Author(s): Margaret Slade with Francine Lafontaine
Article Title: Vertical Integration and Firm Boundaries: The Evidence
Year of publication: 2007
Link to published version: http://dx.doi.org/10.1257/jel.45.3.629

Vertical Integration and Firm Boundaries:
The Evidence[1]

Francine Lafontaine
Stephen M. Ross School of Business
University of Michigan
Ann Arbor, MI
USA
email: laf@umich.edu

and

Margaret Slade
Department of Economics
University of Warwick
Coventry CV4 7AL
UK
email: m.slade@warwick.ac.uk

May 2007

Journal of Economic Literature classification numbers: L22, L24

Keywords: Vertical integration, firm boundaries, vertical mergers, firms versus markets

[1] We would like to thank Roger Gordon and two reviewers for helpful comments, and Desmond (Ho-Fu) Lo for his assistance. Margaret Slade would like to acknowledge financial support from the Leverhulme Foundation.

1 Introduction

Understanding what determines firm boundaries and the choice between interacting in a firm or a market is not only the fundamental concern of the theory of the firm, but it is also one of the most important issues in economics. Data on value added, for example, reveal that in the US, transactions that occur in firms are roughly equal in value to those that occur in markets.[2] The economics profession, however, has devoted much more attention to the workings of markets than to the study of firms, and even less attention to the interface between the two. Nevertheless, since Coase's (1937) seminal paper on the subject, a rich set of theories has been developed that deal with firm boundaries in vertical or input/output structures. Furthermore, in the last 25 years, empirical evidence that can shed light on those theories has been accumulating.

The empirical literature on vertical integration has focused on two main, interrelated questions: First, what types of transactions are best brought within the firm, or, put differently, under what circumstances do we observe that an input or service is produced in house? And second, what are the consequences of vertical integration for economic outcomes such as prices, quantities, investment, and profits? The answers to those questions are important in that they can inform managers' decisions directly. But they are also important ultimately as input into the development of sensible vertical merger policy and related government intervention in vertical relationships.

In this paper, we review the findings of empirical studies that have examined either or both questions. Recent articles have surveyed the theories of vertical integration (see, e.g., Holmstrom and Roberts (1998), Whinston (2003), and Gibbons (2005)) with an eye towards highlighting similarities and differences among the theories, while others have surveyed the evidence that relates to a particular theory (see, e.g., Klein (2005) or Shelanski and Klein (1995), and Lafontaine and Slade (1997, 2001) respectively for surveys of tests of transaction–cost and moral–hazard models). However, we are aware of no prior survey of the evidence on vertical integration that encompasses the different approaches and tests of various models.[3] [4]

[2] For manufacturing the ratio is about one third, whereas for services it is twice that. Calculated by authors from Census bureau data.

[3] See Cooper et al (2004) and Lafontaine and Slade (2005) for reviews of the empirical literature on vertical restraints as opposed to vertical integration.

[4] Note that we focus mostly on empirical research published in economics. TCE-based empirical research published in marketing and in management or strategy journals in particular is quite voluminous, and though we discuss some of these studies, a complete overview of this literature is beyond the scope of the present article. But see Klein (2005) for example for a broader coverage of TCE-based research.

Our reasons for reviewing the evidence on this topic at this time are twofold. First, we believe that enough evidence has accumulated by now and it is time to assess what the empirical regularities can tell us about the predictive power of existing theories, as well as how they can guide the development of future theories. In particular, we are interested in highlighting areas of potential cross-fertilization, namely how tests of one type of theory might be relied upon to inform us as to the validity of other theoretical approaches. Second, we also believe that it is important to examine what the evidence can tell us about the efficacy of public policy towards vertical mergers and divestitures, as well as how it can guide future competition policy.

The paper is organized as follows. In the next section, we present theories and evidence about the decisions that firms make concerning their boundaries. We begin with decisions to integrate forward into retailing, and then discuss backward integration into input production. We treat these separately largely because the models that authors have relied upon to derive testable implications have been different for these two sets of decisions. Specifically, most of the empirical literature on firms' decisions to integrate forward into retailing relies on incentive and moral–hazard type arguments, whereas the empirical literature on backward integration, otherwise known as the "make or buy" decision, mostly tests predictions derived from transaction–cost arguments. In our treatment of both forward and backward integration, we begin with an overview of some stylized facts from the literature, and then present simple versions of the relevant models that highlight the predictions that have been taken to data.[5] We then organize our presentation of the evidence around the predictions derived from the models. In particular, the empirical studies are organized into tables according to model tested (e.g., moral–hazard) and issue addressed (e.g., riskiness of transactions). Each table contains information on the industry examined, the empirical technique used, and the author's interpretation of the findings. We conclude this section with some thoughts on the potential for cross fertilization, that is, how evidence relating to one model or context can also shed light on other models of vertical integration. In Section 3, we review the theories and evidence concerning the consequences of vertical integration for a number of different outcomes, including prices, costs, profits, and investment. The sections on both incidence and consequences include brief discussions of the main econometric problems that authors face, as econometric and data issues have been major challenges in this literature. Finally,

[5] Some might say that our models are simplistic. Nevertheless, the simple models, while neglecting much of the richness of the theories, are capable of capturing most of the comparative statics that empiricists have focused on.

in the last section, we draw some general lessons from the body of evidence, focusing in particular on what it can tell us about the theories of firm boundaries as well as about public policy towards vertical mergers.

2 The Vertical–Integration Decision

In this section, we examine the evidence that relates to circumstances under which firms choose to integrate vertically. This in turn requires that we define precisely what we mean by vertical integration and market transaction. The difference that we emphasize is that, under the former, ownership is joint and control rights are integrated, whereas under the latter, they are separate. As will become clear, we do not distinguish between the entrepreneurial firm and the modern corporation, nor do we discuss issues of governance within firms or modes of market organization. In particular, while we recognize the important role of contracts as potential ways to achieve "almost integration," we rely on authors' institutional knowledge and, as such, do not question the definition of vertical integration and markets that is used in the empirical studies. In most cases, this implies that we equate contracts with arms length transactions, and contrast firms' decisions to rely on such transactions versus vertical integration.[6]

The empirical literature on the vertical–integration decision is easily divided into two major segments: those papers that consider the decision whether to integrate forward into retailing, and those that examine the "make or buy" decision, which is the decision whether to integrate backwards. [7] We discuss the evidence on these below in that order. In both cases, we begin by describing some stylized facts, followed by a simple version of the type of model that authors have relied upon to derive the

[6] We very briefly mention potential contractual alternatives to vertical integration in Section 3, where we present market power arguments for vertical integration. The notion that contracting and vertical integration are equivalent is a recurring theme in particular in the literature on vertical restraints. However, in the models of vertical integration we focus on, what is and is not vertically integrated is well defined, but differs across theories. For example, agency theory focuses on differences in residual claims - with vertical integration involving no such claims for the agent - while property right theory defines vertical integration as common ownership of assets and associated control rights. A complete treatment of contracting and when it is similar or not to vertical integration is beyond the scope of the present paper. However, see Lafontaine and Slade (2006) and Lafontaine and Slade (2008) for reviews of the empirical literature on the effects of vertical restraints and on inter-firm contracts respectively.

[7] Of course, decisions to integrate forward into retailing are also a form of "make or buy" decision in that the firm is choosing to make or buy downstream distribution services. However, the literature typically reserves the expression "make or buy" to contexts where firms integrate backward, and we follow this convention here.

hypotheses that they test. Our motive for discussing the theories is not to produce a comprehensive survey of their richness but rather to provide us with a framework within which to present the evidence. In particular, the evidence that we present is organized around the predictions of simple bare–bones models.

Throughout our presentation we discuss some of the measurement challenges that authors face but mostly ignore econometric problems. We do this to keep the overview tractable and of reasonable length. This is not to say that the econometric problems are unimportant. To make this point clear, in each of the two segments we highlight some of the econometric issues that researchers must confront. Since those issues are not always dealt with satisfactorily, one can be skeptical about some of the conclusions that authors have reached. Nevertheless, taken as a body, the evidence is often so strong that it can overcome much of our skepticism.

2.1 Forward Integration into Retailing

The empirical literature on forward integration generally considers a manufacturer's decision to sell her outputs to consumers herself – that is reaching customers through premises she owns and operates directly – versus using independent retailers. This question, in turn, arises in contexts where manufacturers produce a set of outputs that can be sold by themselves in branded stores. For that reason, this literature has been concerned with distribution under exclusive dealing, as in the case of franchising, rather than common agency, such as sales through department or grocery stores.

Franchising commonly takes two forms: traditional and business–format. The former involves an upstream manufacturer and a downstream retailer (e.g., gasoline or automobile sales), whereas the latter does not involve upstream production. Instead, the franchisor sells a business format — a way of doing business — to the franchisee and allows him to use the trademark (e.g., fast–food sales or hotel services).

