

EPPP DP No. 2010-08

**Monitoring Policy and Organizational
Forms in Franchised Chains**

Thierry Pénard
Emmanuel Raynaud
Stéphane Saussier

Monitoring Policy and Organizational Forms in Franchised Chains*

Thierry PENARD

CREM, University of Rennes I
Faculté des Sciences Economiques, 7 Place Hoche,
35065, Rennes cedex, France
thierry.penard@univ-rennes1.fr

Emmanuel RAYNAUD (corresponding author)

INRA SADAPT & CES (University of Paris I)
INRA SADAPT, 16 rue Claude Bernard,
75231, Paris cedex 05, France
emmanuel.raynaud@agroparistech.fr

Stéphane SAUSSIÉ

CREGOR, IAE University of Paris I Panthéon-Sorbonne,
21 rue Broca, 75005, Paris, France
saussier.iae@univ-paris1.fr

* We would like to thank James Brickley, Eric Brousseau, Michael Raith, Joel Ruet, Vivian Dos Santos Silva, Josef Windsperger, the editor and the two referees for their helpful comments. A previous version of this paper was presented at ATOM (University Paris I), INRA SAE2 (Dijon), the ISNIE conference held in Boston (MIT), the EMNET conference in Vienna and at the fourth IIOC Conference in Boston (Northeastern University).

Monitoring Policy and Organizational Forms in Franchised Chains

ABSTRACT:

Franchising is nowadays a prominent way to organize the distribution sector. While previous literature suggests that monitoring issues are a critical determinant of organizational choices, it is rather silent on the optimal monitoring strategy once the organization of the chain is set. In this article we analyze the monitoring policy of chains with both franchised and company-owned units. We develop a model in which a chain monitors its outlets under asymmetric information on local demands and managers' efforts. We show that partial monitoring (*i.e.* when the franchisor monitors only a subset of its outlets) represents an optimal monitoring policy. Second, we identify the units that should be monitored. Finally, we discuss the impact of information technologies and outlet location on monitoring policy and how it may affect the proportion of franchised and company-owned units within the mixed chains.

KEYWORDS: asymmetric information, dual distribution, franchising, monitoring, moral hazard.

JEL classification: D23, L14, L22

1. INTRODUCTION

Franchising is nowadays a prominent way to organize the distribution of goods and services worldwide. According to a recent study on the economic impact of franchising commissioned by the International Franchise Association (IFA), the US franchise market generates over \$1.4 trillion a year in revenue and thus accounts for over 45% of all retail sales. Franchising is also widespread in other parts of the World.¹ The economic importance of franchising partly explains why this contractual relationship has received a significant amount of attention in the economic literature on contracting (see Blair and Lafontaine, 2005; Lafontaine and Slade, 2007, for surveys). Franchising is a “hybrid” organizational form, which lies between vertical integration and spot markets (Williamson, 1991). In addition, franchising is one of the few types of contractual relationships about which significant amounts of data are available from public sources. Thus insights gleaned from the study of franchise contracts have allowed researchers to develop a better understanding not only of this organizational form, but also of organizational issues more generally, both within and across firms.

Most works on franchising emphasize the role of agency problems in shaping organizational choices. Two prominent dimensions of the contractual agreements have been extensively studied. The first one is the design of the franchise contract with much of the emphasis on the level of royalty rates and franchised fees (Mathewson and Winter, 1985, Lafontaine, 1992, Sen, 1993, Wimmer and Garren, 1997, Vazquez, 2005). The second one analyzes the agency factors likely to affect the trade-off between franchising and company

¹ For example, more than 1100 different chains were active in France in 2007 with an amount of revenues over €45 millions (source: French Franchise Federation).

ownership (Brickley and Dark, 1987; Lafontaine, 1992; Brickley, 1999).² An important stylized fact is however the coexistence of franchised and company-owned units within the same chain, known as “plural form” or “dual distribution” in the literature (Lafontaine and Shaw, 2005).

In all these works, the central idea is that by delegating to the units’ manager (franchisee or salaried manager) the right to use its brand, a chain will incur agency costs (Brickley and Dark, 1987, Rubin, 1978). Aligning incentives through contractual provisions and monitoring managers’ behavior is thus a critical issue for the chain. For instance, Brickley and Dark (1987) found that chains rely more often on franchised units when the direct costs of monitoring behavior (effort or input monitoring) is large.³

While previous literature suggests that monitoring issues are a strong determinant of organizational choices, it is rather silent on the optimal monitoring strategy once the organization of the chain is set. In this article we analyze the monitoring policy of chains composed of franchised and company-owned units. The monitoring policy is made of a probability of inspecting individual units as well as a level of penalty. The probability of monitoring may represent the intensity of monitoring expenditures (for instance how many times a given unit will be audited) or the proportion of units that will be inspected. We develop a model in which a chain monitors its outlets under asymmetric information on local demands and managers' efforts. Outlets are located in various commercial areas and local demands are subject to random and idiosyncratic shocks. Two kinds of opportunistic behavior may appear in our model depending on the manager's status. In a company-owned

² A more recent stream of papers analyzes the performance implications of various dimensions of the chain governance (see for instance, Azoulay and Shane, 2001, Kalnins and Mayer, 2004, Shane, 2001).

³ In the empirical literature, variables such as the distance from headquarters, the number of states in which the chain operates or the outlet density are used as measures for monitoring costs (or as measures inversely related to monitoring costs like for instance, outlet density) (see Lafontaine and Slade, 1997, 2007, for a survey).

unit, the salaried manager is tempted to shirk (Brickley and Dark, 1987).⁴ In contrast, within a franchised unit, the franchisee shirking is less an issue as he is the residual claimant.⁵ But, the franchisee's opportunism takes the form of under-declaration of actual sales in order to extract rents ("royalty evasion"). Bradach (1997) study of restaurant chains partially supports this assumption by claiming that "*the only financial information that the chain received was a revenue number each month from which the royalty was calculated*" (1997, p. 288). Combs *et al.* (2004) also cite under-declaration of sales as one way for a franchisee to hurt the chain.⁶ One could claim that chains have similar or better information on local markets than prospective franchisees because they can rely on information on sales data provided by their units with comparable attributes and similar local conditions. There are however two elements supporting our assumption of private information on local sales in favor of the local managers. First, Minkler (1992) argued that better knowledge on local markets by franchisees is an important driver for franchising. Second, while the Federal Trade Commission regulation in the US requires that franchisors disclose a lot of information to prospective franchisees, reporting on potential earning (called "earning claims") is optional (Price, 1995). When chains disclose such information, they are very cautious and claim that it should be not taken for granted as each local market has specific features.⁷

Managers' opportunism (either under-declaration of sales or shirking) can be circumvented by monitoring the units. However, we assume that monitoring the actual sales

⁴ While this is not the only contrasting differences between company-owned and franchised units, it is quite often stressed in the literature (see for instance Blair and Lafontaine, 2005, Lafontaine and Slade, 2007). We take this difference for granted in our analysis.

⁵ As owners of their business, franchisees have a claim on the profits generated by their outlet(s) net of the usual sales-based royalties and advertising fees they pay to their franchisors. As these payments normally represent 6 to 10% of revenues, franchisees obtain the bulk of every additional dollar of sales generated within their outlet(s).

