R. Betancourt: (Draft, April/May 2003/8/03)

Ch. 9: Franchises

Franchises are organizational forms that lie in between vertically integrated firms and arm's length market exchange. Firms exist to internalize transaction costs, according to Coase (1936) and Williamson (1985), and they satisfy a need for long run contracts to facilitate exchange. Hence, at this most elementary level one can view franchises as an alternative to both vertical integration and arm's length market exchange due to the existence of substantial transaction costs that are difficult to internalize and the need for long run contracts to facilitate exchange.

As an organizational form franchising encompasses two very different types of franchise systems: product trade name or authorized franchise systems and business format franchises. In the former a physical good is always one of the basic products transacted at all levels of the channel structure and the franchisees are primarily a distribution system. In the latter a process or way of doing things is always one of the basic products transacted between the franchiser and the franchisee and a service is often one of the basic products transacted between the franchisee and the consumer. While the franchisees in this case also provide a distribution system, they often have to do much more than distribute because production and distribution are frequently inseparable in business format franchising. Finally, product trade name franchising can be traced back to the late nineteenth century whereas business format franchising becomes a significant activity after World War II. While the former type still dominates the latter type in terms of actual economic size, business format franchising is the more dynamic and fastest

growing type, Khan (1999, Ch.1).

In the next section, 9.1, some quantitative information on franchising is provided. This type of information is very difficult to obtain on a comprehensive and consistent basis because government agencies don't collect systematic information on this organizational form. In Section 9.2 the main economic characteristics of one of the two most important product trade name retail franchises, gas stations, are identified and discussed in detail. In particular, I stress their relation to the provision of distribution services by the franchiser directly to consumers. Similarly, in Section 9.3 the main economic characteristics of the other main type of product trade name retail franchises, new automobile dealers, are identified and analyzed, together with their relation to the provision of distribution services by the franchisees to consumers. Finally, in Section 9.4 the main economic characteristics of business format franchising are also identified and analyzed together with their relation to the provision of distribution services by both the franchiser and the franchisee. In general our aim is to bring out the role of distribution services in determining important features of franchise arrangements. In identifying the latter we rely heavily on the existing literature.

9.1. Quantitative Information on Franchises.

According to Stern and El-Ansary (1992, pps.343-344) sales of goods and services by franchising companies totaled about $680 billion in 1989. Most of these sales are retail sales.[1] They constituted about 33% of all retail sales in 1989, according to

[1] These sales figures include, for example, sales to soft drink bottlers (3.6% of the total),who are wholesalers. They also include some categories that are difficult to classify such as business aids and services or maintenance and cleaning. In any event, over 90% of the sales of these franchising companies are clearly retail sales.

the same source. The following figures, which are taken from Table 7.1 in the above source, indicate the principal retail sectors in which these organizational forms have taken hold. These five sectors represent about 80 % of outlets and sales of franchising companies and a greater proportion of outlets and sales of retail franchises. Furthermore, product trade name franchises, i.e., the first two categories in Table 9.1, represent at least 68 % of sales. Unfortunately, the latest edition of this text doesn't update these figures.

Table 9.1: Distribution of Sales and Outlets of Franchise Companies, 1989

SECTOR	% of ESTABLISHMENTS	% OF SALES
Automobile and Truck Dealers.	5.5	52.
Gasoline Service Stations.	22.5	16.
Restaurants (all types)	18.9	10.2
Convenience grocery stores	3.5	2.1
Hotels, Motels, Camps	2.0	3.1

Data on retail franchises are next to impossible to obtain from standard official government statistics for two reasons. First and foremost, the Census typically does not gather data on the basis of this organizational form. Second, while franchised establishments dominate sectors such as the above five they are not, in general, the only form of operation in these sectors. Hence, looking at the behavior of the aggregate of the sectors can be misleading with respect to what is going on in the franchised part.

Perhaps the only exception to the first reason is the category automobile and truck dealers. The U.S. Statistical Abstract publishes one table every year on Franchised New Car Dealerships. For instance, the 2000 issue (Table 1278) shows that the number of

dealerships declined steadily between 1980, when there were 27, 900 of them, and 1999, when there were 22, 400 of them. Interestingly, the number of new cars sold went from 8.98 millions to 8.70 millions during the same period. Even in this case, however, the source of most of the data in Table 1278 is the National Automobile Dealers Association.

Despite these difficulties, it is useful to report some data on two of the above categories published by the Census. Convenience grocery stores in the Census are part of food stores, but they consist of two types: convenience food stores and convenience food stores with gas stations. Gasoline sales must be less than 49% for the latter category or it would be classified as a gas station. Similarly, gasoline stations in the Census are of two types: gas stations with convenience store or gas stations/other, usually with a repair shop. Why does this matter? The entities in these different classifications are evolving in very different fashions.

In Table 9.2 we show what has happened to the number of outlets in these four different categories between 1987 and 1997.[2] Similar trends affect sales. While convenience stores have grown during this period, the growth is entirely due to the growth in the subcategory that has expanded the assortment to include gasoline sales.

