent with a standard incentive-cum-insurance theory of organization, while others are not. We briefly review some theoretical extensions that seem promising in terms of reconciling seeming conflicts between theory and practice.

I. INTRODUCTION

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 usually choose some form of franchising.¹ Many theories attempt to explain the choice of retail-organizational form. Moreover, a considerable body of empirical evidence concerning this choice is accumulating. We attempt to evaluate this econometric evidence and relate it to the theories. To accomplish this task, we assess studies of franchising arrangements as well as studies of sales-force organization in industrial settings.

The paper is organized as follows. In the next section, we highlight a number of regularities that arise in the empirical literature on retail contracting. In particular, we find that:

(i) Risk is positively related to the use of high-powered incentives (vertical separation).
(ii) The importance of the agent’s effort in production is positively related to the use of high-powered incentives.

¹ Margaret Slade’s research was supported by a grant from the Social Sciences and Humanities Research Council of Canada. We would like to thank Erin Anderson, Joe Farrell, Bruce Lyons, Nancy Lutz, Scott Masten, two anonymous referees and participants in a session of the European Economic Association meeting, Prague, for helpful comments and suggestions.

¹ There are two forms of franchising: traditional and business-format. Traditional franchising, which involves an upstream producer and a downstream seller (e.g., gasoline), is the most common form. Business-format franchising, however, is the faster growing category. 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, typically takes place at the retail outlet (e.g., fast-food).
(iii) Larger units tend to be company-operated.
(iv) High costs of monitoring output or sales are positively related to company operation.
(v) High costs of monitoring effort directly are negatively related to company operation.
(vi) Retail prices are lower under company operation than under franchising.

Some of these regularities are consistent with the familiar agency model of retail contracting that is based on the need for franchisee insurance and incentives, but a number of them are not. Furthermore, certain regularities (e.g., points four and five above) appear to be mutually contradictory. After highlighting those areas where the data support the theory, as well as those where they do not, we discuss extensions to the agency model and other explanations that could account for the patterns that are revealed by the data.

II. EMPIRICAL EVIDENCE

Our discussion of the empirical evidence is organized around the factors that theory suggests should be important determinants of retail-organization form. Before turning to the evidence, we therefore give a brief overview of the principal relevant theories.

Most theories of the relationship between manufacturer and exclusive retailer or reseller are based on the idea that an unconstrained first-best outcome cannot be achieved due to agency costs. For example, the manufacturer or principal faces a tradeoff between providing the retailer or agent with insurance against risk and with incentives to work hard. Alternatively, incentive problems can arise on both sides of the principal/agent relationship, in which case the contract must trade off incentives for one party against incentives for the other. In general, the constrained-optimal form of organization balances the costs and benefits associated with production, transaction, and agency considerations.

Recently, theorists have also examined the implications for manufacturer/retailer contracts of imperfect competition in retail markets that leads to strategic behavior on the part of manufacturers. In an oligopoly context, it has been shown that principals can often benefit from delegating the authority to set the retail price to an agent who purchases the product at a wholesale price that is typically above marginal cost.

---

The relative agency costs of integration and separation are analyzed by Coase [1937], Williamson [1971], Alchian and Demsetz [1972], Klein, Crawford, and Alchian [1978], Rubin [1978], Fama and Jensen [1983], Mathewson and Winter [1985], and others.

Indeed, whereas vertical integration is frequently seen as a way of harmonizing the interests of up and downstream firms, vertical separation can harmonize the interests of competing manufacturers.4

The comparative statics from the standard incentive-cum-insurance model are fairly well understood. We refer to these in our review of empirical regularities below where relevant. What is perhaps less well known is the fact that the comparative statics from an agency model extended to include the possibility of strategic vertical separation are very similar to those from the original model.5 This is both good and bad. On the one hand, it strengthens the theoretical predictions, because they hold under a wider set of circumstances. On the other hand, it makes it difficult to distinguish between the two models empirically.

Much of the agency-theoretic literature on retail contracting focuses on explaining the size of the share parameter in a franchise contract and how this parameter should vary as other aspects of the problem change, where the share parameter determines the partition of residual claimancy between principal and agent.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 franchise 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 this dichotomous choice between company operation (vertical integration, which is associated with lower-powered incentives) and franchising (vertical separation, which is associated with higher-powered incentives) using arguments developed to explain how firms

4There is a related literature that examines the strategic choice of managerial incentives (e.g., Fershtman [1985], Fershtman and Judd [1987], Skilvas [1987], and Reitman [1993]) as well as one that assesses divisionalization (e.g., Baye, Crocker, and Ju [1996]), in which strategic motives lead to more rivalrous behavior.