⁶ The other hazards respectively are: (i) not adhering to quality standards, and (ii) divulging the chain's proprietary information.

⁷ A standard warning at the beginning of a typical earning claim is as follow: "*The following results of operations at various average daily vehicle count levels should not be considered as the actual or potential income or results of operations of any particular franchise. The franchisor does not represent that the franchisee can expect to attain these volumes*".

of individual outlets is costly and this cost increases with the number of monitored units (or with the intensity of monitoring an individual unit). As extensively shown in the literature on auditing in the context of tax evasion, utility regulation or antitrust enforcement (Baron and Besanko, 1984; Mookherjee and Png, 1989; Khalil, 1997), it may be sub-optimal to monitor all units (or to monitor all of them with the same intensity). Among this body of literature, our model heavily relies on Besanko and Spulber (1989) that looked at the optimal monitoring policy of an antitrust authority that imperfectly observes firms' production costs and pricing behavior. They showed that it might be optimal to tolerate a limited amount of collusion (*i.e.* to adopt a *laissez-faire* policy for slightly harmful price agreements). Applying this framework to franchising, we obtain two main results. First, we show that it is never optimal to systematically monitor all the outlets. Second, chains should monitor only the units that declare or reach sales below a threshold that could differ for franchised and company-owned units. Indeed, it is efficient to tolerate some level of under-declaration in franchised units or shirking in company-owned units that declare high sales in order to concentrate costly monitoring efforts on the units that declare low local demand. These results allow us to discuss how information technology and market characteristics could affect monitoring costs in franchised and company-owned units. Finally, we analyze how our findings on monitoring policies might influence dual distribution.

Our paper is related to the literature on the organization of franchised chains. Most of the papers assume, explicitly or implicitly that chains can freely measure sales (output) in order to calculate the amount of royalties paid by franchisees. In contrast, we explicitly assume that monitoring output is costly (Mathewson and Winter, 1985, Gal-Or, 1995, are exceptions assuming also costly monitoring of outputs).⁸ In addition, only a few papers deal

⁸ From such a perspective, this cost can also be viewed as the expenses chains have to incur in order to make sales or revenues verifiable. Most of the literature on franchising, and more generally on contract theory, distinguishes observable from unverifiable variables by assuming that the former are observed at zero cost,

with private information issue in the franchising context. Mathewson and Winter (1985) proposed a setting wherein managers can take advantage of information asymmetry at the actual level of local demand. They are however primarily concerned with the efficiency of sharing contracts and do not deal with global monitoring strategy of chains. Hempelmann (2006) also analyzed a setting where franchisees have private information about a relevant variable (marginal cost of sale in his analysis). Here again, his paper is about the design of franchise contract. Gal-Or (1995) is more directly interested in monitoring strategy and the “make or buy” issue. She examined how some local markets characteristics can influence the incentives of the chain to monitor its outlets. Like in our model, she found that it is sometimes in the chain's interest to monitor just a subset of outlets. In her paper however, the extent of monitoring is equivalent to the extent of vertical integration since it is assumed that monitoring a unit will allow the chain to dictate the desired level of effort. In our model however, monitoring is not equivalent to ownership. Some units are monitored while remaining franchised and others are company-owned but not monitored.

This paper will proceed as follows. Section 2 will present the theoretical model, while the set of propositions will be displayed and discussed in Section 3. Section 4 provides some managerial implications. Conclusion follows.

2. THE MODEL

2.1 Model framework

The model considers a chain with outlets located in different geographical areas. Each local market is identically characterized by a random fluctuating demand. The positive or

whereas the latter are too costly to observe. We might argue in the real world that verifiability is an agent's decision based on a cost/benefit analysis. Some variables are costly to verify, but this does not necessarily deter parties from contracting on them (Khalil and Lawarée, 1995). Recent literature both in contract theory and in law and economics has begun to explicitly introduce the notion of verifiability as a decision based on the cost and incentives to produce evidence for a third-party enforcer (see for instance Bull and Watson, 2004, Scott and Triantis, 2006).

negative shocks affecting local demand are observed by store managers and not by the chain. Moreover, random shocks on local markets are independent from each others.⁹ Let θ be the demand level experienced by an outlet, where θ is distributed between $\underline{\theta}$ and $\bar{\theta}$ according to a probability distribution function f and a cumulative distribution function F . A negative shock tends to push demand towards $\underline{\theta}$ while a positive shock increases demand towards $\bar{\theta}$.

Each outlet is run by a manager who can either be a franchisee or an employee. All managers are assumed to have the same abilities in running a store. Store performance will depend however on the level of local demand.¹⁰ Since managers are perfectly informed on local market conditions, the chain must thus rely on them for knowing the true level of local demand.

We now define $V(\theta)$ as the expected revenue of a store experiencing a level of demand θ , with $V'(\theta) > 0$ and $V''(\theta) < 0$. However, the actual revenue can be less than $V(\theta)$ because sales also depend on the effort expended by the manager. For the sake of simplicity, this effort e is assumed to take a value between 0 and 1, and the cost of efforts is given by δe . The actual revenue of a store experiencing demand θ and run by a manager deploying effort e is defined by: $eV(\theta)$. A manager who expends no effort will achieve zero sales (regardless of the level of demand), whereas the same manager expending the highest level of effort ($e=1$) will generate the maximum revenue $V(\theta)$.¹¹ Moreover, the current operating costs of a store, net of manager earnings, are constant and equal to C , regardless the store's status.¹² We will

⁹ This assumption rules out the possibility of implementing relative performance incentive schemes by comparing information across local markets. Furthermore, we are not aware of any case where chains explicitly rely on such a scheme.

¹⁰ Independently of the level of demand, the expected revenue of a store should be positively influenced by other variables like the value and reputation of the chain and other franchisor-related behaviors.

¹¹ We will see later on that the chain is always able to enforce an effort level above zero with a salaried manager, ruling out the possibility of actually observing units with no sales.

¹² This is a simplifying assumption. One could argue that, because he is the residual claimant of the unit's profits, a franchisee has stronger incentives to minimize costs than a chain employee. The operating costs of a franchised unit might thus be lower. Qualitative results of our model would not change if this assumption is relaxed.

now define both manager and chain payoffs depending on the status of the outlet.

2.2 Payoffs within a Franchised Unit

For all franchised units, we assume an identical contract involving the payment of sales-based royalties β .¹³ This assumption is consistent with franchisor practices where contract customization is rarely observed (see Bhattacharyya and Lafontaine, 1995; Lafontaine, 1992). Franchisee profit with demand θ is thus given by $(1 - \beta)eV(\theta) - C - \delta e$, while the franchisor receives revenue $\beta eV(\theta)$. Franchisees can potentially display two types of opportunistic behavior: (i) they can shirk and put forth limited effort¹⁴, (ii) they can under-declare the actual value of sales and save on the amount of royalties paid to the chain. This last form of opportunism could be costly to detect since a low declared sales level can result either from an honest franchisee experiencing low θ or from a franchisee experiencing high θ but under-reporting sales. In order to better focus on this under-declaration problem, we rule out franchisee's shirking by assuming that the marginal benefit of additional sales effort is always higher than the marginal cost. The franchisee will thus always expend the maximum level of effort ($e=1$):

$$\text{Assumption 1: } (1 - \beta)V(\theta) > \delta \tag{1}$$

It is not a strong assumption because as a residual claimant, franchisees have strong incentives to make efforts.¹⁵ We will further define the cost of monitoring franchisees' sales and discuss how franchisees react to monitoring.