A franchise, whether traditional or business–format, is an independent business under the law and is thus not vertically integrated with the upstream firm. Nevertheless, transactions are often not completely arm's length. Indeed, business–format franchise contracts are normally long term, and involve the payment of royalties (ρ) and fixed franchise fees (f) to the principal. The agent then obtains $(1-\rho)q - f$ where q is the value of output, as compensation for his effort. Traditional franchises, on the other hand, are dealer networks where franchisors, instead of charging fees to franchisees directly, earn a return on the products they sell to them.[8]

[8] The distinction between business–format and traditional franchising is partly a matter of degree

Most franchisors operate some outlets directly, while they franchise others. Due to data constraints, many empirical studies of business–format franchising have focused on the proportion of company owned, or vertically integrated, outlets across chains as their main dependent variable.[9] Studies of traditional franchise relationships more often have looked at the vertical integration decision outlet by outlet.

The literature has revealed a number of consistent patterns. In particular, there is systematic evidence that franchisors and manufacturers rely on independent retailers or franchisees to a greater extent the more important is the effort of the franchisee, or the more geographically dispersed the operations of the firm are. Authors have also found a positive relationship between risk or sales variability and the use of franchising. On the other hand, these firms vertically integrate more when the inputs provided by the franchisor, namely the value of the brand, is greater. They also integrate a greater proportion of their outlets when their outlets are larger.

In what follows, we provide the detailed results behind these stylized facts. In the majority of cases, authors in this literature have relied on agency theoretic arguments, and more specifically incentive or moral–hazard models of franchise relationships, to derive predictions to take to their data, and perhaps most importantly, organize and interpret their results. In the next section, we present a simple moral–hazard model that generates many of these predictions. At the same time, the model helps in pinpointing the type of empirical model of vertical integration that arises from this approach, and the similarities and differences between this approach and others we discuss further below.

2.1.1 The Moral–Hazard (MH) Model

The idea that risk and uncertainty are important determinants of firm size and scope dates at least as far back as Knight (1921), who emphasized the need to insure workers and consolidate managerial decision making. The problem with insurance, however, is that workers who are fully insured do not necessarily have appropriate incentives to supply effort. Moral–hazard arguments for firm boundaries thus emphasize the tradeoff between providing workers with insurance, which firms do well, and with

as some business–format franchisors, e.g. Baskin Robbins, also sell inputs to their franchisees and earn a return on such sales, and the payment scheme in traditional franchising may also include some payments beside per unit markups (e.g. rental fees in gasoline retailing).

[9] A few studies examine also, or instead, factors that affect the proportion of revenues to the principal (royalty rate). Given our focus on vertical integration, we do not review those results, but simply note that they are generally consistent with those found for the proportion of outlets that are integrated.

effort incentives, which markets do well.

To illustrate, it is common for firms to pay workers fixed wages that are independent of performance, at least in the short run.[10] Within the firm, therefore, incentives tend to be low powered whereas insurance is high, since worker pay does not fluctuate. Independent contractors, in contrast, are entrepreneurs who receive the profits that remain after variable costs have been paid. In other words, they are residual claimants. When a transaction occurs in a market, such as when an input is procured via an independent contractor, incentives thus tend to be high powered. However, the independent contractor also bears much risk, since his pay fluctuates in response to both demand and production shocks.

We use a standard principal/agent moral–hazard model of worker compensation to derive some testable hypotheses.[11] A slight modification of that model yields a theory of vertical integration.

Since our goal is to make predictions about forward integration, in our model, the principal is a manufacturer (M) while the agent is a retailer (R). Assume that both principal and agent must exert effort, a_M and a_R, respectively. Examples of such efforts would include advertising the brand, using high–quality inputs, and performing services at the point of sale. Output is produced (sales are realized) according to the production function

$$q = f(a_M, a_R, u), \tag{1}$$

where u is a random variable that captures uncertainty in the production process. For simplicity, we assume that this function is linear, which implies that retailer and manufacturer effort are additively separable, and that u is normally distributed with zero mean and constant variance, so that

$$q = \beta_0 + \beta_M a_M + \beta_R a_R + u, \quad u \sim N(0, \sigma^2). \tag{2}$$

In this production function, β_M and β_R, which are the marginal products or returns to manufacturer and retailer efforts, are assumed to be non–negative. In other words, effort is not unproductive. [12]

The principal would like to design an optimal payment scheme for the agent. However, agent effort is not observable and, due to the presence of u, it cannot be

[10] Clearly there are other payment schemes, such as piece rates, that are available to firms. However, fixed wages are more common, at least for nonmanagerial staff. Of course, career concerns provide incentives to forward–looking workers.

[11] Variants of this model can be found in Lafontaine and Slade (2001).

[12] With this formulation, q can be negative and hence may best be thought of as profit or returns than sales. However, the probability that $q < 0$ can be made arbitrarily small by our choice of β_0.

inferred by the principal.[13] We assume that the compensation scheme is based on realized output, q, which we take to be observable.[14] We also assume that the agent compensation scheme is linear, i.e. $s(q) = \alpha q + W$, where α is a parameter that determines the intensity of incentive pay, and W is a fixed wage that is independent of effort.[15] Finally, the private cost of effort is $c(a_i) = \frac{1}{2}(a_i)^2, i = M, R$, and there are no other costs.

The parameter, α, plays a key role in the analysis as it determines the agent's share of residual claims. Two limit cases are of interest. When $\alpha = 0$, the agent is a salaried employee who is perfectly insured, whereas when $\alpha = 1$, the agent is the residual claimant who bears all of the risk. One expects that, in general, $0 \leq \alpha \leq 1$, but, as we discuss below, the firm may not find it optimal to use non–limit values for α. Still, we identify α with the power of the agent's incentives.

We assume that the principal is risk neutral, whereas the agent, who is risk averse, receives utility from income y according to the constant absolute risk aversion, or CARA, utility function, $u(y) = -e^{-ry}$, where r is his coefficient of absolute risk aversion. Note that agent risk aversion was required to generate a sharing arrangement in early agency models. More recently, and partly because of some of the evidence that we summarize below, models that do not require agent risk aversion have become more common place. Moreover, even our simple model yields output sharing without risk aversion. Nevertheless we introduce agent risk aversion to generate many of the predictions that have been tested in the literature.

The first–best solution is the set of effort levels that maximize the joint surplus. Under our assumptions, the first–best efforts are $a_i^{**} = \beta_i, i = M, R$.

The second–best problem has two incentive constraints: given the payment scheme, the principal chooses effort to maximize her expected income, $E(\pi)$, and the agent chooses effort to maximize his certainty–equivalent income, $E(y) - \frac{r}{2}\text{VAR}(y)$, where $E(\cdot)$ and $\text{VAR}(\cdot)$ are the expectation and variance functions, and the term $\frac{r}{2}\text{VAR}(y)$ is the agent's risk premium.[16]

[13] In other words, this a moral–hazard model.

[14] Difficulties in either measuring output or inferring effort from output will reduce its appeal as a compensation basis, and increase the appeal of alternatives such as direct quality monitoring or other signals of effort. See the discussion of *costly monitoring* in the next subsection on this issue.

[15] The assumption of a linear compensation scheme is motivated by an empirical regularity: a large fraction of real–world contracts take this form. Optimal contracts are in general more complex. However, see Bhattacharyya and Lafontaine (1995) for a discussion of the optimality of linear contracts in cases where both principals and agents exert unobservable effort, as is the case here.

[16] Given our assumptions on functional forms, this expression for the risk premium is exact.

The first–order conditions for those maximizations are $a_M^* = \beta_M(1-\alpha)$ and $a_R^* = \beta_R\alpha$. Note that, in general, the situation is one of underinvestment in effort by both parties relative to the first best. Furthermore, as α increases (falls), the agent's (principal's) effort moves towards first best, but the principal's (agent's) effort moves towards zero.[17]

Finally, the principal chooses α to maximize the joint surplus, taking into account both incentive constraints and the agent's participation constraint. This maximization yields[18]

$$\alpha^* = \frac{\beta_R^2}{\beta_R^2 + \beta_M^2 + r\sigma^2}. \tag{3}$$

One can transform equation (3) into an empirically tractable model of the share parameter, α, as a function of a set of variables that capture the fundamentals of the technology and the agent's utility, namely the β's, r and σ, by appending a random variable ϵ with cumulative distribution function $F(\cdot)$ to (3).[19] We interpret that variable as representing those factors that affect desired compensation but are unobserved by the econometrician.[20]

The model thus far has focused on agent compensation, specifically incentive pay. A slight modification yields a theory of vertical integration. For this version, we use the participation constraint, which will bind in equilibrium (i.e., $u(y) = 0$), to obtain W as a function of the model parameters. This function is substituted into the principal's expected profit under vertical separation, which becomes

$$E(\pi)^{VS} = (1-\alpha)E(q) - c(a_M) - W = \frac{1}{2}\beta_M^2 + \frac{1}{2}\beta_R^2(\alpha^* - \epsilon). \tag{4}$$

Profit under integration ($\alpha = 0$) is

$$E(\pi)^{VI} = \frac{1}{2}\beta_M^2. \tag{5}$$

Suppose that a fixed cost is associated with writing and administering a contract, a transaction cost T. Then the principal will choose integration if

$$E(\pi)^{VI} - [E(\pi)^{VS} - T] = -\frac{1}{2}\beta_R^2(\alpha^* - \epsilon) + T \geq 0. \tag{6}$$

[17] We interpret zero effort as some minimal level.