5These results are derived formally in Lafontaine and Slade [1995].

6See for example Rubin [1978], Mathewson and Winter [1985], Lal [1990], and Bhattacharyya and Lafontaine [1995].

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 hired 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 [1993]).
should choose the terms of their contracts. In what follows, we focus on the findings from the literature that examines this dichotomy. We return briefly to the issue of standard contract terms within firms at the end of the section.

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 agents choose to integrate vertically (i.e., the 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 a number of narrowly defined retail sectors (e.g., Brickley and Dark [1987], John and Weitz [1988], Martin [1988], Norton [1988], Lafontaine [1992a], and Scott [1995]). 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 (e.g., Anderson and Schmittlein [1984], Barron and Umbeck [1984], Anderson [1985], Brickley and Dark [1987], Minkler [1990], Muris, Scheffman and Spiller [1992], Shepard [1993], Slade [1993 and 1996], Graddy [1995], and Lafontaine [1995]). The two sets of studies also differ in that the first involves business-format franchising, whereas the second includes some traditional-franchising industries in which the principal is a manufacturer.

Table I summarizes the findings of studies that assess the integration/separation choice. In this table, the signs in the final columns show the observed effect of a variable of interest on the tendency towards vertical integration. A minus sign thus indicates a negative correlation with the use of company operation in a franchised chain or of “direct sales” in the sales-force-integration problem. Moreover, an asterisk next to a plus or minus sign indicates that the finding in the original paper is statistically significant at the 0.05 level based on a two-tailed test. In what follows, each portion of the table is discussed in turn.

II(i): The Effect of Risk

The traditional agency model of retail contracting suggests that increases in uncertainty, or more precisely in demand variability, increase the need for agent insurance and thus the desirability of vertical integration. Inside the firm, the agent is given lower-powered incentives and is thus more fully

---

8 Lafontaine [1992a and 1993] and Sen [1993] have examined instead, or also, the determinants of the financial terms of business-format franchise contracts, namely the royalty-rate and franchise-fee combination. Their results have been generally consistent with the conclusions one can draw from examining the franchise versus own decision.

9 This assumes that the principal, here the upstream firm, is risk neutral or at least less risk averse than the agent, or outlet manager.
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. Available data are often at the level of the sector rather than at the franchisor or upstream-firm level.

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

II(ii): The Importance of the Agent’s Effort

The degree to which an increase in agent effort affects output or sales is also an important parameter in traditional agency models and should therefore affect the upstream firm’s decision to use high or low-powered incentives. Specifically, the theory predicts that increases in the importance of the retailer’s input should be associated with more separation and higher-powered contracts. From a practical point of view, the measures that are 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, in this case employees/sales or capital/labor ratios, as the agent is the one who must oversee the provision of labor. Researchers have also used a measure of the agent’s value added 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 I(ii) summarizes the results from five studies that assess agent importance. In every case where agent importance is statistically insured.

---

10 We take the empirical regularity that company employees tend to be given lower-powered incentives as a given and do not attempt to explain its causes. For a possible explanation, see Holmstrom and Milgrom [1994].

11 For a discussion of the relative merits of these two measures, see Lafontaine and Bhattacharyya [1995].

12 Usually the dependent variable and measures of risk are at the same level of aggregation. The exception is Lafontaine [1992a] who uses firm-level franchising data coupled with sector-level risk data.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Data</th>
<th>Measure</th>
<th>% Co Owned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson and Schmittlein</td>
<td>1984</td>
<td>Electronics Components by Product Line and Territory</td>
<td>% Forecast Error of Product-Line Sales by Territory</td>
<td>–</td>
</tr>
<tr>
<td>John and Weitz</td>
<td>1988</td>
<td>Industrial Firms with Sales above $50 million</td>
<td>Index Capturing Environmental Uncertainty</td>
<td>+</td>
</tr>
<tr>
<td>Martin</td>
<td>1988</td>
<td>Sectoral Panel</td>
<td>Coefficient of Variation of Detrended Sector Sales</td>
<td>–*</td>
</tr>
<tr>
<td>Norton</td>
<td>1988</td>
<td>Restaurants and Motels by State and Sector</td>
<td>Variance of Detrended % Change in Sector Sales by State</td>
<td>–*</td>
</tr>
<tr>
<td>Lafontaine</td>
<td>1992</td>
<td>Bus. Format Franchising Firms from All Sectors</td>
<td>Fraction of Outlets Discontinued in Sector</td>
<td>–*</td>
</tr>
</tbody>
</table>

Note: * indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.
significant, its relationship with company operation is negative, 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, separation is more likely.