¹³ For the sake of simplicity, we have assumed that the franchisor does not claim initial fees from his franchisees. Our results do not depend on this simplification.

¹⁴ The severity of this moral hazard problem increases with royalty rates (Lafontaine, 1992).

¹⁵ If we relax assumption 1, the chain should provide incentives both to mitigate shirking and under-declaration; moreover, it will significantly complicate the model without adding any insight. One way to reinforce assumption 1 is to suppose that franchisees' efforts also reduce the cost of sales, thereby lowering franchisee incentive to shirk.

2.3 Payoffs within a company-owned unit

How are managers in company-owned units paid in real life? Most, if not all the literature on franchising assumes that they are paid a fixed wage. In their recent survey on the boundaries of the firm, Lafontaine and Slade (2007) equate vertical integration in franchising with a low-powered incentive contract where the agent receives a fixed wages. This is probably driven by the fact that previous empirical studies of franchised chains found a few reliance on explicit “pay for performance” contract. For instance, in a detailed study of the structure of compensation in the US fast-food industry, Krueger (1991) found that the long-standing practice in this industry has been to pay managers of company-owned units a fixed annual salary not directly related to the unit’s profitability. Similarly, Bradach (1998) found that managers of company-owned units in the fast-food industry were mainly compensated on the basis of a fixed salary, with sometimes a small bonus component. The main incentive instrument he described is a promotion-based system where “good” managers move up to higher positions in the organization.¹⁶ In addition, bonuses were largely based on meeting some uniformity standards instead of being based on financial performances (see Bradach, 1998, p. 37-38, for a more detailed description). We follow the existing literature and assume that the chain only offers a fixed wage w . For a demand level θ and manager effort e , the chain's expected profit is $eV(\theta) - w - C > 0$, and the manager's utility is $w - \delta e$. We also assume that all outlets are profitable regardless of demand level, provided the manager is exerting maximum effort.

$$\text{Assumption 2: } V(\underline{\theta}) - w - C > 0 \tag{2}$$

With a fixed compensation scheme, salaried managers have strong incentives to shirk. The chain is not able to directly observe the manager's efforts. However, it can use the sales

¹⁶ See Brown Jr. (1998) for an analysis of chains governance based on the implementation of promotion-based incentive schemes.

or revenues from its outlets to infer the effort level. As an example, if sales below $V(\underline{\theta})$ are observed, the chain knows with certainty that the manager has shirked. The chain can therefore always enforce a minimum, strictly positive, level of effort. For a demand θ it is clearly in the interest of the salaried manager to select an effort at least equal to $e_{\min}(\theta) = \frac{V(\underline{\theta})}{V(\theta)}$ such that sales exceed $V(\underline{\theta})$.¹⁷ In accordance with Assumption 2, the franchisor is guaranteed to obtain positive profits with this minimum sales level.

2.4 Monitoring policy and opportunistic behavior

To overcome opportunistic behavior stemming from salaried managers or franchisees, the chain may proceed by monitoring them (e.g. auditing their accounts and/or conducting visits). Our understanding of monitoring encompasses all activities to accurately measure sales and assess behavior. Monitoring technology and monitoring cost have no reason to be the same for franchised and company-owned units, because opportunistic behavior in the two setting is of a different type. As stressed before, the main problem with company-owned units is shirking. Without monitoring, the only information on the manager effort is the actual sales. Chains can complement this information by direct supervision for instance by sending a company representative in order to obtain indirect information on the level of effort exerted by the employee, like information concerning quality of service, etc. This is a form of what Lafontaine and Slade (2007) called “behavioral monitoring” (or input monitoring). In franchised units, the main problem in our model is under-declaration of actual sales. Here again, without monitoring, the only information on effort available to the chain is the declared sales. Monitoring in this situation corresponds to devices used to discover the true value of sales. For instance, chains may audit the account. Lafontaine and Slade (2007) called this “output” monitoring. We denote K as the cost of auditing a company-owned store and Q as

¹⁷ Without monitoring, the optimal effort for the salaried managers is clearly e_{\min} .

the monitoring cost of a franchised store. We assume that in both cases, monitoring technologies are costly but "perfect", *i.e.* an audit enables the franchisor to discover the true level of demand and consequently to know whether the manager has been cheating (see Gal-Or, 1995, for a similar assumption).

In order to determine how the chain can efficiently monitor its managers, let us start by noting that it is always better to design a monitoring policy that takes advantage of available information rather than opting for a myopic or random policy. In the present case, the existing information is franchisees' declared sales and observed sales from company-owned units. Through such sales, the manager is more or less announcing the level of demand. When a franchisee experiencing demand θ declares a revenue R to the franchisor, this is therefore equivalent to announcing a demand θ_R with $\theta_R = V^{-1}(R)$, where $V^{-1}(\cdot)$ is the inverse revenue function.¹⁸ If the franchisee is behaving honestly, then $\theta_R = \theta$ whereas if he is behaving opportunistically (*i.e.* $R < V(\theta)$), then $\theta_R < \theta$. In any event, the franchisor's revenues are given by: $\beta V(\theta_R)$. Likewise, when a salaried manager experiencing demand θ yields revenue R , it is equivalent to announcing demand θ_E , with $\theta_E = V^{-1}(R)$. If the salaried manager is behaving honestly ($e=1$), this means $\theta_E = \theta$.

The optimal monitoring policy will thus be conditional upon the observed level of sales θ_R and θ_E . Let $\chi(\theta_R)$ denotes the probability of monitoring a franchisee who declares sales of $V(\theta_R)$ (*i.e.* who announces a demand of θ_R) and let $\varpi(\theta_E)$ represent the probability of monitoring a salaried manager who generates sales of $V(\theta_E)$ (*i.e.* who announces a demand of θ_E). If $\chi(\theta_R)=1$, then the franchisee knows with certainty that he will be monitored and if $\chi(\theta_R)=0$, no monitoring will occur. In the intermediate case, the franchisee

¹⁸ This is because the franchisor knows that the franchisee's maximum effort is ($e=1$).

will be inspected with probability or frequency $\chi(\theta_r)$.¹⁹ The same applies for $\varpi(\theta_e)$.

The monitoring policy must also specify punishment schemes. In company-owned units, if the chain detects an effort below 1, it imposes a penalty on the salaried manager. These punishments can take different forms, ranging from a fine to dismissal. The termination of a labor contract, for example, represents a credible punishment if the manager receives an “efficiency wage” (Shapiro and Stiglitz, 1984) or if the manager has invested in specific human capital, whether partly or entirely financed by him.²⁰ Most of the time the punishment corresponds to an opportunity cost for the salaried manager (a bad reputation, less perspective of inside or outside promotions,...), rather than a monetary fine.²¹ Similarly, if the franchiser detects under-declaration of revenues, he penalizes the franchisee either by terminating the contract or renegotiating the contract or introducing a fine.²²

We consider that the penalty inflicted to the franchisee or the salaried manager is bounded and corresponds to a kind of (exogenously given) maximum liability. Following Becker’s seminal contribution on crimes and punishment (Becker, 1968), if managers are risk neutral, it is efficient to set the penalty at the maximum level in order to save on monitoring costs.²³ We denote a (respectively b) as the franchisee’s maximum liability (respectively as the salaried manager’ maximum liability). Therefore, the optimal penalties are a for a cheating franchisee and b for an opportunistic salaried managers. Moreover, we assume that:

$$\text{Assumption 3: } a \geq \beta[V(\bar{\theta}) - V(\underline{\theta})] \quad (3)$$

¹⁹ An alternative interpretation would be to consider that over period T, the franchisee receives $\chi(\theta_r)$ T visits.