Table 9.2. Outlets of Gas Stations and Convenience Stores: 1987,1997

SECTOR	OUTLETS 1987	OUTLETS 1997
Gasoline/Convenience Store	15,625	53,641
Gasoline/Other	99,123	45,205

[2] Due to the change in classification from SIC to NAICS, 1997 is the last year for which this comparison can be made based on the publicly available data.

Convenience Stores	30,864	27,081
Convenience Food/Gasoline	18,650	28,043

Along similar lines, the decline in the number of outlets classified as gasoline stations obscures two dramatically opposite trends. Namely, a rapid decline in what used to be the standard format, i.e., with a repair shop associated with the station, and an even more rapid increase in a new format, i.e., with a convenience store associated with the station.[3]

Convenience stores are normally classified as business format franchises and gasoline stations are normally classified as product trade name or authorized dealership franchises. Hence, these trends suggest the possibility of direct competition between the two types of franchises in the current economy.[4] These trends also cast some doubt on Oi's (1992) interpretation of the evolution of the gasoline station industry as one in which assortments have narrowed. While the repair function has been declining, as he demonstrates and the above data confirms, the addition of a convenience store represents a move to broader assortments.

Last but not least is the information from a recent study of franchising sponsored by the International Franchise Association (IFA), namely The Profile of Franchising: A Statistical Abstract of 1998 UFOC Data. It was released in 2000 and is available on IFA's website. It is not comprehensive, since it is limited to franchise systems that are offering franchises in a particular year in one of 12 states requiring annual registration.

---

[3] Incidentally, these stores usually sell a narrower line of convenience products than the traditional convenience store.

[4] Furthermore these trends have given rise to a new term 'dual concept franchising'[Khan,(1999), Ch.1], which describes arrangements where two different franchise systems, for example a gas station and a restaurant or convenience store, share the same location.

All of the data comes from the Uniform Franchise Offering Circular (UFOC) required by the FTC for those companies that meet its definition of a franchise. Nonetheless, it is the most comprehensive source that exists at the moment. It covers 1,226 franchise systems operating in 1998. All of the data refers to business format franchises.

This source breaks down the franchise population in its data base into 18 'industry' categories deemed relevant by the data collection organization.[5] It finds that four food related industries account for 34% of the population of franchise systems in its data base. These industries are : Baked Goods, Fast Food Industry, Restaurants and Retail Food Industry. Indeed, the Fast Food category contains the largest percentage of franchise systems in the data base (18%). It is followed by General Retail (11%), which comprises beauty related products, computer products and services and clothing and accessories and party related goods and services. This is then followed by a tie between Restaurants (9%) and Service Businesses (9%). The latter comprise health and fitness, publications, security related services and general services. Automotive Products and Services comes in next at (8%) and these five categories account for 55% of the franchise systems in the data base.

Limited but useful information is provided on several fundamental characteristics of franchise operations. For instance, 44% of the franchise systems in this data base have over 90% of their units as franchises. At the other end, 9% of the systems have only company owned stores. Presumably the latter observations correspond to franchise

[5] This classification of industries does not correspond necessarily to standard classifications such as SIC or NAICS.

systems in the early stages of development. About 53% of the systems have less than 50 units and 47% have more than 50 units. Incidentally most of these systems have been in operation for a while, since 62% of them have been in business over 12 years and 41% of them have been in franchising over 12 years.

With respect to the basic economic characteristics of a franchise contract, these data offer interesting insights. The initial franchise fee for 82% of these systems is over $10, 000 and for 7% of them it is over $50,000. The average initial fee is greater than $20,000 in 17 of the 18 industry categories identified in the data. These fees seem substantial, especially when one takes into account that large companies are unlikely to be in the data base. For instance, 95 % of the systems are listed as privately held companies.

A related item is initial investment in the franchise. For 43% of the systems it is less than $100,000 and for 34% of them it is between $100, 000 and $249,999. The average initial investment is less than $100,00 in 5 of the 18 'industry' categories in the data set. The lodging industry is an outlier where the average initial investment is over $2 million. While this initial investment is supposed to exclude real estate costs, it is not certain that respondents always did so in their answers.

Data on royalty fees reveal that 84% of these systems charge a royalty fee in percentage terms[6], 6% charge a flat royalty fee in dollar terms and 4% charge no royalty fee.[7] Of the 84% that charge a royalty fee in percentage terms, 93% do so as a percentage

[6] If a percentage fee had a dollar cap, it was still counted as a percentage fee in these data.

[7] The remaining 6% would be royalty fees related to the categories: inventories, transactions or other.

of sales or revenues and this percentage of sales lies between 3 and 6 % for 71% of them. Most of these systems require frequent payments of the royalty fee, for example 72% of them require the royalty fee to be paid on a weekly or monthly basis.

To conclude, this report provides data on three other interesting aspects of the franchise contract: advertising fees, territorial exclusivity, and contract length. 52% of the systems charge a national advertising fee as a percent of sales, 5% charge a flat rate, 28% charge no advertising fee and 12% charge only a regional or local advertising fee. Among those charging a national advertising fee as a percentage of sales, 83% charge between 0.01% and 3%. 71% of the systems offer grants of territorial exclusivity. Finally, 91% of the contracts offer the possibility of renewal and the average contract length for original contracts is 10.3 years and for renewals is 8.2 years. Thus, franchise contracts are long term contracts.