II(iii): The Effect of Outlet Size

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, when it is treated as a separate parameter affecting contract choice, a standard agency model predicts that an increase in size should also be associated with separation and higher-powered incentives.

Unlike the factors discussed above, the measurement of size is fairly straightforward. Common measures are average sales per outlet and the initial investment required. Table I(iii) shows that, with one exception, greater size leads to increased company ownership. In other words, contrary to what a standard agency model predicts, people responsible for large outlets tend to be company employees who receive low-powered incentives.13

II(iv): The Cost of Monitoring

The traditional agency model is based on the assumption that the principal cannot observe the agent’s effort directly. Empirical researchers have taken this to mean that the difficulty the principal faces in monitoring the agent will affect the extent of vertical integration, and they have tested this proposition.

Finding appropriate proxies for monitoring costs tends to be difficult. For two studies of the electronic-components industry, as well as a study of distribution channels for industrial-goods producers, which are labeled Group I in Table I(iv), 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,” and in Anderson [1985] the measure was tabulated from responses to i) “team sales are common,” ii) “sales and cost records tend to be inaccurate at the individual level,” and iii) “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

---

13 Consistent with the above evidence, Muris, Schellman and Spiller [1992] also 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.
### I(ii): The Importance of the Agent’s Effort

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Data</th>
<th>Measure</th>
<th>% Co Owned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norton</td>
<td>1988</td>
<td>Restaurants and Motels by State and Sector</td>
<td>Employees/Sales</td>
<td>−*</td>
</tr>
<tr>
<td>Lafontaine</td>
<td>1992</td>
<td>Bus. Format Franchising Firms from All Sectors</td>
<td>(Sales—Franchisor Inputs)/Sales for Sector Previous Experience Required</td>
<td>−</td>
</tr>
<tr>
<td>Shepard</td>
<td>1993</td>
<td>Gasoline Service Stations in Massachusetts</td>
<td>Full Service</td>
<td>−</td>
</tr>
<tr>
<td>Scott</td>
<td>1995</td>
<td>Bus. Format Franchising Firms from All Sectors</td>
<td>Capital/labor Ratio</td>
<td>(+*)</td>
</tr>
<tr>
<td>Slade</td>
<td>1996</td>
<td>Gasoline Service Stations in Vancouver</td>
<td>Full Service</td>
<td>−*</td>
</tr>
</tbody>
</table>

Note: 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.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Data</th>
<th>Measure</th>
<th>% Co Owned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brickley and Dark</td>
<td>1987</td>
<td>Selected Franchising Firms</td>
<td>Initial Investment</td>
<td>+*</td>
</tr>
<tr>
<td>Norton</td>
<td>1988</td>
<td>Restaurants and Motels by State and Sector</td>
<td>Sales/Outlet</td>
<td>-*</td>
</tr>
<tr>
<td>Martin</td>
<td>1988</td>
<td>Sectoral Panel</td>
<td>Sales/Outlet</td>
<td>+*</td>
</tr>
<tr>
<td>Lafontaine</td>
<td>1992</td>
<td>Bus. Format Franchising Firms from All Sectors</td>
<td>Initial Investment</td>
<td>+*</td>
</tr>
<tr>
<td>Lafontaine</td>
<td>1995</td>
<td>Fast-Food in Pittsburgh and Detroit Metropolitan Areas</td>
<td>Sales/Outlet for Sector</td>
<td>+*</td>
</tr>
<tr>
<td>Scott</td>
<td>1995</td>
<td>Bus. Format Franchising Firms from All Sectors</td>
<td>Initial Investment</td>
<td>+</td>
</tr>
</tbody>
</table>

Note: * indicates a result that is significant in the original study at the 0.05 level based on a two-tailed test.
long lag between actions and market responses makes it difficult to attribute output to effort. Using scores thus obtained as measures of monitoring costs, these researchers found that higher costs were associated with more vertical integration.

Another frequently used measure of monitoring difficulty is some notion of geographic dispersion (captured in one case by whether the retail unit is more likely to be in a rural area) or of distance from a monitoring headquarters. In addition, outlet density has been used as an inverse measure of monitoring cost. The studies in Table I(iv) that are labeled Group 2 show that, when monitoring costs are measured either directly by dispersion or distance or inversely by density, in all cases where coefficients are significant, higher monitoring costs have a negative effect on the fraction of company-operated outlets. Thus, higher costs lead to less, not more, integration.