²⁰ See Krueger (1991) for empirical evidence of the efficiency wage for employees within company-owned units in the fast food industry.

²¹ Thus, we consider that penalty is not a pure transfer from the salaried manager to the chain and has no direct effect on the chain’s profit. In other terms, the money that the chain could expect by suing the opportunistic manager is assumed to compensate the litigation costs.

²² See Klein (1995) and Lafontaine and Raynaud (2002) for insights on self-enforcement in franchising where contract termination (coupled with the prospect to earn expected ex post rents) acts as an incentive device.

²³ If the managers are risk averse, increasing the penalty to its maximum level would not usually be optimal (Polinski and Shavell, 1979).

$$\text{Assumption 4: } b \geq \frac{\delta(V(\bar{\theta}) - V(\underline{\theta}))}{V(\bar{\theta})} \quad (4)$$

Assumptions 3 and 4 indicate that if a manager is certain to be monitored ($\chi(\theta_R) = 1$ and $\varpi(\theta_E) = 1$), the penalties (a and b) are sufficiently high to deter cheating, regardless of the level of demand. The left-hand side represents the cost of being punished and the right-hand side the highest benefit gained from cheating (*i.e.* the gain derived by a manager experiencing the most favorable demand conditions while declaring the lowest demand).²⁴ Under such assumptions, a chain has always the possibility to prevent every manager from behaving opportunistically. A natural question then is whether or not it is optimal to systematically monitor all the units.

Figure 1 displays the timing sequence. In Step 1, the chain announces its monitoring policy to its franchised and company-owned units. During Step 2 and depending on the monitoring policy, the franchisee declares a sales level $V(\theta_R)$, and the salaried manager chooses a level of effort that induces sales of $V(\theta_E)$. In the last step, the chain implements its monitoring program.

²⁴ When a salaried manager is certain to be monitored, he will be deterred from shirking if $w - \delta \geq w - \delta e_{\min}(\theta) - b$. This condition is equivalent to $b \geq \delta(1 - e_{\min}(\theta)) = \delta \left(\frac{V(\bar{\theta}) - V(\underline{\theta})}{V(\bar{\theta})} \right) = b_{\min}(\theta)$ with $b_{\min}(\theta)$ increasing with θ . Assumption 4 ensures that this condition holds for any level of demand.

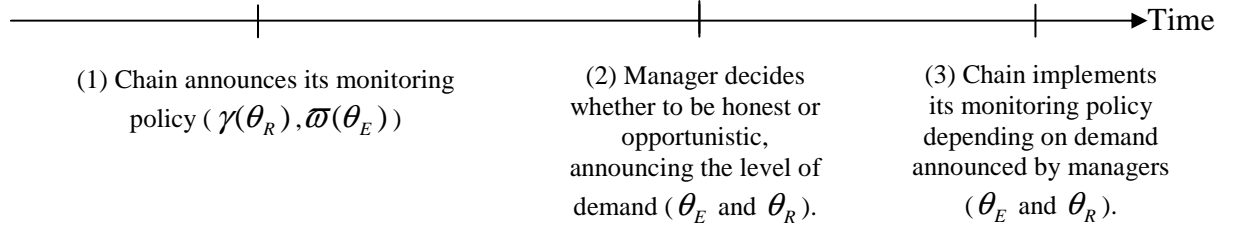


Figure 1: The timing of decisions

3. THE EXTENT OF MONITORING

In this section, we solve the strategic game and determine the optimal monitoring policy under different organizational forms. As a benchmark, we analyze the full deterrence case and then consider the optimal monitoring policy in franchise and company-owned units.

3.1 Full deterrence

A franchised unit

Let us first consider the case of a franchised unit. If the chain is seeking to prevent a franchisee experiencing demand θ from declaring $\theta_R < \theta$, then he must announce a monitoring probability $\chi(\theta_R)$, such that:

$$(1 - \beta)V(\theta) - C - \delta \geq V(\theta) - \beta V(\theta_R) - C - \delta - \chi(\theta_R)a \quad (5)$$

Equation (5) means that the expected payoff of a franchisee who truthfully reports his sales must be higher than his payoff when misreporting. By misreporting, he saves on royalties yet runs the risk of being detected and punished by the chain ($\chi(\theta_R)a$). After rearrangement, $\chi(\theta_R)$ should be at least equal to:

$$\chi(\theta_R) \geq \frac{\beta(V(\theta) - V(\theta_R))}{a} \quad (6)$$

If the chain wishes to deter a franchisee from behaving opportunistically whatever θ , then she must announce a monitoring policy defined for any θ_R by:

$$\gamma(\theta_R) = \frac{\beta(V(\bar{\theta}) - V(\theta_R))}{a} < 1 \quad (7)$$

Since market with local demand $\bar{\theta}$ provides the highest incentive to under-declare, such monitoring intensity is sufficient to deter any franchisee from cheating. For any θ , the franchisee will prefer to be honest and pay the required royalties than to under-declare and be exposed to a penalty (even if he is not certain of being monitored).

A company-owned unit

If the chain is seeking to deter a salaried manager with a local demand of θ from cheating, his probability of supervision $\varpi(\theta_E)$ must be defined by:

$$w - \delta \geq w - \delta e - \varpi(\theta_E)b \quad \text{with}^{25} \quad e = \frac{V(\theta_E)}{V(\theta)} < 1 \quad (8)$$

After rearrangement, the intensity of monitoring $\varpi(\theta_E)$ should be at least equal to:

$$\varpi(\theta_E) \geq \frac{\delta(V(\theta) - V(\theta_E))}{bV(\theta)} \quad (9)$$

If the chain is seeking to deter a salaried manager from shirking for any θ , then she must implement a monitoring policy $\varpi(\theta_E)$, characterized for any θ_E by:

$$\varpi(\theta_E) = \frac{\delta(V(\bar{\theta}) - V(\theta_E))}{bV(\theta)} < 1 \quad (10)$$

With such monitoring intensity, a salaried manager experiencing demand $\bar{\theta}$ is deterred, as are all the salaried managers facing lower level of demand. We now examine

²⁵ Exerting an effort $e = \frac{V(\theta_E)}{V(\theta)}$ leads to a level of sales $eV(\theta) = V(\theta_E)$ instead of $V(\theta)$.

whether the full deterrence monitoring policy (*i.e.* monitoring every managers systematically whatever the level of sales they declare) is optimal. We will first consider the case of company-owned units.