[18] See Lafontaine and Slade (2001) for this and other calculations.

[19] We assume that F is differentiable and that $F' > 0$. Furthermore, in deriving our model of vertical integration, we assume that $F(.)$ is symmetric, an assumption that can easily be relaxed.

[20] See for example Lafontaine (1992) and Lafontaine and Shaw (1999) for empirical analyses of franchise–contract sharing terms that rely on this approach.

This will be true if

$$\frac{2T}{\beta_R^2} - \frac{\beta_R^2}{\beta_R^2 + \beta_M^2 + r\sigma^2} \geq -\epsilon. \tag{7}$$

The probability of observing integration then is

$$PROB[VI] = F\Big[\frac{2T}{\beta_R^2} - \frac{\beta_R^2}{\beta_R^2 + \beta_M^2 + r\sigma^2}\Big]. \tag{8}$$

which is our discrete–choice model of vertical integration. The model predicts that, in cases where the costs of incentive contracting outweigh the benefits, the principal will offer the agent a uniform–wage contract ($\alpha = 0$ or VI). Furthermore, when contracting is desirable, the model can be used to predict the power of the incentives that will be offered. Finally, in some circumstances, arm's length transactions ($\alpha = 1$) will be chosen.

This moral–hazard model of vertical integration yields a number of testable predictions.[21] First, equation (8) implies that the probability of vertical integration should be lower when the retailer's effort is more productive (i.e., the partial derivative of the right–hand side of (8) with respect to β_R is negative). On the other hand, that probability is expected to be higher when the manufacturer's effort is more productive, when risk (σ^2) and/or retailer risk aversion (r) is greater, and when the cost of contracting is higher.

To rephrase these predictions, when the marginal return to an individual's effort becomes larger, that individual should be given higher powered incentives (a higher fraction of residual claims). In our context, this will mean more or less vertical integration depending on whose effort we are considering. However, when risk or agent risk aversion increases, insurance considerations become more important and the agent – who is the only risk averse party in the model – should be given lower powered incentives, which implies that we should see a greater tendency towards vertical integration in the data.[22] Finally, if contracting were costless ($T = 0$), vertical integration would be a limit case ($\alpha^* = 0$) that was rarely observed.

[21] The same predictions can be obtained from (3) or from (8). This means that we can simultaneously discuss the power of an agent's incentives in, for example, a revenue–sharing contract and the choice between interacting in a firm or a market.

[22] But see Lafontaine and Bhattacharyya (1995) and Prendergast (2002) who point out that an agent's optimal use of private information may lead to a positive association between observed risk and agent incentives. Also see Ackerberg and Botticini (2002) for an argument that less risk averse agents may be attracted to riskier contracts, thereby negating the expected correlation between risk and incentives in our simple model.

Now consider adding another variable, x, to the production function. For example, x might be the size of the retail outlet, the size of the market in which it is located, or any other characteristic of the principal, the agent, the outlet, or the market. Whether or not x makes a difference to the vertical–integration decision depends on how it enters equation (2). The following is a fairly general formulation,

$$q = \beta_0 + (\beta_M + \beta_{Mx}x)a_M + (\beta_R + \beta_{Rx}x)a_R + (\gamma + u)x. \tag{9}$$

Under our assumptions above, which are fairly typical in this literature, γ has no effect on agent incentives or the decision to integrate.[23] Indeed, unless x affects the marginal return to the effort of the principal or agent in the model, or unless it affects risk, it is irrelevant. This means that x must interact with a_M, a_R, or u to become relevant. Furthermore, the comparative statics for the interaction terms are the same as those discussed above. Specifically, if x, for example, increases the marginal return to the agent's effort while leaving the return to the principal's effort unchanged, then increases in x will be associated with a greater tendency to use markets rather than firms.

The model that we have developed can be extended to accommodate many complications. For example, we have considered only one agent and can thus say nothing about team production. Yet the notion of team production, which occurs when individuals working together are more productive than when each works alone, is central to the moral–hazard model of the firm (see e.g., Alchian and Demsetz (1972) and Holmstrom (1982)). Team production is a technological characteristic of the production process — similar to those emphasized in neoclassical theories of vertical integration — whereby individual marginal products are enhanced by the efforts of others. This, in turn, makes interaction within a firm more desirable. Unfortunately, it also makes the allocation of rewards more difficult. Indeed, if wages are based on marginal productivities, a shirking worker can lower the wages of everyone in the team.

In our simple model, team production could be introduced by adding a second agent and allowing agent efforts to interact in equation (2). In other words, the marginal product of each agent's effort could depend on the effort of the other one. Unfortunately, the departure from linearity caused by allowing efforts to interact would make the solution of the model more difficult. A similar difficulty would arise if we allowed synergies between the efforts of principal and agent.

[23] Note that this would still be true if γx were replaced by an arbitrary function $g(x)$.

In addition, we have not allowed for the possibility that the principal could monitor the agent's activities at a cost. Whether or not costly monitoring leads to more integration depends on the type of information that the principal can gather via this monitoring (see Lafontaine and Slade (1996)).

Finally, in our simple model, the agent performs only one task. A multitask model is much richer, but comparative–static derivatives can be signed only in special cases (see Holmstrom and Milgrom (1991 and 1994)). In other words, designing reward systems that provide high-powered incentives for multiple tasks is quite difficult. This difficulty can lead to more vertical integration, where internally firms rely on subjective performance evaluations instead of explicit task specific incentives (see Holmstrom (1999) and Azoulay (2004)).

In spite of these limitations and its simplicity overall, the model embodied in (3)–(9) is useful as it yields the types of predictions that are most often tested in the empirical moral–hazard literature.

2.1.2 Evidence on Predictions from Moral–Hazard Models

The theoretical moral–hazard model above identified a number of factors that should affect the vertical–integration decision. Unfortunately, some of those factors do not easily lend themselves to empirical assessment (e.g., the degree of risk aversion, r). In what follows, we limit attention to factors that can be assessed more readily — the importance of: risk (σ^2), downstream effort (β_R), upstream effort (β_M), and outlet size (an x). In addition, we discuss factors that require slight modifications to the basic model: monitoring difficulty, spillovers within a chain, and multitasking.

We summarize the empirical evidence in a series of tables that are organized by factor (e.g., risk). For each study, the relevant table indicates the author's name, the year of publication, the industry studied, the data type or empirical technique used, the way the factor of interest is measured, and the author's conclusion concerning the effect of that factor. These conclusions are summarized in the final column, where a + (-) indicates that the factor encourages (discourages) vertical integration, and a * indicates that the finding is significant, using a two–sided test and a 5% confidence level.[24] Parentheses in the last column indicate that the variable that is examined is an inverse measure of the factor of interest and is therefore expected to have the opposite effect on vertical integration from a direct measure. For example,

[24] Note that while we follow the author in assessing the importance and significance of different factors, and in interpreting their findings more generally, we do not always interpret their measures as they intended.

'outlet density,' which is an inverse measure of monitoring costs, is expected to have a sign that is opposite from 'distance from headquarters,' which is a direct measure. Finally, given that measurement has proved challenging in much of this literature, our discussion of each factor considers measurement issues as well as empirical findings.

Risk

The standard agency model of retail contracting suggests that, as the level of uncertainty increases, so does the need for agent insurance and thus the desirability of vertical integration with the presumably less risk averse upstream firm. In other words, the lower-powered incentives that are typically used inside the firm protect the agent from the vagaries of the market, a protection that becomes all the more valuable as uncertainty rises.

The notion of uncertainty or risk that is relevant in this context is the risk that is borne by the agent, namely the risk at the outlet or downstream level. Unfortunately, data that measure outlet risk are virtually nonexistent. For this reason, imperfect proxies are employed. The two most common are measures of variation in detrended sales per outlet and measures of failure rates such as the fraction of outlets that were discontinued in a particular period of time. Furthermore, data are often available only at the level of the sector rather than at the level of the franchisor or the retail outlet.

Table 1 gives details of studies that assess the role of risk in determining the tendency towards integration. In all but two of those studies, contrary to prediction, increased risk is associated with less integration. Moreover, the two positive findings are not significant. These results suggest a robust pattern that is unsupportive of the basic agency model. Interestingly, allowing effort to interact with risk in the basic model only makes matters worse. In particular, if a_R is interacted with u in equation (2), higher powered incentives become even more costly, since, by increasing the agent's effort, they also increase the risk that he must bear.

The finding that risk is negatively rather than positively associated with integration is a puzzle that, as noted earlier, has attracted some attention already in the literature. Some authors have concluded from the evidence that franchisors shed risk onto franchisees (e.g. Martin [1988]). This would be optimal, however, only if franchisors were more risk averse than franchisees. Unfortunately, if agents were indeed less risk averse than their principals, there also would be less need to balance the provision of incentives and insurance to those agents. At the extreme, franchisors would simply sell outlets to franchisees outright for a fixed price, a situation that is

rarely observed.