A principal/agent model in which the principal receives two noisy signals of an agent’s effort sheds light on the discrepancy between these two sets of findings. Consider the possibility that the principal can use not only outcome information (i.e., sales or Type 1) to infer something about the agent’s effort, but also a direct signal of the agent’s behavior, which we call Type 2 information. In other words, direct supervision or monitoring of the agent’s behavior provides the manufacturer with an expanded signal of retailer effort that supplements the information contained in the sales data.14 Furthermore, the principal is allowed to base the agent’s compensation on both signals. The informativeness principle (Holmstrom [1979], Milgrom and Roberts, [1992, p. 219]) then suggests that compensation will be based on both types of information. Moreover, the higher the quality/cost ratio of effort information obtained from sales data (Type 1 information) relative to that arising via direct supervision (Type 2), the more the compensation package will emphasize sales-based compensation at the expense of behavior-based compensation. Thus vertical separation, with its focus on sales-based compensation, will be more attractive when Type 1 information is relatively less costly to obtain. The opposite is true of Type 2 information.

Now consider in more detail what the authors of the studies in Table I(iv) are measuring. Anderson and Schmittlein, Anderson, and John and Weitz obtained information on the difficulty of relating sales (Type 1) results to the performance or effort of individual salespeople. The higher their measures of monitoring cost, therefore, the poorer are sales data as a measure of effort. Consequently, companies tend to rely less heavily on

---

14 The type of mechanism that we have in mind is sometimes called “behavior-based” control, as opposed to “outcome-based” control (e.g., Anderson and Oliver [1987]). See Lafontaine and Slade [1996] for a model emphasizing this distinction between these two types of monitoring.
this noisy signal of effort and to provide the agent with more insurance. In other words, they reduce their reliance on output-based compensation; thus the increased degree of vertical integration.\footnote{A different but consistent way to interpret this result is that, in a transaction-cost context, high uncertainty in a transaction leads to more vertical integration. This is the interpretation given by the authors.}

With measures of dispersion or distance, in contrast, researchers are aiming to capture the cost of monitoring the retailer’s effort directly (obtaining Type 2 information). This is accomplished by, for example, sending in a company representative or a pretend customer in order to obtain additional information (e.g., on cleanliness, product quality, and friendly service) that complements sales data. Moreover, in the types of businesses that are franchised, final products are apt to be relatively simple and team sales, uncommon. Hence retail sales figures tend to be good measures of a unit manager’s or a franchisee’s effort. In this context, high monitoring cost means that non-sales data on behavior is costly to obtain, while sales data are a good signal of effort. This leads the upstream firm to rely more heavily on sales in its compensation scheme, or rely more often on vertical separation.

It should be clear that the different measures that are used in the empirical literature capture different types of monitoring costs. When one takes this difference into account, the “contradictory” results obtained by these researchers are in fact consistent with each other as well as with standard downstream-incentives arguments for franchise contracting.

**II(v): Product Substitutability**

The standard incentive-cum-insurance model of retail contracting does not usually consider the competitive environment in which the principal/agent relationship operates. Not surprisingly then, most empirical studies of contract choice also rely solely on attributes of the upstream firm and its outlets and ignore the firm’s competitors. It can be shown, however, that when one introduces the potential for substitution among competing firms’ products into an agency model, one finds that the benefits associated with separation increase with the degree of substitutability across products. This is true in either a strategic or a more standard agency setting.

In a standard agency setting, 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 more important it is for the agent to promote his product in order to prevent the erosion of its sales. As in subsection II(ii), therefore, the principal has an additional motive for emphasizing high-powered incentives relative to other objectives.