3.2 Company-Owned Units and Partial Monitoring

What should the monitoring policy look like when chains suffer from asymmetric information on both agents' behavior and market characteristics? In their analysis of optimal antitrust policy, Besanko and Spulber (1989) have shown that it should take the form of a threshold monitoring program. While we deal with a different issue in this paper, the information problem faced by the antitrust agency and the franchised chain in our model is very similar. Because of limited resources and costly monitoring, the antitrust agency will not monitor all industries and must select some of them. The problem is more complicated with asymmetric information since it will be difficult to distinguish situations with high price due to high marginal costs from situations with high price due to collusive behavior. In our model, when a unit announces a low level of demand (and thus sales), the chain does not know whether this is due to bad luck or to opportunistic behavior. Following Besanko and Spulber (1989), the threshold monitoring program could be defined by:

$$\text{if } \begin{array}{l} \theta_E > \hat{\theta}, \text{ then} \\ \underline{\theta} \leq \theta_E \leq \hat{\theta}, \text{ then} \end{array} \quad \begin{array}{l} \varpi(\theta_E) = 0 \\ \varpi(\theta_E) = \delta \frac{(V(\hat{\theta}) - V(\theta_E))}{bV(\hat{\theta})} \end{array} \quad (11)$$

When the announced level of demand is below the threshold $\hat{\theta}$, then the chain will monitor the unit with a certain probability, whereas no monitoring is implemented for a level of demand above $\hat{\theta}$. Complete monitoring would imply setting $\hat{\theta} = \bar{\theta}$. If the chain adopts a policy of partial monitoring, the threshold will be characterized by $\hat{\theta} < \bar{\theta}$. In the latter case, the chain concentrates its resources on monitoring managers with sales below $V(\hat{\theta})$.

Moreover, the intensity of monitoring tends to decrease with the unit's announced sales.

Given the monitoring policy defined in (11), managers experiencing a demand $\theta \leq \hat{\theta}$ are deterred from shirking and will always choose $\theta_E = \theta$ (*i.e.* an effort $e=1$). Managers

experiencing $\theta > \hat{\theta}$ know that they will never be monitored as long as they are achieving at least $V(\hat{\theta})$. They therefore choose effort level $e = \frac{V(\hat{\theta})}{V(\theta)}$, which enables reaching sales

amount $V(\hat{\theta})$. These managers can never expect a higher payoff by exerting an effort below

$e = \frac{V(\hat{\theta})}{V(\theta)}$, since: $w - \delta \frac{V(\theta_E)}{V(\theta)} - \left(\frac{\delta(V(\hat{\theta}) - V(\theta_E))}{bV(\hat{\theta})} \right) b < w - \delta \frac{V(\hat{\theta})}{V(\theta)}$, where the expression

on the left-hand side represents his expected benefit from cheating in choosing a level of

effort of $\theta_E < \hat{\theta}$ and the expression on the right-hand side the expected benefit if he chooses

$\theta_E = \hat{\theta}$.

With partial monitoring, the chain can be considered as offering a rent to managers with $\theta > \hat{\theta}$. The average profit per company-owned unit is then defined by:

$$G^{co} = \int_{\underline{\theta}}^{\hat{\theta}} (V(\theta) - w - C - \varpi(\theta)K) f(\theta) d\theta + \int_{\hat{\theta}}^{\bar{\theta}} (V(\hat{\theta}) - w - C) f(\theta) d\theta$$

For any demand θ between $\underline{\theta}$ and $\hat{\theta}$, the chain monitors its units with a certain probability to prevent the managers from shirking and the profit is $V(\theta) - w - C$, less the

monitoring costs $\varpi(\theta)K$. For demand θ higher than $\hat{\theta}$, the chain prefers not monitoring its

units and accepts that the managers shirk as long as they reach the revenues $V(\hat{\theta})$. Hence,

they can save on monitoring costs. The question is to determine what the optimal threshold of

the monitoring program is. The following proposition demonstrates that a partial monitoring

strategy is indeed optimal.

Proposition 1: *It is always optimal to tolerate some shirking in company-owned units, i.e.*

$$\hat{\theta} < \bar{\theta}.$$

Proof: Optimal monitoring policy is determined by maximizing the chain's average profit per company-owned unit with respect to the monitoring threshold $\hat{\theta}$:

$$G^{co} = \int_{\underline{\theta}}^{\hat{\theta}} (V(\theta) - w - C - \varpi(\theta)K) f(\theta) d\theta + \int_{\hat{\theta}}^{\bar{\theta}} (V(\hat{\theta}) - w - C) f(\theta) d\theta \quad (12)$$

The derivative with respect to $\hat{\theta}$ is given by:

$$\frac{\partial G^{co}}{\partial \hat{\theta}} = [V(\hat{\theta}) - w - C - \varpi(\hat{\theta})K - V(\hat{\theta}) + w + C] f(\hat{\theta}) + V'(\hat{\theta}) \int_{\hat{\theta}}^{\bar{\theta}} f(\theta) d\theta + \frac{\partial \varpi}{\partial \hat{\theta}} K \int_{\underline{\theta}}^{\hat{\theta}} f(\theta) d\theta$$

After rearrangement:

$$\frac{\partial G^{co}}{\partial \hat{\theta}} = - \int_{\underline{\theta}}^{\hat{\theta}} \frac{\delta V'(\hat{\theta}) V(\theta) K}{b(V(\hat{\theta}))^2} f(\theta) d\theta + V'(\hat{\theta}) \int_{\hat{\theta}}^{\bar{\theta}} f(\theta) d\theta \quad (13)$$

The first expression has a negative sign and corresponds to a *monitoring cost effect*: a higher $\hat{\theta}$ increases the cost of supervising managers since the number of monitored units (or the probability of monitoring units) rises. The second expression has a positive sign and represents an *incentive effect* since a rise in monitoring reduce moral hazard and thus increases sales as well as profits. Monitoring all units entails setting $\hat{\theta} = \bar{\theta}$. Such a policy however is not optimal since:

$$\left. \frac{\partial G^{co}}{\partial \hat{\theta}} \right|_{\hat{\theta}=\bar{\theta}} = - \int_{\underline{\theta}}^{\bar{\theta}} \frac{\delta V'(\bar{\theta}) V(\theta) K}{b(V(\bar{\theta}))^2} f(\theta) d\theta < 0 \quad (14)$$

From (14), we conclude that the chain can raise profits by relaxing supervision on the most efficient units: from a level of $\hat{\theta} = \bar{\theta}$, reducing $\hat{\theta}$ by a slight amount enables the chain to significantly lower its monitoring costs (first-order effect), without considerably weakening the incentive for managers to behave honestly (second-order effect).

Moreover, we find that a *laissez-faire* policy ($\hat{\theta} = \underline{\theta}$) is not optimal:

$$\left. \frac{\partial G^{co}}{\partial \hat{\theta}} \right|_{\hat{\theta}=\underline{\theta}} = V'(\underline{\theta}) > 0 \quad (15)$$

In the neighborhood of $\underline{\theta}$, by increasing $\hat{\theta}$, the chain can stimulate efforts from all its managers (first-order effect), with the expected benefits largely compensating for the additional monitoring cost. ■

Proposition 1 can be explained as follows. Partial deterrence allows for a significant reduction in monitoring and incentives costs. Even if the revenues generated by the chain are less than what would be expected in stores experiencing high demand, this shortfall is more than counterbalanced by the savings on monitoring costs. Partial deterrence has two meanings. First, the chain only monitors a sub-sample of its units. High-level sales units are never monitored. Secondly, for lower levels of demand, monitoring is implemented with a certain probability. At the equilibrium, the chain monitors only the honest managers and tolerates some misreporting.