Several alternative, and we believe more satisfactory, explanations for the observed negative risk/integration relationship have surfaced in the literature. The first stems from the fact that market uncertainty can be endogenous and that the power of incentives can influence sales variability. Indeed, retailers often have superior information concerning local–market conditions. Moreover, since separation gives agents greater incentives to react to those conditions, one is likely to find more sales variability in separated than in integrated units.[25] A second possibility is that differences in risk aversion, which typically are not controlled for in the empirical analyses due to the absence of data, can explain the correlation. With this interpretation, more risk averse agents select safer markets as well as contracts with lower–powered incentives.[26] Finally, we come back to the anomalous effect of risk on the extent of vertical integration below in our discussion of evidence relating to property–rights theory, as this theory provides yet another potential explanation for the empirical regularity that appears in table 1.

Downstream Effort

The moral–hazard model predicts that increases in the importance of the retailer's input should be associated with less integration and higher-powered incentive contracts. In other words, when the agent's job is more entrepreneurial in nature, his compensation should reflect that fact.

From a practical point of view, proxies for the importance of the agent's effort (or its inverse) have included measures of labor intensity (either employee/sales or capital/labor ratios) as the agent is the one who must oversee the provision of labor. Researchers also have used a measure of the agent's value added, or discretion over input choices, and a variable that captures whether previous experience in the business is required. In the context of banking, since managers in rural settings must offer a more complete set of services, locational dummies have been used to capture levels of responsibility. Finally, two studies of gasoline retailing rely on a dummy variable that distinguishes full from self service.

Table 2 summarizes the results from studies that assess the effect of the importance of agent effort. In every case where the coefficient of the agent–importance variable is statistically significant, its relationship with integration with the upstream company

[25] See Lafontaine and Bhattacharrya (1995) and Prendergast (2002) for more on this.

[26] See Ackerberg and Botticini (2002) for an explanation based on selection and some corroborating evidence in the context of sharecropping.

is negative, as predicted by standard agency considerations and other incentive–based arguments. In other words, when the agent's effort plays a more significant role in determining sales, integration is less likely.

Upstream Effort

It is common for MH models to be based on the assumption that only one party, the agent, provides effort in the production (or sales-generation) process. Our model above incorporates the possibility that the principal also provides some effort because, in reality, success at the retail level often depends importantly on the behavior of the upstream firm or principal. For example, franchisees expect their franchisors to maintain the value of the tradename under which they operate (via advertising and other forms of promotion), and to screen and police other franchisees in the chain as well as managers of corporate stores. If this behavior is not easily assessed, there is moral hazard on both sides — up and downstream — and the franchisor, like the franchisee, must be given incentives to perform. Not surprisingly, when the effort of the principal increases in importance, it is the share of output that she receives, or the extent of vertical integration, that must rise.

Table 3 shows results from studies that have considered how the importance of the franchisor's effort affects the probability of integration. The importance of upstream effort is measured by the value of the tradename (proxied by the amount of advertising, the number of outlets in the chain, or the difference between the market and the book value of equity), the amount of training provided by the franchisor, or the number of years spent developing the business format prior to franchising. The table shows that, in all cases, when franchisor inputs are more important, more vertical integration is observed, as predicted.

One proxy for the importance of the franchisor's input that has been used in the literature but is not included in table 3 is the chain's number of years of franchising (or business experience). The idea is that more years in franchising (or business) lead to a better known, and thus more valuable, tradename. However, that variable is also a proxy for the extent to which franchisors have access to capital as well as for learning and reputation effects. Furthermore, cross-sectional evidence relating to this variable is affected by the adjustment process all franchisors go through as they first begin to expand the franchised side of their business. Using panel data at the franchisor level, Lafontaine and Shaw (2005) show that, after the first few years in franchising, the proportion of corporate units within chains levels off — at levels that differ across chains — and then remains stable. They conclude that a firm's years in

franchising is not a major determinant of the "stable" extent of vertical integration in these chains.

Outlet Size

Modeling the effect of outlet size is less straightforward than modeling the previous two factors, and model predictions are more sensitive to specification as a consequence. In particular, in the context of equation (9), size is a characteristic of the outlet (i.e., an x) that can enter linearly or multiplicatively or in some other form. Unfortunately, we can achieve any prediction, as noted above, for such an x variable depending on how we incorporate it in the model. We quite purposely choose a specification whose predictions are consistent with the empirical regularity that we present below. In particular, we model size as interacting with risk (u), in addition to having its own direct effect on output through γ. This interaction with risk captures the idea that the franchisee has more at stake in a larger outlet — the market is not riskier per se, but more capital is now subject to the same degree of risk. As we noted earlier, were it not for this interaction with u, x would have no effect on the optimal contract. With this interaction, it is predicted to have the same effect as risk does. This variant of the MH model thus predicts that vertical integration becomes more likely when the size of the capital outlay increases. Furthermore, vertical integration in this context has the added advantage that it substitutes the principal's capital for the agent's.

Unlike the factors discussed earlier, the empirical measurement of outlet size is fairly straightforward. Common measures are average sales per outlet and the initial investment required. Table 4 shows that, in all but one study, greater size leads to increased company ownership or integration. In other words, people responsible for large outlets tend to be company employees who receive low–powered incentives, as predicted.

While our specification ensures that the model and evidence agree, it is nonetheless possible to argue for the opposite relationship in an equally convincing manner. Indeed, when an outlet is large, the agent has more responsibility. For this reason, outlet size has been interpreted as a measure of the importance of the agent's input in the literature.[27] Not surprisingly then, it is often claimed that an agency model should predict that an increase in size will be associated with less integration and higher–powered incentives. The data, however, contradict that prediction.

[27] In terms of our agency model (9), this is equivalent to interacting size with a_R. If it were interacted with a_M, predictions would be reversed.

Costly Monitoring

The idea that monitoring the agent's effort can be costly or difficult for the principal is central to the incentive–based–contracting literature. In fact, if monitoring were costless and effort were contractible, there would be no need for incentive pay.

Given the centrality of the notion of costly monitoring, it is somewhat surprising that there exists confusion in the literature concerning the effect of an increase in monitoring cost on the tendency towards vertical integration. Indeed, one can find statements that imply that monitoring difficulties should, on the one hand encourage, and, on the other hand discourage, integration.[28]

To reconcile those discrepancies, Lafontaine and Slade (1996) modify the standard agency model to include the possibility that the principal can use not only outcome (i.e., sales) information to infer something about the agent's effort, but also a direct signal of effort. Furthermore, the principal is allowed to base the agent's compensation on both signals. We consider two types of signals because, in most real–world manufacturer–retailer relationships, it is possible to supervise the actions of a retailer directly by, for example, testing food quality, assessing the cleanliness of the unit, and determining work hours. This direct supervision provides the manufacturer with information on retailer effort that supplements the information contained in sales data.

To model this situation in the simplest possible way, we follow Lafontaine and Slade (1996) and replace the effort/sales relationship in a basic agency model – one that involves only franchisee moral hazard – with two functions to denote the fact that the principal receives two noisy signals of the agent's effort, a_R. In particular, the principal observes retail sales, q, and a direct signal, e. We assume that the vector of signals is unbiased and normally distributed with covariance matrix Σ, where $\Sigma = [\sigma_{ij}]$.

The contract that the principal offers the agent is amended to include, in addition to the fixed wage W, not only an outcome–based or sales commission rate, α_1, but also a behavior–based commission rate, α_2, that relates to the direct signal of effort. With the simplest version of the model, the two signals are uncorrelated ($\sigma_{ij} = 0, i \neq j$). Under that assumption, solution of the two first-order conditions yields

$$\alpha_i^* = \frac{1}{1 + r\sigma_{ii} + \sigma_{ii}/\sigma_{jj}}, \tag{10}$$

[28] For example, consider the following statements from the empirical literature: "The likelihood of integration should increase with the difficulty of monitoring performance." (Anderson and Schmittlein [1984 p. 388]). "Franchised units (as opposed to vertical integration) will be observed where the cost of monitoring is high." (Brickley and Dark [1987 p. 408], text in parentheses added).

where r is again the agent's coefficient of absolute risk aversion. This solution is to be compared to the optimal contract from the basic agency model, which yields $\alpha^* = 1/(1+r\sigma^2)$. Equation (10) then shows that with two signals of effort, the optimal contract must be modified to account for the relative precisions of the signals. In other words, the compensation package places relatively more weight on the signal with the smaller variance.

We are interested in the effect of increases in the two sorts of uncertainty on the size of α_1^*, since this is the incentive-based pay that appears in the data. It is straightforward to show that increases in the precision of sales data ($1/\sigma_{11}$) lead to a higher reliance on outcome-based compensation (higher α_1), which corresponds to less vertical integration. However, increases in the precision of the direct signal of effort ($1/\sigma_{22}$) lead to less outcome-based compensation (lower α_1) or more vertical integration.