To understand the strategic setting, it is helpful to consider a two-stage

### I(iv): Monitoring Difficulty

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Data</th>
<th>Measure</th>
<th>% Co Owned</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anderson and Schmittlein</td>
<td>1984</td>
<td>Electronics Components by Product Line and Territory</td>
<td>Index indicating that it is difficult to measure results of individuals.</td>
<td>+*</td>
</tr>
<tr>
<td>Anderson</td>
<td>1985</td>
<td>Electronics Components by Product Line and Territory</td>
<td>Index indicating that 1) team sales are common, 2) records are inaccurate and 3) sales and cost figures are insufficient for a fair evaluation.</td>
<td>+*</td>
</tr>
<tr>
<td>John and Weitz</td>
<td>1988</td>
<td>Industrial Firms with Sales above $50 million</td>
<td>Importance of Non-selling Activities</td>
<td>+*</td>
</tr>
<tr>
<td><strong>Group 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brickley and Dark</td>
<td>1987</td>
<td>Selected Franchising Firms</td>
<td>Distance From Monitoring Headquarters</td>
<td>-*</td>
</tr>
<tr>
<td>Norton</td>
<td>1988</td>
<td>Restaurants and Motels by State and Sector</td>
<td>Fraction of State Population Rural</td>
<td>-*</td>
</tr>
<tr>
<td>Minkler</td>
<td>1990</td>
<td>Taco Bell Restaurants in Northern California and Western Nevada</td>
<td>Distance From Monitoring Headquarters</td>
<td>-*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Outlet Density = Number of Outlets Within a 5 Mile Radius</td>
<td>(-)</td>
</tr>
<tr>
<td>Lafontaine</td>
<td>1992</td>
<td>Bus. Format Franchising Firms from All Sectors</td>
<td>Number of States in which the chain has Established Outlets</td>
<td>-*</td>
</tr>
<tr>
<td>Lafontaine</td>
<td>1995</td>
<td>Fast Food in Pittsburgh and Detroit Metropolitan Areas</td>
<td>Outlet Density = Number of Outlets from the Same Chain in Same Zip Code</td>
<td>(+*)</td>
</tr>
<tr>
<td>Scott</td>
<td>1995</td>
<td>Bus. Format Franchising Firms from All Sectors</td>
<td>Number of States in which the chain has Established Outlets</td>
<td>-*</td>
</tr>
</tbody>
</table>

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.
game—the separation game—and contrast it with a static game—the integration game. With the separation game, retail contracts are offered in the first period, and price competition occurs in the second. Since manufacturers anticipate the outcome of the retail game, one can think of the upstream firms as choosing retail prices. If managers of upstream firms maximize profits conditional on rival behavior, it is optimal for them to set price/cost margins equal to minus one divided by their perceived elasticities of demand. Their perceived elasticities, however, depend on the type of game that they are playing. In the separation game, the perceived elasticity has two components: a direct and a strategic one. The direct component is the own-price elasticity, holding rival price constant, whereas the strategic component is the cross-price elasticity (the percentage change in own sales due to a one-percent change in rival price) times the elasticity of the rival’s reaction function (the percentage change in rival price due to a one percent change in own price). When the two products are substitutes, these two components have opposite signs, their sum is smaller in magnitude than the own-price elasticity, and equilibrium markups are greater than they would be if there were only a direct component. Furthermore, all else equal, as cross-price elasticities increase, so do equilibrium markups.

With the integration game, in contrast, principals or upstream firms choose prices directly; they do not delegate this task to agents. Since there is no fixed fee or wholesale price to choose, the game has only one period, and, as a consequence, the players’ perceived elasticities have only direct components. In other words, since prices are chosen simultaneously in the first (and only) period, players cannot condition their price choices on the terms of own and rival contracts with retailers. They therefore set price/cost margins equal to minus one divided by the own price elasticity of demand, which is evaluated holding rival price constant. Equilibrium markups are then lower than in the separation game.

To our knowledge, Slade [1993] is the only study that looks at contract choice as a function of the demand characteristics that agents face. She relates the contracts under which outlets operate to outlet-level own and cross-price elasticities of demand and to the elasticities of rival-reaction functions. She finds, as predicted, that delegation is more likely when cross-price elasticities of demand are large relative to own-price elasticities and when rival-reaction functions are steep.

**II(vi): Franchisee Free Riding**

Once an agent is given high-powered incentives via a franchise contract,

---

16 For an analysis of the role of direct and strategic effects in two-stage games, see Tirole [1988, pp. 324–336].
another kind of incentive problem can arise—the franchisee can free ride on the tradename (see e.g., Klein, [1980] and Brickley and Dark, [1987]). As these authors note, the free-riding problem is more likely to occur in situations where consumers do not impose sufficient discipline on retailers, namely cases of non-repeat businesses. Brickley and Dark [1987] find that franchisors rely less on franchising in the three industries that they consider to be most susceptible to free riding: hotels and motels, restaurants, and car-rental agencies. However, although the free-riding theory also implies that units of franchised chains located along freeways will tend to be company operated, they find the opposite tendency in their data.

The evidence on free riding is thus mixed (see also Brickley, Dark and Weisbach, [1999a], and Minkler [1990] on this). One explanation for the lack of strong evidence in favor of free riding is that franchisors find ways to control franchisee behavior by, for example, using approved supplier requirements or self-enforcing contracts. The franchisor, unlike the franchisee, internalizes the spillovers that damage the trademark when free riding occurs in transient-customer locations such as freeway exits. If this is so, however, it implies that 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 leads to the issue of franchisor incentives in a double-sided moral hazard model of franchise contracting.