We can also reinterpret partial monitoring policy in a dynamic setting. If a company-owned unit experiences at each period a level of demand that is independent of the level in the previous periods, then the chain is likely to monitor this unit only during the periods of low demand ($\theta \leq \hat{\theta}$) and adopt a *laissez-faire* policy during the periods of boom ($\theta > \hat{\theta}$). Then, the salaried managers tend to reduce their efforts moderately when demand is favorable (just to reach the required level of sales $V(\hat{\theta})$). The sales performance distribution of each company-owned store thus should be upward bounded, with several units displaying the same level of sales ($V(\hat{\theta})$).

It should be noted that the "informational rent" captured by salaried managers

$\left(\delta - \delta \frac{V(\hat{\theta})}{V(\theta)}\right)$ is an increasing function of monitoring costs and a decreasing function of penalty severity (as $\hat{\theta}$ decreases in K and increases in b). The range of realized demand for which the chain tolerates some shirking tends to be reduced as monitoring costs decline towards zero or when punishments become extremely severe. Yet by raising the penalties, two opposite effects are introduced on monitoring intensity at the individual level. A unit subject to a negative demand shock is now less intensively monitored²⁶, whereas the same unit submitted to a positive shock is more likely to be monitored (even if the probability is infinitesimal in presence of high demand). Even under these extreme conditions (severe punishments, monitoring costs close to zero) however, it is still efficient for the chain to renounce auditing the most profitable units.

3.3 Franchised Units and Partial Monitoring

As previously, the optimal monitoring policy for the franchised units is a threshold program that can be defined as follows:

$$\begin{array}{ll} \theta_R > \hat{\theta}, & \text{then } \gamma(\theta_R) = 0 \\ \text{if } \underline{\theta} \leq \theta_R \leq \hat{\theta}, & \text{then } \gamma(\theta_R) = \beta \frac{\left(V(\hat{\theta}) - V(\theta_R)\right)}{a} \end{array} \quad (16)$$

Complete monitoring entails setting $\hat{\theta} = \bar{\theta}$, whereas when $\hat{\theta} < \bar{\theta}$, the chain only monitors a subset of units and "offers" rent to those franchisees experiencing high demand. From a formal standpoint, in the case of $\hat{\theta} < \bar{\theta}$, all franchisees experiencing demand $\theta > \hat{\theta}$ are incited to under-declare their revenues (by reporting $V(\hat{\theta})$ instead of $V(\theta)$). As long as the royalties paid remain greater than or equal to $\beta V(\hat{\theta})$, they are not monitored. It is therefore

²⁶ Since the manager behavior is only influenced by the expected penalty, a higher penalty enables reducing the probability of audit.

profitable for them to declare revenues equal to $V(\hat{\theta})$.²⁷ Conversely, all franchisees exposed to demand $\theta \leq \hat{\theta}$ are likely to be monitored and are thus incited to truthfully report their revenues. Given the announced monitoring policy, the franchiser's average profit per franchised unit is defined by:

$$G^f = \int_{\underline{\theta}}^{\hat{\theta}} (\beta V(\theta) - \gamma(\theta)Q) f(\theta) d\theta + \int_{\hat{\theta}}^{\bar{\theta}} \beta V(\hat{\theta}) f(\theta) d\theta$$

The following proposition is the “mirror image” of proposition 1 and asserts that complete monitoring is not optimal.

Proposition 2: *It is always optimal to tolerate some under-declaring in franchised units, i.e.*

$$\hat{\theta} < \bar{\theta}.$$

Proof: The optimal threshold $\hat{\theta}$ is given by maximizing the franchiser's average profit per franchised unit with respect to managerial threshold monitoring $\hat{\theta}$:

$$\text{Max}_{\{\hat{\theta}\}} G^f = \int_{\underline{\theta}}^{\hat{\theta}} (\beta V(\theta) - \gamma(\theta)Q) f(\theta) d\theta + \int_{\hat{\theta}}^{\bar{\theta}} \beta V(\hat{\theta}) f(\theta) d\theta \quad (17)$$

Given $\gamma(\theta) = \frac{\beta(V(\theta) - V(\hat{\theta}))}{a}$, the derivative of the profit with respect to $\hat{\theta}$ is given by:

$$\frac{\partial G^f}{\partial \hat{\theta}} = -\beta \frac{V'(\hat{\theta})}{a} Q \int_{\underline{\theta}}^{\hat{\theta}} f(\theta) d\theta + \beta V'(\hat{\theta}) \int_{\hat{\theta}}^{\bar{\theta}} f(\theta) d\theta \quad (18)$$

²⁷ For franchisees of the type $\theta > \hat{\theta}$, the benefit from declaring revenues $V(\theta_R)$ below $V(\hat{\theta})$ is given by:

$$V(\theta) - \beta V(\theta_R) - C - \delta - \left(\frac{\beta(V(\hat{\theta}) - V(\theta_R))}{a} \right) = V(\theta) - \beta V(\hat{\theta}) - C - \delta$$

Hence, they are perfectly indifferent between reporting $V(\hat{\theta})$ and any lower revenue.

The first expression represents the negative *monitoring cost effect* if the chain decides to tighten monitoring of its franchisees (to increase $\hat{\theta}$). The second expression corresponds to the *incentive effect* (more intensive monitoring induces franchisees to declare their actual sales). Complete monitoring consists of setting $\hat{\theta} = \bar{\theta}$. Such a policy however is not optimal since the franchisor is able to increase profit by relaxing monitoring efforts on the most efficient units. Marginal profit is found to decrease with $\hat{\theta}$ at $\hat{\theta} = \bar{\theta}$:

$$\left. \frac{\partial G^f}{\partial \hat{\theta}} \right|_{\hat{\theta} = \bar{\theta}} = -\beta \frac{V'(\bar{\theta})}{a} Q < 0 \quad (19)$$

We can also notice that a *laissez-faire* policy $\hat{\theta} = \underline{\theta}$ is sub-optimal:

$$\left. \frac{\partial G^f}{\partial \hat{\theta}} \right|_{\hat{\theta} = \underline{\theta}} = \gamma(\underline{\theta}) Q f(\underline{\theta}) + \beta V'(\underline{\theta}) > 0 \quad (20)$$

■

The rationale for partial monitoring in franchised units is the same as in company-owned units. Because of monitoring costs, chains refrain from auditing the most profitable franchised units. A franchisee experiencing demand $\theta > \hat{\theta}$ will thus benefit from an "informational rent" equal to $\beta(V(\theta) - V(\hat{\theta}))$.²⁸ Offering a rent to the higher-earning franchisees represents a cost-saving method for the franchisor to deter franchisees from engaging in more harmful cheating (a more detrimental action would be to declare the lowest level of sales $V(\underline{\theta})$).

