Abstract

Theoretical and empirical work on franchising has developed from agency theory and from ideas about asset specificity and opportunism associated with transaction-cost analysis.

I begin by considering some traditional arguments about the capital-structure function of franchising. Next, I consider agency and transaction-cost theoretical explanations of franchising. An interesting special case is where the franchisor also runs company stores. Econometric work supports the view that franchise contracts protect against reciprocal opportunism. I also examine several arguments concerning the possible nature of 'hostages' in franchise contracts.

JEL classification: K23, L22

Keywords: Franchising, Agency, Search Theory, Hostages

1. Introduction

Franchising is an organizational form lying between markets and hierarchies, and can follow either a business-format or a simpler dealership model. It is a symbiotic relationship between businesses (Schanze, 1991). Business-format franchising, in which the franchisor supplies a brand name and also a model business for the franchisee to copy, is the growing sector of franchising and covers businesses like vehicle rental and fast-food restaurants. Many of the differences between business-format franchising and dealerships (for example, cars or petroleum) are disappearing over time as manufacturers provide a wide range of support for their dealers. Theoretical and empirical work on franchising has developed from agency theory and from ideas about asset specificity and opportunism associated with transaction-cost analysis.

I begin by considering some traditional arguments about the capital-structure function of franchising. Next, I consider agency and transaction-cost theoretical explanations of franchising. An interesting special case is where the franchisor also runs company stores. Econometric work supports the view that franchise contracts protect against reciprocal opportunism. I also examine several arguments concerning the possible nature of 'hostages' in franchise contracts.
2. Franchising as a Method of Raising Capital

An early argument is that firms franchise to raise capital for expansion (Caves and Murphy, 1975). Rubin (1978) argues that this makes no sense unless we assume that the franchisor is more risk averse than the franchisee, which is implausible. Even if franchisors could not use normal capital markets, they could sell shares in a portfolio of all outlets. The shares would diversify risk for the buyers but impose no costs on the franchisor. Franchisees would pay less for undiversified investments if they are risk averse, which implies smaller returns for franchisors. Any capital-market advantages from franchising must come from shifting risk to the franchisee, which only makes sense if the franchisor is the more risk averse.

Rubin’s argument that capital raising does not explain franchising depends upon an assumption of zero transaction costs. Franchising can be a capital issue under less restrictive assumptions. However, empirical work generally supports organizational costs rather than capital-market influences as the driving force behind franchising. Lafontaine (1992) discovered that increases in the capital cost of opening stores reduced the proportion of franchised outlets, which is contrary to the capital-raising story.

3. Franchising as a Problem of Monitoring and Control

Rubin explains the features of franchising in terms of solving monitoring problems. In retail networks where the satellite business is remote from the head office monitoring is difficult and it pays to develop an incentive system that encourages the avoidance of shirking. A profit-sharing agreement gives the franchisee sufficient residual profits to make shirking too costly. The franchise chain will show more total profit if shirking is controlled. Franchisors will not pay any more profit to franchisees than is necessary to remove the incentive to shirk. A competitive supply of prospective franchisees should be willing to pay lump sums equal to the difference between franchise profits and what they could earn as managers in similar occupations.

We do not usually observe franchise contracts of this kind. Instead franchisees pay a lump-sum initial fee, and a continuing royalty payment related to sales, in return for residual profits. The most plausible explanation is that the franchisee requires protection against poor post-contract performance by the franchisor. The franchisor’s duties cover such things as providing managerial support and the monitoring of standards of operation throughout the franchise system. Monitoring of the system is necessary to control a classic externality problem: if one franchisee allows quality to deteriorate, he benefits by the full amount of the savings from reduced quality but incurs only part of
the costs as other franchisees will suffer some of the loss of business. This type of externality is described by Mathewson and Winter (1985) as horizontal free riding.