9.2. Retail Product Trade Name Franchises: Gas Stations.

For simplicity we begin the analysis by focusing on gasoline stations. A useful starting point is Shepard's (1993) seminal article on contractual form in this industry. She characterizes refiners of branded gasoline as having three basic choices of contract form to distribute their product through gas stations. One is company owned stations where the manager is hired by the refiner, which corresponds to the vertically integrated channel. Another is a lessee-dealer arrangement where the land and the immobile capital are owned by the refiner and the manager is self-employed, which corresponds to a franchise contract. The two main features of this contract are that the refiner sets the wholesale price and the annual rental or royalty fee, which is aimed at being proportional

to the net income that the station generates. Finally, the last one is an open dealer contract where the refiner has no investment in the station and controls only the wholesale price, except for restrictions on product purity and labeling and minimum purchase requirement.[8]

An interesting aspect of the franchise contract for lessee-dealer stations is that it does not rely on most of the incentive devices that we saw in Section 9.1 such as franchise fees, initial investment, advertising fees or territorial agreements. The only incentive device used from the list in Section 9.1 is a royalty fee as a percentage of net income or profits. The Appendix to this chapter presents a simple model that shows how royalty fees as a percentage of sales or profits (and a wholesale price set by the manufacturer) are sufficient to replicate the vertically integrated solution when the manufacturer provides a distribution service to the consumer. An essential feature of the interaction between refiners and their distributors is the provision by the refiner of a distribution service to consumers through the brand name.

More generally, an insightful paper on channel coordination by Lal (1990) shows that royalty fees are not necessary for channel coordination when the retailer provides a distribution service to the consumer or when franchisee retailers can free ride on each other. But Lal (1990) also shows that when the franchiser affects the demand of its franchisee retailers by providing a distribution service to the consumer royalty fees are necessary as a mechanism for channel coordination. The distribution service can be

---

[8] While Shepard does not discuss this option there are also independent stations that sell gasoline under their own name even if it is branded gasoline, for example see Hastings (2001). This latter category fits more closely the notion of arm's length contracts than the open dealer contracts.

provided, for example, through investments that maintain or expand a brand name. One way Lal motivates this result is by noting that the royalty payments assure the franchisee of profits by providing incentives for the franchiser to invest in the brand name.

One of the arguments established in Chapter 8, Section 4, was that retailers have profit incentives to expand into new products that are accentuated by the existence of distribution services. Thus, it is not surprising that one sees retail gas stations adding new product lines such as repair services or convenience stores, depending on the judgement of which local markets support one or the other profitably. What is of interest in the setting of choice of contractual form by refiners considered by Shepard is that repair services are far more difficult to monitor than convenience stores, in terms of revenues generated and a level of effort that produces an outcome for consumers with the appropriate level of assurance of product delivery in the desired form. This feature of repair services creates or enhances a principal agent problem for the refiner in his/her choice of contractual form.

An implication of this characteristic is that the principal agent problem favors the choice of a lessee-dealer contractual form or an open dealer form relative to a company owned form for gas stations that operate a repair shop. This is what Shepard finds in fitting a multinomial logit to data on the choice of contractual form by refiners for gas stations in Massachusetts. In each of three different specifications, Shepard (1990, Tables 4 and 5), the presence of a repair shop substantially lowers the probability of choosing a company owned contract form relative to a lessee-dealer form and this result is statistically significant at the 1% level in all three empirical specifications reported.

Similarly, the presence of a repair shop substantially lowers the probability of choosing a company owned form relative to an open dealer form, but not by as much as with respect to the lessee-dealer form.[9] The opposite result holds for the presence of a convenience store at a gas station: It increases the probability of choosing a company owned contractual form relative to the other two forms.[10]

A more recent study by Blass and Carlton (2001) uses nation wide data, limits the choice of contractual form to company owned and lessee-dealer forms, and finds similar results. The probability of having a lessee-dealer contract form in 1988 for 513 new stations built by refiners between 1984 and 1987 is larger the greater the number of service bays the gas station operates. The result is statistically significant at the 1% level and robust to limiting the sample to the 204 stations built in 1984-1985. These authors go on to estimate the implications of these results for the costs of extending the divorcement laws introduced in some states to the whole country.[11] Finally, a recent analysis based on a quasi-experiment in California, provided by ARCO's acquisition of over 260 independent gas stations from Thrifty in 1997, supports the case against divorcement. In this study Hastings (2001) found that the elimination of an independent

---

[9] Given the form in which the results are presented it is impossible to ascertain the statistical significance of this second result, but if one is willing to assume orthogonal regressors these results are also statistically significant at standard levels.

[10] Shepard explains the choice between lessee-dealer contracts and open dealer contracts in terms of capacity volume being sufficiently high under lessee-dealer contracts to justify the royalty payments. Her empirical results are consistent with this explanation. In turn this implies that the vertically integrated solution was not feasible in the markets served by open dealers. Open dealer contracts can't replicate the vertically integrated solution due to double marginalization.