II(vii): 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 of the 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 for the goods that are sold through separated and integrated units.17

There are a number of reasons why prices might be higher at separated outlets. First, some transactions are more costly in a market than inside a firm. For example, the types of contracts written with franchisees are often more complex and thus costlier to write than those written with employees. Second, because separation involves two firms rather than one, it may introduce an additional administrative layer. Third, when retailers have market power, double marginalization (i.e., successive output restrictions) can arise. Finally, in a strategic model of contracting, separation lowers retailers’ perceived elasticities of demand and thus increases retail markups (see Rey and Stiglitz, [1995]).

17 Barron and Umbeck [1984] also consider the effect of different contractual arrangements on hours of operation.
Table II summarizes results from seven studies that are relevant to this issue. Three deal with retail prices of gasoline in the US, another deals with prices charged by retailers of separated and integrated soft-drink bottlers, still another involves beer sold in public houses in the UK, and the last two are concerned with fast-food franchising in certain US submarkets.

Barron and Umbeck [1984] and Slade [1995] 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. 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.

II(viii): Within-Firm Contract Uniformity

Most theoretical contracting models allow the principal to tailor the terms of the contract to suit the characteristics of the agent, the outlet, and the market. Contracts that are observed in practice, in contrast, are remarkably insensitive to variations in individual, outlet, and market conditions. Indeed, many firms use a standard business-format-franchising

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Data</th>
<th>Price Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barron and Umbeck</td>
<td>1984</td>
<td>Gasoline Service Stations in Maryland</td>
<td>+*</td>
</tr>
<tr>
<td>Muris, Scheffman and</td>
<td>1992</td>
<td>Prices of Retailers Served by</td>
<td>+*</td>
</tr>
<tr>
<td>Spiller</td>
<td></td>
<td>Integrated or Separated Soft-Drink Bottlers</td>
<td></td>
</tr>
<tr>
<td>Shepard</td>
<td>1993</td>
<td>Gasoline Service Stations in Mass.</td>
<td>+*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(and significant for one product)</td>
<td></td>
</tr>
<tr>
<td>Slade</td>
<td>1995</td>
<td>Beer in the UK</td>
<td>+*</td>
</tr>
<tr>
<td>Lafontaine</td>
<td>1995</td>
<td>Fast-Food in Pittsburgh and Detroit</td>
<td>+*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metropolitan Areas</td>
<td></td>
</tr>
<tr>
<td>Graddy</td>
<td>1995</td>
<td>Selected Fast-Food Chains in</td>
<td>+*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New Jersey and Western Pennsylvania</td>
<td></td>
</tr>
</tbody>
</table>

Note: * indicates a result that is significant in the original study at the 0.05 level, based on a two-tailed test.
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.\(^{18}\) 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 operate those with characteristics that require less high-powered incentives, and to franchise those that require more. This again explains the focus in empirical work on the choice between integration and separation rather than on the actual terms of the contract.

Interestingly, in the few empirical studies of the effects of the variables described above on the terms of the uniform franchise contract (the share parameter, or royalty rate, and the up-front fixed payment, or franchise fee), authors have found that variables that should affect the share and fixed-fee parameters are better at explaining the dichotomous decision to franchise or not.\(^{19}\) Thus it appears that firms, in responding to risk, incentive and monitoring-cost issues, adjust by changing how much they use franchising more than by altering the terms of their uniform franchise contract. In that sense, the theoretical models seem to be missing some important aspects of the upstream/downstream relationship.

It has also been found that franchise fees do not extract all downstream profits given the royalty rate, as the standard principal-agent model suggests. Instead, fixed fees tend to be set at levels that compensate the franchisor for expenses incurred in setting up a franchise.\(^{20}\) Finally, once set, franchise contract terms have been found to be quite persistent over time.\(^{21}\)

\(II(ix):\) Asset Specificity

Asset specificity is an important area of the theoretical literature that we have, up to now, had little to say about. This is because we believe that it is far less important for retail contracting than for the purchase and sale of

\(^{18}\) In the US, the Robinson-Patman Act requires wholesale-price uniformity, at least locally. This is not true, however, in Canada. Nevertheless, uniformity across buyers is common there as well (e.g., in gasoline markets; see Slade [1993 and 1996] 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 generally as the main source of contract uniformity in business-format franchising.

\(^{19}\) See, for example, Lafontaine [1992a].

\(^{20}\) See Lafontaine [1992a] and Does [1993].

\(^{21}\) See Lafontaine and Shaw (1996). The authors suggest that this is due in part to adjustment costs, and in part to persistent firm effects.
intermediate inputs. As a result, we don’t think it sheds much new light on the empirical regularities highlighted herein. Nevertheless, as this issue occasionally surfaces in the empirical literature, we discuss how we arrive 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 level of investment. This investment, however, is again not specific; hotel banners are routinely changed with little effect on property value.