The theory generates several predictions. Increasing the geographical density of outlets should make operating company stores more attractive. Also, franchisors should buy back their outlets as their chains become more mature, the density of outlets increases, and distance-related monitoring costs become lower per outlet. Buy-backs are observed in mature chains. Much econometric work supports the importance of geographical density in explaining franchising (Lafontaine, 1992; Brickley and Dark, 1987; Norton, 1988). Lafontaine also finds evidence that increases in the importance of the franchisor’s inputs increase the royalty, which supports the view that franchise contracts are partly constructed to control the franchisor’s opportunism.

4. Modelling Franchising as an Agency Relationship

Mathewson and Winter (1985) argue that horizontal externalities are not necessary to explain franchise contracts. Monitoring difficulties arise for the franchisor even when there is only one territory. However, vertical externality (chiselling on the franchisor’s standards) is an ever-present problem. Risk aversion on the part of the franchisee is also not a sufficient condition for the emergence of a franchise contract. The franchisor could impose a large penalty if the franchisee were caught cheating, making the franchisee’s income the same across different demand conditions and giving a pure risk-sharing contract with no profit-sharing. However, the penalty may be infeasible owing to wealth constraints affecting the franchisee and this gives profit sharing.

In their model local demand at a franchised outlet is subject to uncertainty and may take a high or low state. The franchisor cannot costlessly identify any ruling state of demand. The franchisee has better local information and may attempt to reduce the quality of his effort in high demand states and try to pass off the resulting low output as due to a low demand state, reflecting a problem of franchisee moral hazard. The franchise contract specifies the franchisee fee schedule (lump sum plus royalty) plus the quality of the franchisee’s input in good and bad demand states.

Mathewson and Winter agree with Rubin that the first-best contract between franchisor and franchisee would lease the brand name in return for a lump-sum payment. The franchisor would be encouraged by the incentive of maximizing the fee to find the joint-profit-maximizing monitoring arrangements. Each franchisee would pay a fee conditional on the value of the brand name and therefore dependent on the optimal amount of monitoring, and could enforce the contract ex post.
If it is infeasible to cover all aspects of the franchise relationship in an explicit contract, profit-sharing emerges. In their basic model, Mathewson and Winter attribute this to a constraint on the wealth of franchisees that prevents them from sinking large investments into franchises. This empirically relevant constraint makes franchisors rely on rewards rather than the penalty of termination to maintain franchisees’ standards. An incentive-compatibility constraint in their model ensures that the profit accruing to the franchisee from correctly declaring the better demand state and applying the correct effort level exceeds the profit from wrongly declaring the poor state and adjusting effort downward. A participation constraint ensures that the contract gives sufficient profit for the franchisee to pay a royalty fee. Mathewson and Winter derive the franchise fees, franchise effort in each state, the level of brand-name investment by the franchisor (including advertising) and the frequency of monitoring.

The removal of the wealth constraint from the model opens up the possibility that franchisees could post bonds to guarantee good performance. Mathewson and Winter agree with Rubin that bond posting is problematic as the franchisor might behave opportunistically. The expected value of the lump sum must be less than the profits accruing to the franchisor from the proper delivery of services. Otherwise, there will be an incentive for the franchisor to abscond with the lump sum, possibly by contriving some reason for contract termination. The royalty, or its equivalent, is always the engine for rent extraction.

5. The Organizational Mix

‘Dual distribution’ is an important phenomenon. Gallini and Lutz’s (1992) model shows that both dual distribution and the use of a sales royalty may be methods by which a new franchisor signals the profitability of the franchise chain by making franchisor returns dependent on the revenues of company stores.

Consider the case where a franchisor with a fixed number of outlets knows that demand is favourable so that stores should be unusually profitable. The problem is to convey this information in a credible manner to prospective franchisees. The high-profit franchisor chooses the proportion of company stores, the lump sum and the royalty to establish a separating equilibrium defining a contract that a low-profit franchisor would never offer. A separation constraint ensures that a low quality franchisor will always make more profit from truthfully declaring quality and franchising all stores, compared with emulating the dual-distribution strategy of the high-profit franchisor.