[11] These divorcement laws ban or restrict refiners from having company owned stations.

(Thrifty) gas station increases prices in its local market, but these increases were unrelated to the choice of contractual form. That is, they were unrelated to whether the gas station was converted by ARCO into a company operated station or a dealer operated station.

9.3. Retail Product Trade Name Franchises: New and Used Automobile Dealers.

A useful starting point to discuss this type of franchise is the set of incentive features characterizing the franchise contract in this retail sector. On the negative side one finds that, just as in gas station franchises, there are no explicit franchise fees or advertising fees. For instance, Dnes (1992) reports this to be the case for the two auto dealer franchises (Austin Rover and Ford) in his analysis of 19 cases studies of franchise systems in England. In contrast to gas stations operated under lessee-dealer contracts, however, Dnes (1992) reports no royalty fees for the two car dealerships in his case studies of franchises. Similarly, Smith (1982) in his classic study of state restrictions on automobile distribution in the US makes no mention of franchise fees or royalty fees. He does discuss an end to cooperative advertising by GM in 1949 as an attempt to prevent passage of a nationwide act protecting dealers and subsequently as one of the restrictions imposed by state regulations. In the two English case studies, there are restrictions imposed by the manufacturer on local advertising but they refer to territory in which the dealer can advertise.

On the positive side the two main incentive features of franchise contracts for automobile dealerships are a significant initial investment by the franchisee in showroom, inventory, and repair facilities, and various forms of territorial restrictions. In the US

Smith (1982) notes that an initial investment to the satisfaction of the manufacturer is an explicit feature of franchise agreements. Indeed, Pashigian's (1961) early study of automobile distribution in the US shows how sizable these investments are. For example dealerships marketing a low price make in the late 1950's had an investment requirement of over $200,000 for city dealerships or dealerships expected to sell over 700 new cars annually (Table 30, p.126). These figures are exclusive of investments in real estate and buildings. Similarly, in England the two auto dealers studied by Dnes (1992) make specific investment in signs, tools and service facilities and have special provisions for financing stock of inventories.

Territorial restrictions are intrinsic to the functioning of automobile dealerships in the US. According to Smith (1982, p.130) " The franchisee was expected to provide display and repair facilities in exchange for which he was granted exclusive territory." And this characterized the system from 1903 to 1940. Since the manufacturer could terminate the agreement with some ease, this situation appears to have favored the manufacturer. The thrust of the rest of Smith article is an attempt to show how the balance of power has shifted towards the dealer as a result of legislation, especially by the states, limiting the manufacturer's ability to terminate the agreement, violate territorial restrictions or impose other costs on the dealer. Dnes(1992) also reports the existence of territorial restrictions for the two auto dealerships in his case studies of franchises in England. In these cases, however, the restrictions are with respect to location and marketing activities and the exclusivity restrictions are relaxed for small dealerships.

What are the main operating characteristics of auto dealerships? A differentiated durable good produced by a manufacturer is distributed by a retailer. The distribution process requires a fair amount of difficult to monitor and control effort in order to provide a high level of a distribution service through the repair and service facilities. That is, the retailer has to provide a high level of assurance of product delivery in the desired form in terms of a properly functioning vehicle in both the short-run and the long-run. The dealer also provides information through his or her showroom and demonstration cars. This information is easily appropriable by third parties who can acquire it by visiting one dealer while purchasing from another. Dealers have specialized knowledge of the local market, not easily available to manufacturers, in the form of prices to offer for trade ins. Most of the time, new car dealers also sell used cars and the latter are not subject to the franchise agreement.

In general manufacturers are attracted to a franchise agreement relative to company ownership because of the incentive features of residual ownership rights for the dealers under a franchise agreement. Dnes (1992) states (p.304) " The answer typically given by franchisers emphasizes that franchisees are more committed to their local business success due to having their own capital tied up in the enterprise." In this particular case manufacturers are also attracted to a franchise agreement relative to an arm's length relationship because of the existence of a principal agent problem induced by the need to operate at a high level of 'quality' in providing a distribution service downstream. The large initial investment in the franchise agreement acts as the posting of a bond by the franchisee, which makes him or her a hostage to the franchiser, in that it

is a "...highly firm-specific productive investment which will have only a low salvage value...", Klein (1980, p.358), if the franchise agreement is terminated. This provides an incentive to provide the right level of the distribution service by the franchisee that is absent in the arm's length relationship.[12]

Dealers are attracted to the franchise agreement for several reasons. Dnes (1992) reports that several of the franchisees of Austin Rover and Ford mention that the manufacturer's brand name provides legitimacy to their used car business.[13] The territorial restrictions in the franchise agreements generate monopoly profits. Indeed, Bresnahan and Reiss (1985) model this industry in terms of the double marginalization problem and the use of exclusive territories, i.e., independently of franchise and service considerations, and they show that the main implication of double marginalization is consistent with their data. Finally, territorial restrictions limit the extent of free riding by other dealers.

By the way of a conclusion we note that this type of franchise arrangement has provided the basis for automobile distribution for over one hundred years. Hence, it has passed the market test for durability. Perhaps the most important reason it has done so is its ability to reconcile the interests of the manufacturers and the dealers. This has taken place despite the basic incompatibility of incentives presented by the double

---

[12] Of course, the degree of asset specificity, and thus the strength of this argument, can vary a great deal from one industry to the next.