Corollary: The level of sales $\hat{\theta}$ above which the franchisee is never audited is

²⁸ The greater the franchisee's demand, the higher the rent he may receive.

determined by $\hat{\theta} = F^{-1}\left(\frac{1}{1+Q/a}\right)$

Proof: It is directly obtained from the derivative of the profit with respect to $\hat{\theta}$:

$$\frac{\partial G^f}{\partial \hat{\theta}} = -\beta V'(\hat{\theta}) \frac{Q}{a} \int_{\underline{\theta}}^{\hat{\theta}} f(\theta) d\theta + \beta V'(\hat{\theta}) \int_{\hat{\theta}}^{\bar{\theta}} f(\theta) d\theta = 0$$

■

The optimal threshold $\hat{\theta}$ that measures the extent of under-declaring deterrence tends to increase when penalty is high or monitoring costs are low. But interestingly, there is no correlation between royalty rate and the extent of under-declaring. Modifying royalty rate has no impact on the threshold $\hat{\theta}$ (i.e. on the probability that a franchisee is given the opportunity to under-declare its sales). However, if the chain raises royalty rate, she has to increase her monitoring efforts on the franchisees that announces sales below $V(\hat{\theta})$. Indeed, the franchisee incentives to cheat are higher when royalties claimed by the chain increase and the latter has to announce a higher probability of being monitored to deter under-declaring. To sum-up, raising royalty rates does not increase the intensity of monitoring for the franchisees that face favorable demand conditions. Only those who are in less favorable demand situations will be more audited.

4 TESTABLE HYPOTHESIS AND MANAGERIAL IMPLICATIONS

While our objective in this paper is not primarily to derive a set of prescriptions for managers, we can however identify two categories of managerial implications: (1) the

consequences of our framework for the optimal governance of the chain, namely the optimal mix of franchised and company-owned units (dual distribution), (2) the implications for the design of the franchise contract.

First, consider the governance implication. In our model, we took for granted the coexistence of franchised and company-owned units within the same chain. It has long been argued that moral hazard and related monitoring costs are important drivers of the "make or buy" decision (Brickley and Dark, 1987, Lafontaine and Slade, 2007). Suppose that a chain wants to open a unit in a geographical market. Given the optimal monitoring policy of company-owned and franchised units, then the chain can compare the expected profit of running the unit under a franchise (G^f) or by the means of salaried manager (G^{co}). If the chain concludes that $G^f > G^{co}$ (respectively that $G^{co} > G^f$) on this market, then the unit should be run by a franchisee (respectively vertically integrated). If we use our framework to analyze the organization of chain can we explain the prevalence of dual distribution and the proportion of company-owned (respectively franchised) units? The rationale of dual distribution is market heterogeneity in monitoring costs. In some markets, it is easier to monitor franchised units than company-owned units ($Q < K$) and in other markets, it is the opposite ($Q > K$). Indeed, monitoring is more focus on efforts in company-owned units (input monitoring) whereas it is more focus on sales in franchised units (output monitoring). As suggested by the empirical literature (Brickley and Dark, 1987, Lafontaine and Slade, 2007), a key determinant for input monitoring is the distance between the chain's headquarter and the unit because monitoring costs tend to increase with distance, proportionally more for company-owned units than for franchised units. In other terms, monitoring sales (output) is less distance-sensitive. Then it is likely to have $G^{co} > G^f$ for units close to the headquarter (company-owned units are more profitable) and $G^f > G^{co}$ for distant units (franchising is optimal).

The cost of monitoring sales can also depend on the activity sector and the proportion of regular customers. Indeed, the cost of monitoring sales may be close to zero for many types of activities (product sales) yet remain high for other activities (services), for which sales monitoring is more complex, e.g. in a fast-food franchise, the franchisee can declare that a certain number of hamburgers were discarded when, in reality, they were sold. This may explain why the proportion of company-owned units can vary among activity sectors. Similarly, it is easier to predict or infer demand on market with loyal customers, than on markets with irregular customers.

In addition, a chain will rely more intensively on franchising when it possesses more advanced information technologies because such technologies are more adapted to monitor sales (output) than effort (input). Conversely, the proportion of company-owned units should decrease with the monitoring costs of salaried managers and increase with the severity of punishments that the chain is able to inflict upon opportunistic employees.

Finally, if manager effort is not critical to increasing sales, then concern over the shirking hazard becomes less important, and we should observe more vertical integration in order to concentrate on mitigating the under-declaration hazard. This premise would seem to be consistent with empirical results. For instance, Lafontaine (1992) found that the proportion of franchised units tends to rise when the franchisee's effort counts; conversely, the extent of vertical integration is more prevalent for chains where franchisee efforts are less crucial. Similarly, if the prospect of sales under-declaration is smaller than the shirking problem or is more easily controlled, we should then observe greater reliance on franchising.

Monitoring intensity may also depend on other contractual provisions. For instance, it has been argued in the literature that specific provisions such as tying may lower monitoring costs or improve monitoring technology (Klein and Saft, 1985; Michael, 2000).²⁹ This feature

²⁹ In tying provisions, the franchisees must buy another good(s) in addition to the franchise license. For instance,

can be illustrated in the case where the chain provides one or several inputs to franchisees. If the production function at the store is of a fixed proportion, then the chain is able to infer the actual level of sales from the quantity of inputs sold to a particular outlet (*i.e.* the chain can infer the true level of output and avoid "excessive" discrepancies between output and declared sales). Since the discretionary behavior of franchisees is limited, we would expect the monitoring intensity of franchised units to be lower with greater reliance on franchising in a chain with such a tying provision.

Finally, a few chains include in the franchise contract a provision on minimum royalty payment required (see the table on page 765 in Bhattacharyya and Lafontaine, 1995). For sales below a given level S , the franchisee pays a fixed amount and for sales above S , the franchisee pays a value of β *sales. Is this an alternative to monitoring policy to curb the under-reporting problem we are referring to in the paper? This instrument provides an effective lower bound to under-declaration because even if a franchisee declares a level of sales below S , it will pay the same fixed amount. It partly solves the severity of under-declaration but not its probability of occurrence. Franchisees with high sales will still have an incentive to declare S as the realized sales.

5. CONCLUSION

We study in this paper the optimal monitoring policy of a chain under conditions of asymmetric information about local demands and managers effort. While a salaried manager of a company-owned unit can shirk, the franchisee opportunism takes the form of under-declaration of sales. Monitoring local units may mitigate these contractual hazards but chains must incur monitoring costs. We show two main results: first, it is not optimal to monitor all

McDonald's might require its franchisees to buy specific inputs from him or from a designated third party (see Blair and Lafontaine, 2005, chap. 6 for more empirical evidences on tying provision).

the outlets; second, the chain should concentrate monitoring efforts on the units declaring low level of demand (and thus sales). Even if franchisees might be tempted to under-declare their actual sales, this would not necessarily imply that they all systematically will. Their decision will depend on the risk of being detected and punished. Some may decide to cheat (if they estimate that the gain is higher than the expected cost) while others may behave honestly. Even if all franchisees are tempted to cheat, revenue chiseling at equilibrium will be minimal as long as monitoring costs remain low.