A number of predictions may be made on the basis of signalling theory, but they are not supported by empirical work. To take one as an example, the high profitability of some franchises would be recognized over time and there would be no need for franchisors to operate company stores as a signal. We should see
mature franchise chains concentrating on franchising, rather than the operation of company stores. Whilst there is possibly some support for this hypothesis, for example Martin (1988) observes that older units are often franchised, economists often observe a buy-back phenomenon (Thompson, 1994) as the chain matures. Lafontaine (1993) reports econometric results showing no support for a range of hypotheses suggesting that franchisors use their organizational mix as a method of signalling.

6. A Search Theory of Franchising

Minkler (1992) has suggested that franchising is a device through which the franchisor gathers and uses local information. The theory is Austrian in character and emphasizes the key role played by information in the competitive process. There is a dark side to franchising in Minkler’s approach: franchisees are useful temporary tools, rather as in some of the small-business literature (Hoy, 1994; Bates, 1995).

According to agency-based theories of franchising distance of the satellite business from the mother company, which makes monitoring more difficult, should be associated with an increased reliance on franchising. However, Minkler cites examples where franchised and company stores operate in close proximity to each other. For example, in Sacramento, California, 34 Taco Bell restaurants covered a 30-mile radius, of which seven were company owned.

Minkler argues that franchisors draw on the local knowledge of franchisees, which concerns local tastes and market conditions. The franchisor might be unable to direct the satellite business, even if monitoring costs were zero, because of ignorance. Franchising allows the use of the trade mark to be exchanged for the franchisee’s local entrepreneurship, which is defined as noticing and acting upon opportunities. The franchisee’s local entrepreneurship reduces the cost of search for new business.

How reasonable is the search-cost theory? Empirical work by Minkler shows that older outlets are more likely to be franchised than newer ones, which is consistent with the theory and with Martin’s (1988) results, although it is not consistent with Thompson (1994). A problem is that the buy-back phenomenon is consistent with many theories: for example older stores may be easier to monitor owing to experience effects unconnected with distance. It is difficult to imagine an empirical test to distinguish Minkler’s theory from others.
7. Vertical Restrictions and Franchising

Within the mainstream industrial-organization literature there are papers which show that a firm with monopoly power supplying an intermediate product into a competitive industry has an incentive to exercise vertical control if downstream input substitution is possible. Vertical restrictions principally comprise refusal to supply, tied-in sales, and exclusive-dealing contracts. The arguments of several economists that there are efficiency reasons for all of these practices are reflected in the specialist economic analysis of franchising, and in the benign view taken by European competition law towards franchising (Dnes, 1991c). For example, against simple claims that a monopolist could foreclose a downstream market by refusing to supply unless buyers were tied into a restrictive contract, it may be argued that it is profitable to allow access to inputs at monopoly prices to more efficient downstream firms. Although, to be fair, some recent analysis has revealed conditions under which refusal to supply (Bolton and Whinston, 1993) is a credible policy committing a firm to compete aggressively in the downstream market and deterring entry.

Analyses of franchising based on monopoly-power explanations of vertical restrictions are typically less general than theories based on the economics of organization. As a very simple example of lack of generality, note that monopoly-power theories of vertical restrictions usually deal with product franchises, when most business-format franchises are based on services, and would seem to have relevance only for brand-and-trade-name franchising. The relevance of the market-power approach is further questioned by a lack of supporting empirical evidence: for example Lafontaine (1992) found that the proportion of franchised outlets decreased as franchisor input sales increased.

Blair and Kaserman (1982) formulate a two-period model that does represent the franchise contract as a mixture of vertical controls. The model predicts use of both a lump sum and a royalty whenever the franchisor’s and franchisee’s discount factors diverge (reflecting perceptions of uncertainty). Blair and Kaserman avoid regarding individual controls like resale price maintenance (RPM) and franchise fees as perfect substitutes for one another. In general, franchising firms use a mix of contractual devices and cannot be indifferent between them.