While the above model does not explicitly include monitoring costs, it should be clear that, when the cost of increasing the precision of sales data as an indicator of effort is low, we should observe more reliance on sales data in the compensation scheme, which means less vertical integration. On the other hand, when the cost of behavior monitoring is low, the firm will perform more of that type of monitoring. A low σ_{22} will lead the firm to choose a lower α_1, which amounts to more vertical integration.

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

Turning to the empirical evidence, the first part of table 5, under *'Outcome Monitoring'*, shows results obtained in the sales–force–compensation literature, where the focus has been on the usefulness of observed sales data in assessing agent effort. In the first two studies, researchers asked managers to respond to statements such as 'it is very difficult to measure equitably the results of individual salespeople' or 'team sales are common.' Other measures of the usefulness of outcome measures of effort include the length of the selling cycle (on the basis that a long lag between actions and market responses makes it difficult to attribute output to effort), as well as a measure of environmental uncertainty that captures the extent to which agents control sales

outcomes. Using scores thus obtained as measures of the cost of monitoring sales and inferring effort from it, researchers found that higher monitoring costs lead to more vertical integration, as predicted by our model.

The second part of table 5, which is labeled *'Behavior Monitoring'*, contains empirical results that come mostly from the franchising literature. Here, authors have focused on the cost of direct monitoring of behavior, that is information that is used to supplement data on sales outcomes. Frequently used measures of behavior–monitoring costs include some notion of geographic dispersion or of distance from monitoring headquarters. Those measures are proxies for the cost of sending a company representative to visit the unit to obtain data on cleanliness, product quality, etc. Other measures are inversely related to costs. These include outlet density and, in the case of trucking, the presence of onboard computers. The table shows that, regardless of whether behavior–monitoring costs are measured directly or inversely, in all cases where coefficients are significant, higher monitoring costs lead to less vertical integration.[29] Again the evidence is consistent with the model.

In sum, the two types of measures that authors have relied upon in the empirical literature have captured different types of monitoring costs: the fit of sales data to individual effort versus direct monitoring of behavior that is a substitute for sales data. Taking this difference into account, the "contradictory" results obtained and claims made by researchers are in fact consistent with each other as well as with standard incentive arguments.

Spillovers Within a Chain

One reason for the prevalence of chains rather than single outlets in the retail and service sectors is that there are externalities that are associated with the brand or chain name. Although such spillovers are meant to be beneficial, they can also create problems for both up and downstream firms. For example, one form that a spillover can take is a brand–loyalty demand externality. With that sort of spillover, a low price at one outlet in a chain increases demand not only at that outlet but also for other retailers in the same chain. Conversely, a high price at one outlet can cause customers to switch their business to another chain rather than merely seek a different unit of the same chain. When this sort of externality is important, integration becomes more desirable. The reason is that the chain internalizes the spillover that is external to the individual unit.

Franchisee free riding can take a variety of forms. For example, franchisees can

[29] Recall that the inverse measures (i.e., of the ease of monitoring) should have the opposite sign.

use lower quality inputs or not abide by various rules — such as a requirement that baked goods be disposed of if not sold within a certain time period — that are good for the chain as a whole but impose costs on individual franchisees. Indeed, once an agent is given high-powered incentives via a franchise contract, he can shirk and free ride on the value of the tradename (see e.g. Klein, (1980) and Brickley and Dark, (1987)). The problem is that the cost of the agent's effort to maintain the quality of the trademark is private, whereas the benefits of his activities accrue, at least partially, to all members of the chain. In this case, the spillover works through effort or product quality, not price.

Whether the externality works through price or effort, the free–riding problem is exacerbated in situations where consumers do not impose sufficient discipline on retailers, namely in cases of non–repeat business. The franchisor, unlike the franchisee, can internalize spillovers that damage the trademark by operating units in transient–customer locations, such as freeway exits, herself.

Table 6 summarizes the evidence from studies that have examined the effect of non–repeat business on the propensity to integrate. This table shows that the evidence on this effect is mixed. One explanation for the lack of strong evidence that vertical integration is used to overcome the free–riding problem is that franchisors can find other methods of controlling retail behavior by, for example, using approved–supplier requirements and imposing minimum advertising requirements. The lack of "highway" effect in particular probably reflects the fact that franchisors often contract with very large companies to operate units along freeways. These large franchisees, in turn, have incentives to maintain quality to the extent that they also internalize spillovers among all their units.[30]

Multiple Tasks

In many retailing situations the agent performs more than one task. For example, a service–station operator might repair cars as well as sell gasoline, a publican might offer food services as well as beer, and a trucker might perform cargo–handling services as well as drive a truck. Generally, when this is the case, the optimal contract for one task (and thus the propensity to integrate) depends on the characteristics of the others (see Holmstrom and Milgrom 1991 and 1994).

There are many possible variants of multi–task models. We discuss a very simple

[30] See e.g. Brickley (1999) and Kalnins and Lafontaine (2004) for evidence that franchisors grant multiple units within the same markets to the same franchisees. In both studies authors argue this is done at least in part so franchisees internalize more of the demand externalities.

version that illustrates our point. Since this version abstracts from any moral hazard issue on the part of the principal, we compare the solution here to that of the basic one-sided moral hazard version of our model, which as noted earlier, yields $\alpha^* = 1/(1 + r\sigma^2)$.

Suppose that there are two tasks and that the agent exerts effort, a_{R_i}, on the ith task. Output on each task, q_i, is a noisy signal of effort, a_{R_i}. Suppose further that the signals are unbiased, they have covariance matrix Σ, and the agent's cost of effort is $({a_R}^T a_R)/2$. As before, the principal chooses the vector of commissions, α, to maximize the total surplus subject to the incentive constraints. In the symmetric case where $\sigma_{11} = \sigma_{22} = \sigma^2$, the first-order conditions can be manipulated to yield

$$\alpha_i^* = \frac{1}{1 + r(\sigma^2 + \sigma_{12})}, \quad i = 1, 2. \tag{11}$$

If one compares (11) to the solution of the basic model, it is clear that, when a second task is added, the propensity to integrate rises (falls) if the associated risks are positively (negatively) correlated. This occurs for pure insurance reasons. In other words, positive correlation means higher risk, whereas negative correlation is a source of risk diversification for the agent.

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

A model that incorporates these three effects is developed in Slade (1996). She shows that, when an agent has full residual-claimancy rights on outcomes for a second task, the power of incentives for a first task should be lower when the tasks are more complementary. Intuitively, since the second task already has high–powered incentives, if tasks are substitutes and incentives are low for the first, the agent will spend most of his time on his own activity (working in the back court). Her empirical application to gasoline retailing supports the model's prediction. Specifically, she finds that when the second activity — the one for which the agent is a full residual claimant — is repairing cars, an activity that is less complementary with selling gasoline than is managing a convenience store, then vertical integration of the gasoline–selling task is less likely.

Similarly, Baker and Hubbard (2003) look at multitasking in for–hire trucking, where the two tasks are shipping and cargo handling. Since onboard computers

(OBC) in trucking facilitate coordination and lower the cost of multitasking, the comparative–static predictions they derive work through costs rather than risk or demand. They find that adoption of OBC results in more integration, particularly in situations where multitasking is important.

For his part, Azoulay (2004) provides evidence that, in the pharmaceutical industry, firms outsource clinical trials that are data intensive while they keep in house those trials that are also knowledge intensive. He argues that firms thereby ensure that incentives on both data and knowledge production are balanced for those projects that really involve both tasks. For those projects involving mostly data production, in contrast, outsourcing provides appropriate high–powered incentives for that main task. Given that pharmaceutical firms rely on flat explicit incentives, combined with some subjective performance evaluations, to evaluate worker performance internally, he interprets this evidence as consistent with multitask agency models (see Holmstrom 1999).

Finally, our double–sided moral hazard model above assumes that franchisees are responsible for local service provision while franchisors manage the brand and its value. Instead of viewing principals as active in the goodwill–production process, it is possible to model agents as those who put effort into both local service and brand value. In this context, free-riding is a situation in which the principal would like her agents to engage in various activities to support the brand, or at least not reduce its value, whereas those agents, when paid residual claims, choose to put too much effort into increasing their own profits.[31] The results from Nickerson and Silverman (2003) and Lafontaine and Shaw (2005), described in Table 3, where higher brand values are associated with more vertical integration, can thus be interpreted as supporting a multitasking view of incentive provision for agents. In particular, vertical integration, which corresponds to lower–powered incentives for local effort, is appropriate when brand protection is more important. Similarly, Yeap (2006) finds evidence that more complex production activities in restaurant chains, in particular onsite food production and table service, are associated with more company ownership. She interprets her results in terms of multitasking, arguing that the chains do not want to rely on high–powered incentives for agents when, if some of their tasks are not tended to properly, this can have a large detrimental impact on the chain as a whole.

[31] Bai and Tao (1999) propose a multitask model in this spirit where outlet managers are responsible for both local service provision and goodwill value.