[13] The website of the National Association of Automobile Dealers (NADA Data) reports that in the US 17% of the percentage of operating profits of car dealers in 2002 came from sales of used vehicles (28.6% of sales), 48% of operating profits came from sales of parts and service (11.8% of sales) and 35% of operating profits came from new vehicle sales (59.6% of sales).

marginalization problem inherent in a differentiated durable product distributed through a system of spatial monopolies.

Some insight into this reconciliation is provided by Mathewson and Winter's (1994) analysis of territorial restrictions in franchise contracts. They develop a model based on maximizing the joint profits of franchisee and franchiser from the business. On this basis, they show that territorial restrictions should be positively associated with other restrictions in the contract designed to "...constrain upwards the retailer's efforts by fixing minimum quantities to be sold or inventories to be maintained....". This is certainly the case for automobile dealers who: must post the suggested retail price by the manufacturer and, thus, usually sell for a lower price, meet the franchise agreement's restrictions associated with the initial investment in repair inventories and parts, and face sales quotas or sales targets or sales bonuses that are restrictions or inducements designed to increase output.[14]

9.4. Business Format Franchises.

Once again a useful starting point for the analysis of this type of franchises is the set of characteristics of the franchise contract. Just as reported in Section 9.1, a typical business format franchise contract entails payments of an initial franchise fee, percentage royalty fees and a substantial initial investment by the franchisee. Two features of these payments by the franchisee are noteworthy. First, the differences in magnitude between the initial franchise fee and the initial investment are quite substantial, for example (as

[14] Recently, Marx and Shaffer (2001) have shown that nondiscrimination clauses in the original contract lead to maximizing the joint profit from the business if sunk costs are sufficiently low. In this case, they provide an alternative to territorial restrictions.

noted in Section 9.1) in 17 out of 18 industries the average initial franchise fee is greater than $20,000 but in 13 of these 18 industries the average initial investment is greater than $100, 000. Second, over 80% of business format franchise systems have all three kinds of payments by the franchisees specified in the franchise contract.

Interestingly, one of the fundamental theoretical articles on franchise contracts, Mathewson and Winter (1985), does not mention explicitly the initial investment in the franchise, but concentrates instead on the initial franchise fee as part of a nonlinear royalty fee. This article focuses on explaining the essentials that give rise to the existence of franchise contracts where profits are shared. Their conclusion can be summarized as (p.525) "...some form of binding franchisee wealth constraint or limitations on commitments and the incompleteness of contracts are both necessary and sufficient for franchise contracts. Vertical externalities are essential for franchises; horizontal externalities are not."

Dnes(1992) builds on the following Mathewson and Winter's formulation of the principal agent problem. The franchiser's profits, π_F, are given by

$$(9.1)\ \pi_F = FF + \alpha \{\Sigma\, t_i\, Q_i\, [D_i\,(dn), D(up)]\} - C(p) - D(up) - G,$$

where FF is the initial franchise fee, α is the royalty fee as a percentage of sales in both demand states (Q_i). Since prices are set at unity, sales (X_i) and output coincide and both depend on efforts provided by the franchisee (which can be interpreted as distribution services provided downstream), $D_i\,(dn)$, during the low (i=1) and the high demand (i=2) period. t_i is the probability that state i occurs, i = 1, 2. Sales also depend on the efforts of the franchiser (which can be interpreted as distribution services provided upstream),

D(up). The level of these services is assumed to equal the franchiser's investment in the brand. C(p) are the franchiser's monitoring costs and p is the probability of detecting chiseling or reporting falsely the state of nature by the franchisee. G are the franchiser's sunk costs.

Profits of the franchisee, given by π_E below, must be nonnegative.

(9.2) $\pi_E = \Sigma t_i \{(1-\alpha)Q_i [D_i (dn), D(up)] - D_i (dn)\} - FF \geq 0.$

Maximization of (9.1) subject to (9.2) yields the first best solution ignoring that the franchisee may chisel about the state of demand or that the franchiser may abscond with the franchise fee. Here is where Dnes formulation differs from Mathewson and Winter. Dnes (1992) adds two constraints to capture these issues in the form given below.

(9.3) $(1-\alpha)Q_2 [D_2 (dn), D(up)] - D_2(dn) \geq (1-p)\{(1-\alpha)Q_1 [D_1 (dn), D(up)] - D_1 (dn)y\}$

(9.4) $\pi_F + G \geq rFF.$

That is (9.3) requires that profits of the franchisee in the good state of demand be greater than or equal to profits that result from presenting the good state as the bad one when chiseling takes place (1-p) and the franchisee's effort is adjusted when chiseling (y). Mathewson and Winter also include the franchise fee in this constraint. Similarly, (9.4) requires that sunk costs plus the profits from franchising to the franchiser exceed the expected value of absconding with the franchise fee, where r is the probability that the franchiser absconds with the franchise fee. If we assume the worst about franchisers, r =1, (9.4) implies that when franchisers make normal profits, ($\pi_F = 0$), the franchise fee will cover the franchiser's sunk costs. Mathewson and Winter's formulation of this

constraint implies in the same extreme case that the franchise fee will cover half of the sunk cost, i.e., $2FF \geq G$.