It has also been argued that franchisors own units in cities, or at least own the right to leases for city-unit locations, because these locations are highly specific assets (Dnes, [1993]). We believe that the difficulty of obtaining “good locations” in city centers is simply a reflection of the economic value of those locations, where the economic value is understood by all potential tenants and thus is clearly not specific. Hence we again fail to see a direct relationship between high asset value and high asset specificity.

III. EXPLAINING THE DISCREPANCIES

Our analysis has uncovered fairly robust patterns in the empirical evidence concerning retail contracting. While some of these are consistent with particular aspects of standard contracting models, in particular with the emphasis on incentive and monitoring issues, it is also clear that the data are inconsistent with other aspects of these models. In what follows, we discuss extensions to the basic theoretical framework that show promise in addressing the thus-far unexplained regularities.

22 Examples of empirical papers that deal with relationship-specific assets in intermediate input markets include Monteverde and Teece [1982], Masten [1984], Goldberg and Erickson [1987], Joskow [1988], Crocker and Masten [1991], and Crocker and Reynolds [1993].

23 See also Klein (1995) for an excellent discussion of how self-enforcement in retail contracting does not rely on the amount of specific assets invested by the parties, but rather how franchisee or agent opportunism is controlled by the amount of rents the agent expects to receive from the continuation of the relationship, and how franchisor opportunism is controlled by the costs of operating the whole chain directly.
III(i): Risk

The first problem noted above is the inconsistency of the evidence related to risk with the insurance argument embedded in the traditional agency model. The empirical evidence reveals a positive effect of risk, or at least of sales variability, on franchising. Some authors have concluded from this that franchisors shed most risk onto franchisees (e.g., Martin [1988]). This could only be optimal, however, if franchisors were more risk averse than franchisees, in which case there would be less need to trade off franchisee incentive and insurance needs, and hence less need to use a share contract. 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 risk/franchising phenomenon surfaces when one considers that the power of incentives can influence sales variability. Indeed, franchisees often have superior information concerning local-market conditions. Moreover, since franchising gives retailers greater incentives to react to these conditions, one might well find more sales variability in franchised than in company-owned units, even when exogenous risk is the same across the two groups.24 In that sense, the positive relationship between risk and franchising may best be understood as further support for incentive-based arguments for franchising.

III(ii): Outlet Size

The second inconsistency between theory and practice is the fact that outlet size is positively correlated with integration rather than separation, which is difficult to reconcile with standard agency models of retail contracting. The positive relationship between size and company ownership, however, interacts with the urban-location phenomenon that is used to measure monitoring costs in a robust way. Indeed, large-volume outlets located in urban areas tend to be company operated, whereas small-volume rural outlets are much more apt to be run by entrepreneurs.25 Indeed, franchisors themselves explain the larger size of company-owned units by stating that it reflects “a higher concentration of company-owned outlets in major urban centers, and with it a higher investment cost per outlet” (IFA Educational Foundation and Horwath

24 See Lafontaine and Bhattacharyya [1995] for more on this. Also, although a positive relationship between incentives and output variability is likely in many cases, whether it occurs depends in general on the form of the function that maps local information, effort and the random shock into output.

25 The same relationships that are shown in the tables, which are mostly for business-format franchising, are also found in traditional franchising such as gasoline retailing and in public houses that can be operated by a company employee or by a tenant.
International, 1990, at 1). While lower direct-supervision costs suggest that urban units should more often be company operated, their larger size implies the opposite tendency. We suggest two further explanations for why firms might opt for integration for large urban units despite the size effect.

First, there are economies of scale in promotion that can be captured at the local level. For example, when a company has many outlets in the same city, the regional manager, as opposed to the individual-unit manager, can occupy herself with city-wide promotional activities. This in turn reduces the importance of the unit manager’s inputs, making company ownership more attractive to the firm. In addition, the coordination of promotional activities city wide is made simpler by company ownership. When outlets are isolated, in contrast, comparable local-promotional efforts must be undertaken separately by each agent and need not be coordinated to the same extent. This makes the entrepreneurial role of the individual-unit manager extremely important to the success of the business, leading firms to rely on high-powered incentives in these units.

Our second explanation is based on the observation that, when franchise contracts are not fully adjusted to outlet differences, franchisors are apt to find that relatively more rent is left downstream to large franchised outlets than to small. When large outlets are company operated, in contrast, upstream firms are more likely to be able to capture these rents.