Summary

Two central predictions of the moral–hazard model of forward integration have been confirmed by the empirical evidence. These are that as the importance of local or downstream effort grows, integration becomes less likely, whereas as the importance of company–wide or upstream effort grows, integration becomes more likely, where importance is measured by the marginal productivity of effort. Moreover, the idea that monitoring the agent is costly is also central to the moral–hazard model of contracting. Nevertheless, there has been some confusion in the literature concerning the effect of higher monitoring cost on vertical integration. We showed that once one recognizes that there are two sorts of monitoring that the principal can perform — outcome and behavior monitoring — the evidence again is highly supportive of the agency model. On all these fronts, the moral–hazard model performs very well.

Model predictions concerning the effects of other factors, such as outlet size, spillovers, and multitasking, are more sensitive to the specification of how those factors enter the output/effort relationship. However, there are reasonable model formulations that lead to predictions that are supported by the evidence. Moreover, here again the evidence leads one to conclude that aligning incentives is a central concern.

One important prediction from basic agency models, however — that increased risk makes integration more likely as insurance considerations begin to dominate — is not supported by the data. One possible explanation for the discrepancy between theory and evidence is that output variability is likely to be endogenous in situations where agents have private information about local–market conditions. However, a similar finding surfaces in the sharecropping literature (see Allen and Lueck (1995) for a survey), a context where exogenous output fluctuations are more apt to dominate. Allen and Lueck suggest that measurement costs, i.e. the possibility that tenants might try to underreport output, may explain the anomalous risk effect in sharecropping. Another possible explanation mentioned above relies on selectivity in a situation of heterogeneous risk preferences.[32] We suggest a third possible explanation below in our discussion of evidence pertinent to the property–rights model.

Finally, many of the predictions from the MH model have received significant attention in the empirical literature, allowing us to draw conclusions from several studies and different contexts. But although theory and evidence seem well integrated

[32] Ackerberg and Botticini (2002) find that risk is associated with more sharing relative to fixed rent contract once they control for endogenous matching in their data, a result that is consistent with predictions from basic agency theory.

in this area, there remains a need for further detailed empirical analyses to test implications derived from various extensions of the basic model in similar and new institutional contexts.

2.2 Backward Integration into Input Production

The empirical literature on backward integration is concerned with a manufacturer's decision to integrate either partially or completely with its suppliers of parts or equipment, or, put differently, the decision to make or buy its supplies. Most of this literature has addressed predictions derived from transaction–cost economics (TCE) even though property–rights theories also have aimed to explain when firms might integrate backward. As noted by Joskow (2005) "the TCE framework has stimulated much more empirical work than [...] the more recent property–rights literature. This is to the credit of the scholars who have done theoretical work in the TCE tradition since they have produced testable hypotheses and endeavored to provide guidance to empirical researchers regarding how to measure relevant attributes of transactions affecting market contracting and internal organization."

The large body of empirical research in the area has found considerable support for the notion, derived from TCE, that specific investments are economically and statistically important when it comes to the decision to organize the production of a given input internally or externally. It also has established that backward integration is more likely for more complex inputs and when the environment within which the firms operate is more uncertain. In some cases, this same evidence has been interpreted as providing support for property–rights models of vertical integration. Whinston (2003) has shown, however, that the property–rights approach generates a distinct set of predictions. We discuss this in some detail below.

In what follows, we review the empirical literature on the make–or–buy decision, organizing the evidence along the lines suggested by the theories. With this in mind, we first present the theoretical arguments, starting with transaction–cost economics, followed by the property–rights approach. Moreover, since transaction–cost arguments are usually informal, our overview of that set of arguments is also informal. Also, like our simple moral–hazard model above, the property–rights model that we present is a bare-bones or skeleton version of the theory that it represents, and we do not attempt to portray the richness that this class of models can embody. Nevertheless, we believe that the simple skeleton that we discuss captures the main predictions needed to organize our discussion of the evidence.

2.2.1 The Transaction–Cost Model

Transaction costs (TC) are the costs of establishing and administering business relationships within and between firms or individuals, including those costs associated with opportunistic behavior and haggling *ex post*. TC theories of firm boundaries can be traced back to Coase (1937), who focused on the costs of transacting under different organizational forms, particularly the costs of writing and enforcing contracts. Those theories have been developed further, notably by Williamson (1971, 1975, 1979, 1985), Klein, Crawford, and Alchian (1978), and others.

The fundamental insight of TCE concerning vertical integration is as follows. Parties to a transaction often make investments that have greater value inside than outside the relationship. In other words, the value of the assets in their intended use is higher than their value in alternative uses. Examples include specialized tools that can only be used to produce the products of one manufacturer, training that increases worker productivity exclusively in using those tools, and supplier facilities that have been located in close geographic proximity to purchasers. Specific investments give each party to a relationship a degree of monopoly or monopsony power. Indeed, even when there are many potential trading parties *ex ante*, when investments are specific, parties are locked in *ex post*.

When specific assets are involved, parties can write long–term contacts to protect themselves and their assets. If such contracts were complete, specificity would not create problems. The complete contract would specify exactly what will occur and who will control the assets under all possible contingencies. However, writing complete contracts is costly, and not all contingencies can be foreseen. Thus real–world contracts are normally incomplete. In that context, specific investments generate quasi rents, and each of the parties to a contract has incentives to endeavor to capture those rents. This means that they are likely to haggle with one another, thereby increasing the costs of writing and administering the contract, as well as attempt to renegotiate the contract or, more generally, engage in opportunistic behavior *ex post*.[33] These possibilities, which are the essence of the hold–up problem, clearly pose problems for long–term contracting, and those problems are exacerbated in volatile environments.

TC theories of firm boundaries usually assume that the hold–up problem is mitigated inside the firm. However, they are often silent as to just how that mitigation occurs. Yet even inside firms, workers who have received specialized training can

[33] See Tadelis (2002) for a model that emphasizes the role of transaction complexity in affecting haggling and ex-post adaptation.

attempt to hold their employers up and vice versa. Moreover, employees also can engage in influence activities that are designed to capture quasi rents. Nevertheless, it is probably true that, even if mitigation is not complete, the problem is lessened inside firms. Indeed, relative to markets, firms are more closely related to command economies.[34] Transactions in which opportunistic behavior is known to cause large problems are therefore more apt to occur in house.

In sum, when the problems that are associated with transaction costs are important, TC models suggest that firms will choose governance structures — including vertical integration or separation — to reduce the likelihood and cost of haggling and exploitation. The theory provides a number of implications concerning circumstances under which firms are likely to choose vertical integration. Specifically, firms are expected to rely on in–house production when transactions are complex, when they involve specific investments, when those specific assets are durable, when the quality of those assets is difficult to verify, when the environment is uncertain, and when the quasi rents that are generated by a relationship are large.

2.2.2 The Property–Rights Model

Property–rights theories, which are more recent and more formal than transaction–costs arguments, were developed by Grossman and Hart (1986), Hart and Moore (1990), Hart (1995) and others. Those theories emphasize how asset ownership can change investment incentives. More specifically, they demonstrate how the allocation of property rights, which confer the rights to make decisions concerning the use of an asset when contingencies arise that were not foreseen or not specified in a contract, changes *ex ante* investment incentives.

Unlike the TC literature, the PR literature does not focus on *ex post* haggling, renegotiation, and opportunistic behavior. Instead, it stresses contractual incompleteness and develops formal models that show how *ex post* bargaining affects *ex ante* investment in noncontractible assets. Nevertheless, since PR theories deal with relationship–specific assets, incomplete contracts, and *ex post* bargaining,[35] they are often thought to be closely related to TC models. Whinston (2003), however, shows how the predictions from the two classes of theories can be very different. Indeed, unlike TC predictions, with the PR model the problems associated with specificity

[34] See Masten (1988) for a discussion of the different legal rules that apply inside and outside the firm and how those differences explain the different capacities to mitigate hold up. Also see Baker, Gibbons, and Murphy (1999, 2002) on this topic.

[35] PR theories can therefore also be traced back to Coase's (1937) seminal contribution.

need not be mitigated by bringing a transaction inside the firm. In fact, in the PR literature, vertical integration can exacerbate the problem by reducing investment incentives to levels that are even lower than those provided by markets. Finally, again relative to the TC literature, PR models provide a more rigorous set of predictions concerning the determinants of firm boundaries. Unfortunately, those predictions are also more fragile and thus more difficult to take to data.

Interestingly, although the moral–hazard literature is concerned with residual claims, whereas the property–rights literature is concerned with residual decision rights, their predictions concerning the effects of marginal–productivity changes are frequently similar, bearing in mind that MH models deal with incentives to exert effort and the productivity of those efforts, whereas PR models deal with incentives to invest in physical assets or human capital and the productivity of those investments. We demonstrate the similarities more formally below.

We follow Whinston's (2003) modelling approach to derive some comparative statics from a very simple version of PR theory. However, we specialize his assumptions to the case that most closely resembles Hart's (1995) model.[36] The model is concerned with a manufacturer (M) who must decide whether to buy an input from an independent supplier (S) or to produce it herself (to acquire S).