Blair and Kaserman suggest there may be complementarity between monopoly-power and organizational explanations of common features of franchised businesses. A franchisor with the relatively lower discount factor would not be able to extract all the expected downstream rent from the franchisee. Thus, post-contract tensions would arise as the franchisor saw franchisees enjoying super-normal profits. If franchisor uncertainty over forecasts fell over time, mature franchise chains would open more company stores. Blair and Kaserman also suggest that the franchisor can practice
post-contract opportunism. The franchisor must promise the franchisee a normal return on investment. Afterwards, however, the franchisor may be able to increase his share of sales revenue without provoking the franchisee to close down (if there were worse losses from closing). The franchisor may use strategies like forcing, where quotas push sales past the point of profit maximization for the franchisee. Blair and Kaserman share some of the concerns over franchisor and franchisee incentive compatibility shown in the organizational literature and are not solely motivated by traditional market-power issues.

Efficiency-based explanations of vertical restrictions are descended from Telser’s (1960) analysis of RPM. A retailer could provide service levels like advice and product demonstrations only to find that consumers made use of these and then bought the product at a low price from a no-frills retailer. There is a free-rider problem among retailers implying that no retailer would provide services. If service levels matter in promoting sales for the manufacturing and retailing industries combined, and are not separable, a means like RPM must be found to defeat free riding. Marvel (1982) explains exclusive dealing, which is a common feature of franchising, in a similar fashion. When a manufacturer with a valuable brand supplies an outlet it endorses the retailer’s business and centralized advertising may promote the retailer’s sales more generally. Marvel argues that exclusive dealing prevents retailers from diverting business to other brands and wasting advertising.

Klein and Saft (1985) examine tied-in sales and argue that franchisors use these either to control the quality of the final service, or to measure the sales of franchisees. Where the franchisee cannot substitute away from the input, a mark-up on a tie-in is equivalent to a fixed percentage sales royalty if price is predictable. Tie-ins may also develop where the franchisor wishes to ensure that franchisees use inputs of specific quality. Rather than monitoring the required technical properties of more generic inputs, the franchisor has the much simpler problem of ascertaining whether anything else was used.

8. Hostages in Franchise Contracts

Transaction-cost analysis shows that franchise contractual provisions that are often regarded as unfair in the law have important implications for efficiency (Klein, 1995). Fully contingent, costlessly enforceable, explicit contracts are not usually feasible. Uncertainty implies a large number of possible contingencies and some aspects of contractual performance are difficult to measure. Individuals have an incentive to renege on agreements and hold-up any contracting partner who has made specific investments by taking advantage of unspecified or unenforceable aspects of contracts. Full vertical integration
between trading partners will not always be observed: for example integration of human capital is outlawed by the prohibition of slavery.

One method of safeguarding performance is for a potential cheater to post a bond (a ‘hostage’) possibly in an implicit form if the cheater is required to make an investment in a highly specific form with a very low salvageable value. In both cases, the same purpose is served. Franchise contracts typically require franchisees to pay lump-sum fees to franchisors and to make highly specific investments in equipment. The franchisor usually takes the right to terminate the contract at will if the franchisee is not maintaining quality standards. For any hostage to be effective it must set the franchisee’s expected gain from cheating equal to zero. This implies that hostages will be worth much more than the actual gain when monitoring costs are positive. Cheating by the franchisor is controlled by possible increases in operating costs. A franchisor known to appropriate hostages opportunistically would lose franchisees and find it hard to recruit new ones, forcing him to use more costly organizational forms. As long as the franchisee’s bond is greater than the franchisee’s expected gain from cheating and is less than the cost penalty imposed on the franchisor on moving to some other organizational form, a hostage can support their relationship. The hostage is a low-cost substitute for costly monitoring and enforcement devices.