It is interesting that in Dnes (1992) case studies initial franchise fees were viewed by most franchisers as covering set-up costs and by most franchisees as spent in covering initial set-up costs. These results are generally consistent with the view that the initial franchise fee provides the franchiser with some modest income to cover sunk costs and perhaps a small profit but it can not act as performance bond or hostage for the franchisee in the way the initial investment acts. In general this suggests that the insights derived from our previous two sections continue to be valid in this one. Namely, royalty fees as a percentage of sales are intimately related to the provision of a distribution service to consumers by the franchiser or upstream firm[15], large initial investments are a mechanism to ensure adequate provision of distribution services to the consumer by the franchisee or downstream firm, and initial franchise fees are neither necessary nor sufficient to address these two issues.

Initial franchise fees are too low to act as a mechanism for capturing rents by the franchiser. In a definitive study of this issue in the context of McDonald's franchises Kaufman and Lafontaine (1994) note that the existence of ex-post rents that they observe for McDonald's (Table 1) is consistent with most incentive theories of franchising. That is, the costs to the franchisee of reducing effort and having termination clauses applied to them by the franchiser as a result are made higher by the existence of these ex-post rents.

---

[15] The fact that 93 percent of percentage royalty fees are based on sales rather than profits is most likely due to the fact that sales are far easier to monitor than profits, for example Dnees (1992, p.236).

What these authors' find surprising in the case of McDonald's is that there also appear to be substantial ex-ante rents (Table 2). They explain these ex-ante rents in terms of the existence of liquidity constraints among the potential franchisees that McDonald's would like to recruit, which is consistent with one of the explanations for franchising stressed by Mathewson and Winter. While McDonald's has no territorial restrictions, 71% of the franchise systems considered in the IFA study of Section 9.1 offer territorial exclusivity to their franchisees. Given low initial franchise fees, these territorial restrictions are an explicit source of ex-post rents and a potential source of ex-ante rents for the franchisees.

Are there easy to identify operating characteristics of business format franchises?, and what effect do they have on outcomes? Most of the discussion, thus far, implicitly assumes that the product of the business format franchise is a service where production and distribution are not separated explicitly. This setting is usually described as brand franchising. It certainly applies to the equations defining the principal agent problem, (9.1) - (9.4), above. It is also appropriate as a description of the business of restaurants, including fast food ones such as McDonald. Nonetheless there are business format franchise systems covered, for example, by the data in Section 9.1 or in Dnes (1992) case studies that are not services.

For instance Dnes (1992) identifies three systems that sell a good to consumers other than cars: namely Apollo blinds, Bally shoes and Yves Rocher beauty products. While these systems are classified as business format franchises and rely on initial franchise fees and /or on percentage royalty fees, they have one feature that differentiates them from the pure service ones. The franchiser has an additional instrument to attain

optimal levels of operation for the system in the wholesale price of the product.[16] Interestingly enough, Rao and Srinivasan (1995) find that royalty rates are higher in franchise systems that sell services to consumers than in franchise systems that sell goods to consumers. Furthermore, they explain this empirical regularity as a consequence of the franchiser having two instruments (royalty rate and wholesale price) to induce franchisees to set output and service levels consistent with the system's optimum in the case of products, but only one in the case of services.

More generally, in a careful paper integrating theory and data on agency theory and business format franchising Lafontaine's (1992) main conclusion is that '...the empirical results are broadly consistent with two sided hidden-action or moral hazard explanation of franchising, suggesting that there really are incentive issues on both sides." Our emphasis on the provision of distributions services by both the franchiser and the franchisee is quite consistent with this conclusion. Moreover, it may provide the basis for better empirical measurements of the independent variables than available in the existing literature.[17]

A wide ranging survey of the empirical literature by Lafontaine and Slade (1997) summarizes several empirical regularities with respect to the choice of dependent variables, explanatory factors and inconsistencies with theoretical models. I will argue

---

[16] Incidentally, Dnes(1992) also identifies five systems that rely on a specialized input provided by the franchiser, for example Midas, which raise similar considerations. That is, the transfer price of the specialized input plays a similar role to the wholesale price in giving the franchiser an additional instrument to attain optimal levels of the system.

[17] This may be one of the reasons Lafontaine (1992) was able to explain the proportion of outlets franchised far better than the percentage of royalty rates or the franchise fees, for example see Tables 5 and 6 in her study.

that a critical factor in interpreting these regularities is that the literature tends to ignore the importance of the initial investments in the franchise by the franchisee as an aspect of the contract design.