Assume that integration decisions are agreed upon at $t = 0$. At $t = 1$, (which we denote *ex ante*) each party to the transaction must make a noncontractible investment, i_j, at a cost $c(i_j) = \frac{1}{2}(i_j)^2, j = S, M$. Note that this occurs whether or not the manufacturer has decided to vertically integrate. In other words, under both integration and separation, someone who is not the manufacturer makes the upstream investment decision. One way to think about these investments under integration then is that the upstream division manager decides what equipment to purchase or how much effort is put into maintaining the equipment. [37]

Finally, at $t = 2$ (which we denote *ex post* since investments have been sunk), the players bargain over the surplus generated by those investments. Thus even though the *ex post* bargaining game is efficient (e.g., the Nash bargaining solution is often used), since investments are sunk, the outcome of bargaining is determined by relative bargaining strengths that are often only loosely related to relative investment levels. Furthermore, investment shifts not only the frontier — the size of the pie that is to

[36] Specifically, we do not consider cross investments, and we assume that there is underinvestment under either vertical structure.

[37] For example, in the case of GM and Fisher body, the assumption amounts to the Fisher brothers still getting to make investment decisions in say equipment used or effort put into maintenance in the plant even though the plant and equipment now belong to GM.

be split — but also the threat point — the outcome that occurs when agreement cannot be reached. Players therefore have incentives to behave strategically and use their investments to better their positions in the *ex post* game. Underinvestment is the most usual outcome of this process.[38]

In the *ex-post* game, either a bargain is struck, in which case the joint surplus is independent of asset ownership, or agreement cannot be reached, in which case payoffs differ depending on asset ownership.[39] Suppose that the *ex post* surplus when bargaining is successful is given by

$$\pi(i) = \alpha_0 + \alpha_M i_M + \alpha_S i_S, \tag{12}$$

where $i = (i_M, i_S)$ is the vector of investments. We assume that $\alpha_M > 0$ and $\alpha_S > 0$. In other words, both types of investments are productive.

First–best investments are those that maximize $W(i) = \pi(i) - c(i_M) - c(i_S)$, which implies that $i_j^{**} = \alpha_j$. The first–best *ex ante* surplus is then $W^{**} = W(i^{**})$. This level of surplus, however, simply cannot be attained. The only options available to the manufacturer are integration or separation, and neither of these yields the first-best levels of investment or surplus.

To see this, we focus on backward integration (i.e., ownership of the upstream asset). Let A be an indicator of asset ownership with $A = 1$ denoting manufacturer ownership of i_S (vertical integration) and $A = 0$ denoting supplier ownership (non-integration or market transaction). Manufacturer and supplier disagreement payoffs (i.e., their payoffs in the next best alternative to trading with each other) are[40]

$$w_M(i|A) = (\mu_0 + \mu_{M0} i_M)(1 - A) + (\mu_1 + \mu_{M1} i_M)A, \tag{13}$$

and

$$w_S(i|A) = (\sigma_0 + \sigma_{S0} i_S)(1 - A) + (\sigma_1 + \sigma_{S1} i_S)A, \tag{14}$$

[38] For a formal treatment of the underinvestment result, see Grout (1984). Note, however, that overinvestment can also occur, as, for example, when influence activities are involved. Moreover, in very simple situations, contracts that specify bargaining strengths can overcome the underinvestment problem. To illustrate, if only one party must invest, the problem is overcome by giving all bargaining power to the investing party. However, when the situation is more complex, such as when both parties must invest, this simple solution does not apply.

[39] The fact that ownership only matters when parties disagree is an assumption of the Hart (1995) model. However, in a different context, Baker, Gibbons, and Murphy (2002) derive this result endogenously.

[40] Whinston allows $w_j, j = S, M$ to depend on both investments. We adopt instead the Hart (1995) assumption that disagreement payoffs depend only on own investment as this simpler model is sufficient for our purposes.

where the subscripts 0 and 1 indicate that an outcome is associated with supplier or manufacturer ownership of i_S respectively (i.e. $A = 0$ or 1). Assume further that $\alpha_M > \mu_{M1} > \mu_{M0} \geq 0$, and $\alpha_S > \sigma_{S0} > \sigma_{S1} \geq 0$. In other words, assets are most productive when an agreement is reached (i.e., they are specific). However, when no agreement can be struck, M's asset is more productive when she owns both assets (integration), while S's asset is more productive when he owns it (separation).

With the second–best situation, we assume, as is typical in this literature, that the Nash bargaining solution with side payments is used in the bargaining game with (w_M, w_S) as the threat point. As is well known, this is equivalent to each player receiving his threat payoff plus one half of the gains from trade. Since there are gains from trade, a bargain will always be struck. Nevertheless, despite the fact that players never receive their threats, a distortion in investment decisions occurs because each party's objective function assigns positive weight to his threat. *Ex ante*, players therefore choose their threats strategically, to better position themselves in the *ex post* game.

Second–best investment for the manufacturer is given by[41]

$$i_M^*(A) = \frac{1}{2}[\alpha_M + \mu_{M0}(1-A) + \mu_{M1}A] < i_M^{**}. \tag{15}$$

The solution for the supplier's investment is similar. Finally, the second–best *ex ante* surplus, $W^*(A)$, is obtained by substituting second-best investments into $W(\cdot)$. At time $t = 0$, the manufacturer decides whether to vertically integrate or not by comparing these values.

To obtain a model of vertical integration that is also an estimating equation, we again append an unobserved (by the econometrician) zero–mean random variable, ϵ_A, to the second–best surpluses. Vertical integration will be chosen if

$$W^*(1) + \epsilon_1 > W^*(0) + \epsilon_0, \tag{16}$$

or

$$\Delta = W^*(1) - W^*(0) > \epsilon_0 - \epsilon_1. \tag{17}$$

[41] The manufacturer's objective function is

$$(\mu_0 + \mu_{M0} i_M)(1-A) + (\mu_1 + \mu_{M1} i_M)A + 0.5[\alpha_0 + \alpha_M i_M + \alpha_S i_S - (\mu_0 + \mu_{M0} i_M)(1-A)$$
$$-(\mu_1 + \mu_{M1} i_M)A] - c(i_M),$$

which can be rewritten as $.5[\alpha_0 + \alpha_M i_M + \alpha_S i_S + (\mu_0 + \mu_{M0} i_M)(1-A) + (\mu_1 + \mu_{M1} i_M)A] - i_M^2$. Maximizing this objective by choice of i_M yields the investment level defined by equation 15.

Finally, if $\epsilon_0 - \epsilon_1$ has cdf $F(\cdot)$, the probability of vertical integration is

$$PROB[VI] = PROB[A = 1] = F(\Delta). \tag{18}$$

One can use (18) and the related equations above to derive some testable hypotheses. First, an increase in the marginal return to the manufacturer's (supplier's) investment in the joint surplus, α_M (α_S), makes backward integration more (less) likely.[42]

Second, consider an increase in a marginal return to one party's investment in his disagreement payoff. If that increase occurs under asset ownership A, it makes that form of asset ownership more likely. For example, if the party is the manufacturer, an increase under integration (i.e., in μ_{M1}) makes integration more likely, but an increase under nonintegration (μ_{M0}) makes integration less likely.[43] This occurs because such changes in marginal returns make disagreement under A more profitable while leaving payoffs in other situations unchanged.

Third, if we add a variable x to either the joint surplus (12) or to the disagreement payoffs (13) and (14), it will not affect investment decisions unless it affects some marginal return to investment. Furthermore, when such a variable affects one of those returns, its comparative statics are the same as those outlined above for the marginal return that it affects.

Note the similarities with the moral–hazard model. In both models, increases in the manufacturer's marginal productivity of investment make integration more likely,[44] whereas increases in the other party's marginal productivity make integration less likely. Furthermore, with both models, adding exogenous variables such as characteristics of the parties, the product, or the market makes no difference unless those variables affect marginal productivities.[45] Finally, in both models, some form of underinvestment — be it in terms of effort level or assets — occurs under both integration and separation. Of course, there are also important differences between the models. For example, since there is no risk in this version of the PR model, there are no predictions concerning changes in risk or risk aversion.

In contrast, the implications of the PR model can be quite different from those generated by TC analysis. Specifically, with the PR model, changes that make the gap between π and w_M or w_S larger can be interpreted as increases in quasi rents

[42] This result depends on the assumptions $\mu_{M_1} > \mu_{M_0}$ and $\sigma_{S_0} > \sigma_{S_1}$ (see Whinston (2003, p. 8)).

[43] Whinston (2003) derives other comparative statics for marginal–productivity increases.

[44] Unless the increase is to the productivity of investment under nonintegration.

[45] Or, in the case of the moral–hazard model, risk.

or specificity. TC arguments predict that such increases will make integration more likely. PR theories imply instead that the outcome of such changes in specificity will depend on the source of the increase. For example, an increase in the marginal productivity of the agent's investment in the joint surplus (12), which implies an increase in quasi rent, makes integration less likely in the PR model.