26 The importance of outlet density for franchisors’ decision to own urban units is supported by the following statement from the 1973 10-K report of the McDonald’s corporation: “The 860 restaurants owned by the company on December 31, 1973, are generally concentrated geographically because of economics and managerial efficiencies made possible by their proximity to each other.”

27 See Kaufmann and Lafontaine [1994] for evidence that there are both ex-ante and ex-post rents left downstream at McDonald’s, and that these are larger in absolute and relative terms in larger-volume units. See also Michael and Moore [1995] on the issue of downstream rents. Combined with recent evidence from Williams [1996], which shows that franchisors on average terminate poorly performing units, these results suggest that franchisors create a premium stream to align franchisee incentives, and that they resort to termination when franchisees do not perform. See Klein [1995] on various ways in which franchisor and franchisee incentives might not coincide. See also Brickley, Dark and Weisbach [1991b] for evidence that franchisors in nonrepeat-business industries franchise less when they face restrictions on termination of franchisees.

28 A third potential explanation relates to the idea of outlet diversity and task variety. When a company has many outlets in a city, each may be more specialized. For example, individual real-estate agencies might specialize in apartment rentals, house sales, and so forth. In a small town, in contrast a single agency is more apt to have to handle business of all kinds.

It can be shown that, when an agent performs many tasks that are not highly complementary, the optimal-payment scheme involves higher-powered incentives (Holmstrom and Milgrom [1991] and Slade [1986]). However, we do not know of any empirical study that has contrasted the degree of specialization of individual units in urban and rural areas, a necessary first step in determining the validity of this explanation.
III(iii): Within-Firm Contract Uniformity

The last way in which the data appear to be inconsistent with the theory is the tendency for contracts used by firms to be much more standardized across retailers than most theoretical models would suggest. Models that emphasize incentive issues for both parties—double-sided moral-hazard models—however, 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, but also the principal must maintain the value of the trademark or company logo. Since maintaining the reputation of the chain usually involves costly activities such as agent monitoring and product promotion that are not easily observed or assessed by retailers, the principal can also shirk. With moral hazard on the part of both parties, even when both are risk neutral, an optimal contract involves revenue sharing (Rubin [1978], Mathewson and Winter [1985], Lal [1990], and Bhattacharyya and Lafontaine [1995]).

In 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 explain some of the persistence of contract terms over time. Indeed, in their model, the terms of the optimal contract remain unchanged as the franchise chain grows.

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

IV. SUMMARY AND CONCLUSIONS

Table III summarizes the findings from our survey of the theory and practice of retail contracting under exclusive marks. It lists predictions from the standard agency model, the empirical findings, and possible
reconciliations when the two do not agree. As in Table I, the signs in this table indicate a factor’s effect on the fraction of outlets (or of the sales force) that is vertically integrated by the company, or on the propensity of an individual outlet (or sales person) to be vertically integrated.

As we have noted in our discussion of the individual tables, this summary table reveals a number of surprising facts. Perhaps the most unexpected, however, is that the empirical evidence is fairly consistent across industries and firms. Indeed, the regularities noted in the introduction are robust. We contrast this robustness with the predictions of the theoretical models, which are much more fragile.

It is well known in Industrial Organization that, when one models imperfect information and strategic interactions, virtually any behavior can be rationalized. In particular, any outcome can result from the optimizing choices of rational agents under some set of assumptions. The problem is therefore not to construct models that predict arbitrary outcomes, but rather to build robust models that encompass the stylized facts. We are encouraged to discover that, at least in the retail-contracting area, these facts are themselves robust in that they do not depend on peculiarities of the industries that are assessed in each study.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Prediction</th>
<th>Finding</th>
<th>Reconciliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk and Uncertainty</td>
<td>+</td>
<td>-</td>
<td>Local Private Information</td>
</tr>
<tr>
<td>Monitoring Difficulty</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 1</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Type 2</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Agent Importance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outlet Size</td>
<td>-</td>
<td>+</td>
<td>Entrepreneurial EOS, Need to Extract Rent</td>
</tr>
<tr>
<td>Product Substitutability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-Firm Contract Uniformity</td>
<td>Low</td>
<td>High</td>
<td>Double-sided Moral Hazard, Costs of Customization</td>
</tr>
<tr>
<td>Prices at Delegated Outlets</td>
<td>Higher</td>
<td>Higher</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


Gal-Or, E., 1992, ‘Franchising or Ownership of Retail Outlets as Implied by Competitive Forces’, mimeo, University of Pittsburgh Business School.


SLADE, M. E., 1995, ‘Beer and the Tie: Did Divestiture Cause Higher Prices?’, mimeo, University of British Columbia Department of Economics, Canada.