First, Lafontaine and Slade (1997) note that most of the empirical studies end up focusing on explaining the extent of franchising, or proportion of company owned outlets, rather than dimensions of the franchise contract in terms of initial franchise fees or royalty fees. They also note that the three exceptions that consider these two contract dimensions find results similar to the ones found for the proportion of company owned stores and, thus, restrict their attention in the survey to the latter studies.[18] For all three choices of dependent variables, however, it is customary to include as an explanatory variable some measure of the capital required to open an outlet. This introduces a serious simultaneity problem into the analysis of all three dependent variables if initial investments by the franchisee are an important aspect of the implicit or explicit contract design. To my knowledge, no empirical study has acknowledged let alone addressed this econometric problem. The information in Section 9.1 suggests that the initial investment is as important or a more important dimension of the franchise contract than the franchise fee.[19]

---

[18] One of these studies is the one by Lafontaine (1992) previously discussed. A second one is Sen's (1993) analysis of franchise fees and royalty fees, which concludes that channel control is the main factor influencing payment design. Two of the main determinants of channel control in his view are franchiser's commitment to the brand name and malfeasance by the franchisee, which correspond to the double moral hazard view stressed by Lafontaine (1992). The third study is Lafontaine's (1993) analysis (and rejection) of signaling models as an explanation of these dimensions of contract design in franchising.

[19] An interesting study by LaFontaine and Shaw(1999) is the first one to look at dynamic issues empirically. It attempts to explain variations in the royalty rate and the franchise fee and it instruments for each of them in the relevant equation, but it does not include as an explanatory variable in these regressions the initial investment (with or without instrumenting for it). One wonders if the large amount of explanatory power generated by the

Second, Lafontaine and Slade (1997) identified nine empirical regularities in the studies they reviewed and argued that six of them are consistent with theoretical models of franchising. The validity of this consistency between the theoretical predictions and the six empirical regularities is contingent on assuming that the simultaneity bias from ignoring the endogeneity of the initial investment by the franchisee is immaterial to these results. For, the initial investment by the franchisee, or variables highly correlated with this initial investment such as size of outlet measured by capital or sales, are often used as explanatory variables in these studies.

Third, at least two of the three empirical regularities that Lafontaine and Slade (1997) find that are supposedly inconsistent with theoretical models of franchising can be made consistent with these models once one acknowledges the importance of initial investments by the franchisee as part of the design of the franchise contract.[20] For instance, they argue that outlet size should be inversely related to the proportion of company owned outlets. Because it measures the importance of the agent's input into the final outcome of the business activity and the more important is this input the more one should observe franchise operations. Three of the six studies listed by them measure outlet size by the initial investment. All three find a positive effect on the proportion of company-owned establishments. Instead of contradicting theoretical predictions, this is exactly what one would expect if initial investment is part of the bond or hostage

---

firm specific fixed effects is partly due to the omission of this variable from the regressions explaining royalty rates and franchise fees.

[20] The third supposed inconsistency is that there is a positive empirical association between risk and franchising, but this is an inconsistency with the theory only if one assumes that franchisees are risk averse and franchisers are risk neutral. In any event, the analysis here has nothing to add on this issue.

dimension that the franchisee must contribute to the business in the contract design (and the simultaneity bias does not lead to a change in the theoretically expected sign).

Along similar lines, they list low within firm (franchise system) contract uniformity as one of the empirical findings predicted by the theory that is contradicted by the data. Since they find that most franchise systems offer uniform contract terms to their franchisees in terms of initial franchise fee and percentage royalty fees, this is viewed as a contradiction of the theoretical prediction of high variability. Nevertheless, if one views the initial investment as part of the franchise contract design it is not clear that there is low within firm contract uniformity in the first place. The empirical regularity that one finds is that the franchise fee is not used as an important instrument to drive the system toward optimal values by capturing downstream rents, but Lafontaine's explanation for this phenomenon in her earlier work (1992) is quite sensible and consistent with the principal agent model in, for example, Dnes (1992) or Mathewson and Winter (1985). Namely, (p.281) "...if the franchise fee is chosen simply as a way to remunerate the franchiser for services offered in starting out the franchise, there would be no need for any relationship between the franchise fee and the royalty rate."

More recent work also ignores the importance and endogeneity of the initial investment by the franchisee.[21] For example, Affuso (2002) provides some evidence that, in addition to double sided moral hazard, there is an adverse selection or hidden information problem in the decision to franchise. She finds at the outlet level that the

[21] Some of the more recent work considers new issues where this consideration may be less relevant, for example the determinants of multi-unit ownership as in Kalnins and Lafontaine (2003)

wages of the agent in a prior job are positively related to the probability of franchising, but one of her control variables treated as exogenous is the initial investment by the franchisee.

A similar neglect arises in a recent paper by Blair and Lafontaine (2003). In Section 9.3 we saw how franchising in the new and used car dealers sector revolved around the requirements of significant initial investment by the franchisee in exchange for grants of territorial exclusivity by the franchiser, which in turn led to a substantial amount of legislation by states. Blair and Lafontaine's (2003) paper analyzes conflicts between franchisers and franchisees in the context of business format franchises, since similar legislation has been undertaken recently in this context. While they devote substantial attention to the use of territorial restrictions as an ex-ante solution to these conflicts, their analysis pays no attention to the role of initial investments by the franchisee in generating these conflicts. For instance, it is not explicit in their illustrative model (fixed investment by the franchisee is the same every year) and it is not directly discussed. Indeed, the topic appears only indirectly in a footnote citing Lutz (1995) work on the differential incentives, relative to managers, provided by asset ownership to franchisees.