Our propositions may be related to the discussion on optimal monitoring policy within organizations in general and retail chains in particular. As stated in the introduction, most papers on franchising typically assume that sales or output observation is costless, which incites chains to link franchisee compensation with sales. This constitutes a form of output monitoring. In company-owned units, chain managers are provided fixed wages and directly supervised by the chain. Chains are thus relying on input (or behavior-based) monitoring within their own units. The prevalence of dual distribution implies that it is efficient to exercise both input and output monitoring at the chain level even if individual units will not be monitored at both the input and output levels. Input and output monitoring thereby coexist in a plural chain even if each of these devices is "specialized" for a particular type of individual unit. Those units whose manager is paid with a fixed compensation are input-monitored, whereas those offering variable compensation are output-monitored. This set-up is consistent with the way chains are organized and with previous results by Khalil and Lawarée (1995) on the choice between input and output monitoring in a principal-agent context. In their paper, they found that when the principal (the chain in this instance) is the residual claimant (owns the outlet), he generally prefers input monitoring, but output monitoring would be preferred when the agent (here, the franchisee) is the residual claimant (owns the outlet).

Lastly, our findings can be understood from a more general perspective: an organization experiencing the kind of contractual hazard we have focused on herein is always more efficient by mixing the incentive schemes made available to its members or agents, *i.e.* by combining fixed and revenue-based remuneration. In other words, the policy of generalizing variable remuneration within an organization might be as sub-optimal as relying exclusively on fixed remuneration. If we were to assimilate variable remuneration with external procurement and fixed remuneration with internal procurement, our results could also be used as another rationale for the prevalence of what is sometimes called "tapered integration" in vertical relationships, *i.e.* the simultaneous use of external input suppliers and in-house suppliers.

REFERENCES

- Azoulay, P., S. Shane. (2001) Entrepreneurs, Contracts, and the Failure of Young Firms, *Management Science*, 47(3), pp. 337-358.
- Baron, D.P., D. Besanko. (1984) Regulation, Asymmetric Information and Auditing, *RAND Journal of Economics*, 15, pp. 447-470.
- Besanko, D., D. Spulber. (1989) Antitrust Enforcement under Asymmetric Information, *Economic Journal*, 99, pp. 408-425.
- Bhattacharyya, S., F. Lafontaine. (1995) Double-Sided Moral Hazard and the Nature of Share Contracts, *RAND Journal of Economics*, 26(4), pp. 761-781.
- Blair, Roger D., F. Lafontaine F. (2005) *The Economics of Franchising*, Cambridge University Press.
- Bradach, J.L. (1998) *Franchise Organization*, Harvard Business School Press.
- Bradach, J.L. (1997) Using the Plural Form in the Management of Restaurant Chains, *Administrative Science Quarterly*, 42, pp. 276-303.
- Brickley, J., F.H. Dark. (1987) The Choice of Organizational Form: The Case of Franchising, *Journal of Financial Economics*, 18, pp. 401-420.
- Brown Jr., W.O. (1998) Transaction Costs, Corporate Hierarchies, and the Theory of Franchising, *Journal of Economic Behavior and Organization*, 36, pp. 319-329.
- Bull, J., J. Watson. (2004) Evidence Disclosure and Verifiability, *Journal of Economic Theory*, pp. 118, 1-31.
- Combs, J.G., S.C. Michael, G.J. Castrogiovanni. (2004) Franchising: A Review and Avenues to Greater Theoretical Diversity, *Journal of Management*, 30(6), pp. 907-931.
- Gal-Or, E. (1995) Maintaining Quality Standard in Franchised Chains, *Management Science*, 41(11), pp. 1774-1792.

- Hemplemann, B. (2006) Optimal Franchise Contracts with Private Cost Information. *International Journal of Industrial Organization*, 24, pp. 449-465.
- Kalnins, A., K.J. Mayer. (2004) Franchising, Ownership, and Experience: A Study of Pizza Restaurant Survival, *Management Science*, 50(12), pp. 1716-1728.
- Khalil, F. (1997) Auditing without Commitment *RAND Journal of Economics*, 28(4), pp. 629-640.
- Khalil, F., J. Lawarée. (1995) Input versus Output Monitoring: Who is the Residual Claimant? *Journal of Economic Theory*, 66, pp. 139-157.
- Klein, B. (1995) The Economics of Franchise Contracts, *Journal of Corporate Finance*, 2, pp. 9-37.
- Klein, B., L.F. Saft. (1985) The Law and Economics of Franchise Tying Contracts, *Journal of Law & Economics*, XXVIII, pp. 345-361.
- Krueger, A.B. (1991) Ownership, Agency and Wages: An Examination of Franchising in the Fast Food Industry, *Quarterly Journal of Economics*, 106(1), pp. 75-101.
- Lafontaine, F. (1992) Agency Theory and Franchising: Some Empirical Results, *RAND Journal of Economics*, 23, pp. 263-283.
- Lafontaine, F., E. Raynaud. (2002) “Residual Claims and Ongoing Rent as Incentive Mechanisms in Franchise Contracts: Complements or Substitutes?”, in E. Brousseau and J-M. Glachant, eds., *The Economics of Contracts: Theories and Applications*, Cambridge University Press, pp. 315-336.
- Lafontaine, F., K.L. Shaw. (2005) Targeting Managerial Control: Evidence from Franchising, *RAND Journal of Economics*, 36, pp. 131-150.
- Lafontaine, F., M. Slade. (2007) Vertical Integration and Firm Boundaries: the Evidence, *Journal of Economic Literature*, XLV(September), pp. 629-685.
- Lafontaine, F., M. Slade. (1997) Retail Contracting: Theory and Practice, *Journal of*

Industrial Economics, XLV(1), pp. 1-25.

Mathewson, F., R. Winter. (1985) The Economics of Franchise Contracts, *Journal of Law & Economics*, XXVIII(3), pp. 503-526.

Minkler, A. (1990) An Empirical Analysis of a Firm Decision to Franchise, *Economic Letters*, 34, pp. 77-82.

Michael, S.C. (2000) The Extent, Motivation and Effect of Tying in Franchise Contracts, *Managerial and Decision Economics*, 21, pp. 191-201.

Mookherjee, D., I.P.L. Png. (1989) Optimal Auditing, Insurance, and Redistribution, *Quarterly Journal of Economic*, 104, pp. 339-415.

Price, R. (1999) Voluntary Earnings Disclosures in Uniform Franchise Offering Circular, *Journal of Accounting and Economics*, 28, pp. 391-423.

Polinsky, A.M., S. Shavell. (1979) The Optimal Tradeoff between the Probability and Magnitude of Fines, *American Economic Review*, 69(5), pp. 880-891.

Rubin, P. (1978) The Theory of the Firm and the Structure of Franchise Contract, *Journal of Law & Economics*, XXI(1), pp. 223-233.

Scott, R.E., G. Triantis. (2006) Anticipating Litigations in Contract Design, *Yale Law Journal*, 115, pp. 814-879.

Shane, S. (2001) Organizational Incentives and Organizational Mortality, *Organization Science*, 12(2), pp. 136-160.

Vazquez, L. (2005) Up-front Franchise Fees and Ongoing Variable Payments as Substitute: an Agency Perspective, *Review of Industrial Organization*, 26, pp. 445-460.

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

Williamson, O.E. (1991) Comparative Economic Organization: the Analysis of Discrete Structural Alternative, *Administrative Science Quarterly*, 36(2), 269-296.