Of particular interest is Klein’s argument that the franchisor’s contractual right to terminate the contract at will (for good cause) supports a number of hostages. Given termination at will, the common requirement that franchisees lease their properties from the franchisor can be explained. The franchisee could be forced to move premises and sacrifice valuable leasehold improvements, which would revert to the franchisor as lessor. This gives the franchisor a hostage with which to control franchisee behaviour and enables monitoring to be reduced with an associated cost saving. In recent years Klein has moved to the view that the rents attached to the non-salvageable investment should be the focus in valuing the franchisee’s potential loss, at least in cases where there are no binding legal constraints on the franchisor’s behaviour.

It is important to recognize the rich variety of devices used to support contracts. The use of restrictive covenants in franchise agreements can also be explained in terms of hostages. Covenants usually prevent a franchisee from competing in a market area for some period after leaving the franchise system, implying that the non-availability of an alternative rent stream is used to constrain the franchisee’s behaviour: that is, he cannot cheat and leave for better pastures. A new franchisee’s future level of skill is not known but if he becomes highly adept at his business, he might be tempted to set up on his own. A covenant prevents the franchisee from simply removing the franchisor’s investment in his training. Also, termination by the franchisor can cause the loss of the hostage.
Williamson (1985) makes some suggestions concerning likely hostage selection. Implicit hostages are less vulnerable to opportunistic appropriation by trading partners compared with pecuniary hostages. A hostage can be selected to be unattractive to its holder. An ideal hostage is like an ‘ugly princess’: the medieval king with two equally cherished daughters would be wiser posting the ugly one as a hostage, as she is less likely to be appropriated by the captor.

A number of common observations emerge from studying franchisees’ contracts (Dnes, 1993d). Franchising increases the specificity of investment for the satellite business, compared with independent operation, for example leasehold improvements are trademarked and hard to adapt to other uses. Also, lump-sum fees are typically small in relation to sunk investment for the franchisee and appear to be linked to the franchisor’s costs of establishing the franchisee (training and launch advertising). The implicit aspects of contracts are important and show adjustments that favour the interests of both franchisees and franchisors.

The feasibility of placing disciplinary hostages with franchisors is qualified by the explicit and implicit details of franchise contracts, which often set out conditions under which the franchisor must buy-back assets in the event of termination. Statute law in some countries, like the USA, also makes it difficult to call in a hostage for disciplinary reasons. Principles of common law, such as the prohibition against penal damages for breach of contract, may also make disciplinary hostages illegal in an Anglo-American setting.

It is not surprising that franchisees are careful to avoid hostage penalties in their contracts: investments in such things as leasehold improvements are not ugly princesses but are of potential direct value to the franchisor. There are several questions about the real-world feasibility of disciplinary hostages, regardless of whether these are measured as rent streams or as the book value of sunk investments. Sunk investment by the franchisee may well have mainly a screening function, serving to demonstrate confidence in his own competence.

9. Conclusions

The last decade has witnessed progress in the scientific understanding of franchising. Several theories have been constructed to explain franchising, most of which emphasize savings of monitoring costs in an agency framework. Details of the theories show how opportunism on the part of both franchisors and franchisees may be controlled. In separate developments, writers have argued that franchisors recruit franchisees to reduce information-search costs, or that they signal franchise quality by running company stores.
Empirical studies tend to support theories emphasizing opportunism on the part of franchisors and franchisees. Thus, elements of both agency approaches and transaction-cost analysis receive support. The most robust finding is that franchising is encouraged by factors like geographical dispersion of units, which increases monitoring costs. Other key findings are that small units and measures of the importance of the franchisee’s input encourage franchising, whereas increasing the importance of the franchisor’s centralized role encourages the use of company stores. In many key respects, in result although not in principle, transaction-cost analysis and agency analysis are just two different languages describing the same franchising phenomena.

Bibliography on Franchise Contracts (5890)

Franchise Contracts