To conclude, I discuss briefly the last two dimensions of business format franchise contracts identified in the IFA study: advertising fees and contract length. Advertising fees as a percentage of sales have been treated in the literature as simply another form of royalty fees as a percentage of sales. This is true of both the theoretical literature, e.g., Mathewson and Winter (1985), and the empirical literature, e.g.,

Lafontaine (1992) and Sen (1993). Such treatment seems warranted for the 52% of franchise systems that charge a national advertising fee as a percentage of sales and unnecessary for the 28% that charge no advertising fee. Similarly, not much attention has been paid in the literature to contract length. One exception is Thompson (1994) who found little evidence for a life cycle effect where franchise systems revert to an increasing proportion of company-owned outlets.

Appendix: The Basic Role of Royalty Fees.

A manufacturer that provides a distribution service to the consumer in a vertically integrated channel can be characterized as follows in the simplest possible setting.

(A1) $\pi(I) = pQ - cQ - aD$,

where c is a constant marginal cost with respect to Q and a is a constant marginal cost with respect to D. Q is the consumer's demand function, which is specified as usual, i.e., $Q = f(p, D)$. D is a distribution service. It can be thought of as information provided

through advertising or through the development of a brand name, or more generally as any distribution service provided directly to the consumer by the manufacturer.

The manufacturer maximizes (A1) through his choice of p and D. The first-order conditions will be given by

(A2) $\partial\pi/\partial p = p\,[1- (1/\epsilon)] - c = 0.$

(A3) $\partial\pi/\partial D = (p-c)\partial Q/\partial D - a = 0.$

The retail price is chosen so that marginal revenue equals marginal cost and the level of the distribution service is chosen so that the marginal revenue it generates also equals marginal cost. Thus, this is just a special case of the model in Chapter 2, Section 2.5, and the variables are defined in exactly the same way. << Maximum profits will be given by $\pi^*(I) = p(I)Q(I) - cQ(I) - aD(I)$, or profits at the vertically integrated solution.>>

Under a decentralized arrangement, we can characterize the problem of the retailer as follows:

(A4) $\pi(R) = pQ - wQ - \alpha pQ$,

assuming a royalty fee based on a percentage of sales. The symbols have the same meaning as before, w is the wholesale price charged by the manufacturer and α , $0 \leq \alpha < 1$, is the royalty fee charged by the manufacturer. Profit maximization by the retailer entails choosing p, given w, α and the market demand function $Q = f(p, D)$. The first-order condition will be

(A5) $\partial\pi/\partial p = p[1- (1/\epsilon)] - w/(1- \alpha) = 0.$

If the royalty fee is zero, so that we have an arm's length relationship between manufacturer and retailer, any wholesale price higher than marginal cost will lead the

retailer to charge a higher retail price and the double marginalization problem arises. If the wholesale price charged equals marginal cost, c, the retailer will charge the same retail price as under vertical integration, assuming the manufacturer provides the same level of service. Of course, in this case the retailer obtains all the monopoly profits as well as the revenues the manufacturer would have used to cover the costs of providing the distribution service under vertical integration. In this setting the manufacturer will go out of business or provide zero levels of the distribution service and make zero profits.

Instead of going out of business or making zero profits, however, the manufacturer can obtain the vertically integrated solution by charging the retailer a nonzero royalty fee and a wholesale price, w, that differs from the marginal cost of an additional unit of output, c. Any combination of w and α such that (A6) below holds,

(A6) $w/(1 - \alpha) = c$,

will lead the retailer to charge the same price as in the vertically integrated solution given that the manufacturer is providing the same level of distribution services to the consumer. Thus the higher is the royalty rate the lower will be the wholesale price (below marginal cost) so that (A6) holds.

How should the royalty rate be set? The manufacturer wants to extract all the monopoly profit by replicating the vertically integrated solution. It will set the royalty rate so that the retailer bears the cost of providing the level of distribution services the manufacturer would have chosen under vertical integration <<[plus the maximum profits obtained under the vertically integrated solution, i.e.,

(A7) $\alpha = aD(I)/p(I)Q(I) + \pi^*(I)/p(I)Q(I)$.>>

Given the level of α implied by (A7), the wholesale price must be chosen according to (A6) so that the consumer faces the same level of prices and distribution services as under the vertically integrated solution. All of the retailer's monopoly profits are extracted through the royalty fee and <<part is >> used to cover the costs of providing the distribution services by the manufacturer, who enjoys the same level of profits as under complete vertical integration.

If the royalty fee is based on a percentage of profits, then the retailer's optimal choice of price will be independent of the royalty fee. The manufacturer can set the wholesale price, w, equal to marginal cost, c, and replicate the integrated solution by providing the same level of services as under the integrated solution and setting the royalty fee according to (A8), so the retailer bears the cost of providing the service <<and the manufacturer enjoys the same monopoly profits as under vertical integration.

(A8) $\alpha = a\, D(I)/ \{[p(I) - c]Q(I) + \pi^*(I)/[p(I) - c]Q(I)\}$. >>