2.2.3 Evidence on Predictions from Transaction–Cost Models

Predictions from transaction–cost models, while rather informal, are still well understood. Indeed, asset specificity generates a flow of quasi rents that are associated with *ex post* haggling and opportunism, whereas complexity and uncertainty lead to contractual incompleteness. Thus, vertical integration is predicted to be more likely when assets are specific, when transactions are complex, and when uncertainty is important.

Following Williamson (1983), it is common to divide asset specificity into four main categories based on the source of the specificity.

- *Physical capital specificity* stems from investments that involve tools or other physical assets that have higher value in their intended use.
- *Human capital specificity* results when individuals undergo training or on–the–job learning that is more valuable inside than outside a relationship.
- *Dedicated assets* are ones that would not be acquired if a specific buyer were not intending to purchase a significant fraction of their product.
- *Site specificity* results from colocation. In other words, the flow of quasi rents is generated by savings in inventory and transport costs.

A fifth type of asset specificity, namely *temporal specificity*, has also received some attention in the literature (Masten et al., 1991; Pirrong, 1993; Williamson, 1991). This type of specificity refers to assets that must be used in a given order, or on a particular schedule, such that their unavailability at a point in time can hold up production.

We use these categories to organize our discussion of the evidence on specificity, and then present the evidence on other factors suggested by the theory, namely complexity and uncertainty. As with the tests of moral–hazard models, the evidence is

summarized in a set of tables, one for each factor of interest.[46] Although we focus on backward integration, we include some studies of forward integration in the tables when those studies test TC predictions.

In principle the notion of asset specificity is fairly straightforward, but the measurement of such a concept, and of other factors influencing the make–or–buy decision according to theory, can be quite problematic. Indeed, publicly available data rarely contain useful information concerning such things as specificity or complexity. For that reason, most studies rely on qualitative data, obtained directly by the author(s) from inspection, or, more frequently, through interviews, questionnaires, or postal surveys of firm managers. In addition, some measures are averages of standardized variables that have been constructed from answers to questionnaires. The measures also vary importantly across studies based on context. In our discussion of the evidence, and in the tables below, we discuss some of the measurement challenges and describe the measures authors have relied upon.

Physical Capital Specificity

Authors have used several different measures of physical specificity, usually tailoring their measure to the context. To illustrate, in some instances the measure is a dichotomous variable that equals one if a respondent thought that physical specificity was important. In other cases, it is an index that ranges from 0 to n, depending on the degree of physical specificity. For example, 0 might correspond to 'relatively standard' whereas n might denote 'design specific.' In still other studies, the measure represents a particular feature of an input. For example, it could be a dummy variable that equals one if an input is a gas (which requires pipelines and storage tanks).

Table 7, which summarizes the evidence concerning physical–capital specificity, shows that its effect on vertical integration is always positive and usually significant. In other words, consistent with the predictions of TC analysis, the presence of this sort of specificity makes integration more likely.

Human Capital Specificity

The most common measure of human–capital specificity involves some notion of the amount of training that is required to produce or use an input. In some cases, it is

[46] We focus, in the tables, mostly on papers published in economics. TCE-based research published in marketing and in management or strategy journals is quite voluminous, and though we discuss some of these studies, a complete overview of this literature is beyond the scope of the present article. We also do not include the many articles on the GM-Fisher Body case as these are mostly qualitative.

a direct measure of training. In others, however, it might be a measure of engineering design cost, which is a proxy for the amount of technical know how that must be acquired. This means that the measure can also be a proxy for complexity, and it is difficult to disentangle the two effects. For this reason, in some cases we present the same finding in both tables.

Table 8 summarizes the evidence concerning human–capital specificity. It shows that, with one exception, the effect on vertical integration is positive and significant, which is evidence in favor of the TC model. It is interesting to note that the single negative effect, which is obtained by Woodruff (2002), involves forward integration into retailing. The other studies involve either backward integration with suppliers or forward integration of an industrial sales force.

Dedicated Assets

A few researchers have examined the effect of asset dedication on vertical integration, and the measures used, such as the downstream firm's share of purchases or a dichotomous variable that equals one if only one firm buys the input, are relatively straightforward. Table 9, which summarizes this evidence, shows that when assets are dedicated, vertical integration is more likely. These findings are also supportive of TC arguments.

Site Specificity

The importance of site specificity or colocation has also been assessed. Sometimes the specificity measure is a qualitative scale variable constructed from answers to questions concerning the importance of proximity. In other situations it is a 0/1 variable that equals one if the two facilities are located close to one another (for example, if a plant that generates electricity is located at the mouth of a coal mine that supplies its fuel).

The evidence concerning site specificity, which is summarized in table 10, is not very conclusive. However, the only significant effect on vertical integration is positive, which is consistent with TC predictions.

Temporal Specificity

The importance of temporal specificity, or the need to integrate transactions that can delay other aspects of production, has also been assessed. The measures used include an index that captures how important it is that a given component be available on schedule, or indicators of how atypical a firm's needs may be. The argument behind these measures is that the firm is likely to have difficulty finding alternative supplies at

the last minute, i.e. these input markets are thin and thus more subject to potential hold up by suppliers.

The evidence concerning temporal specificity, as summarized in table 11, shows that vertical integration is more likely when alternative timely sources of supply are likely to be rare, a result that is consistent with TC predictions.

General Specificity

Some empirical tests are not designed to identify the precise nature of specificity. Instead they test for its presence in more general terms. To illustrate, one study (Weiss 1992) assesses residual correlation of share–price returns, under the hypothesis that, when specificity is important, shocks to one firm will affect the other in the same direction. Another study (Ciliberto 2005) assesses how health maintenance organizations (HMOs), which tie physicians to hospitals, affect integration decisions. Finally, González-Diaz *et al.* (2000) examine how the extent of subcontracting by firms in the construction industry relates to the specificity of their product offering, measured by some weighted sum of the other firms offering the same product in the same market each period. Those studies, as summarized in table 12, also show that vertical integration is more likely when assets are specific, as predicted.

Complexity

All forms of specificity are associated with quasi rents that can lead to disputes as each party to a transaction attempts to appropriate those rents. However, problems would not occur if contracts were complete. The next two factors, asset complexity and transaction uncertainty, exacerbate the problem because they increase the difficulties that are associated with writing complete contracts. Although it is the interaction of the two groups of factors — specificity and contractual incompleteness — that is important, the effect of each factor has usually been considered on its own. We follow the literature in this respect and present the effect of each factor separately here and in the tables. For reasons that will become clear later, we postpone our discussion of the few cases where authors have examined the interaction between specificity and complexity — or uncertainty — to the next section.

As with specificity, complexity measures are often based on qualitative information that has been collected through interviews or surveys. For example, respondents might be asked to rank the complexity of an input on a scale from 1 to n, or some notion of design cost or product heterogeneity might be constructed. Moreover, quantitative measures, such as R&D intensity and renegotiation frequency, have also been

used. Table 13, which summarizes the evidence concerning complexity, shows that, with one exception, its effect on vertical integration is both positive and significant. The exception (Acemoglu, *et. al.* 2005) finds that supplier R&D intensity in UK manufacturing is associated with less integration. As before, however, most of the evidence is supportive of the importance of transaction costs.

Uncertainty

Uncertainty also increases the difficulties that are associated with, and lessens the desirability of, writing complete contracts. There are many measures of uncertainty and most can apply to either up or downstream products or markets. For example, uncertainty can be proxied by the variance of sales or of forecasting errors or by the instability of shares in either market. It can also be captured by an indicator of the frequency of specification or design changes for inputs or outputs.

In table 14, which summarizes the evidence from studies of backward integration that have considered this factor,[47] a U (D) indicates that the measure of uncertainty applies to the upstream or supplier (downstream or buyer) market. This table shows that, whenever the effect is significant, higher uncertainty leads to more vertical integration. Furthermore, this conclusion is independent of the market in which the uncertainty occurs. The evidence is therefore consistent with TCE predictions.

Summary

The weight of the evidence is overwhelming. Indeed, virtually all predictions from transaction–cost analysis appear to be borne out by the data. In particular, when the relationship that is assessed involves backward integration between a manufacturer and her suppliers, there are almost no statistically significant results that contradict TC predictions.

2.2.4 Evidence on Predictions from Property–Rights Models

In contrast to the abundance of work that attempts to assess the validity of the predictions from moral–hazard and transaction–cost models, there are very few studies that deal directly with property–rights predictions. This is perhaps due to the fact that the PR models are newer and their predictions are more fragile. Moreover, as noted earlier, many researchers make little distinction between TC and PR models and interpret tests of one as tests of both. Nevertheless, as Whinston (2003) and our

[47] We consider only backward integration here to distinguish this table from table 1.

discussion above have shown, the predictions from the two classes of models can be very different.

A quick reading of tables 7–14 might lead one to conclude that, at least when TC and PR model predictions do not agree, the evidence is not supportive of property–rights theories of vertical integration. Indeed, the level of specificity and quasi rents of any form appear to foster vertical integration. However, this need not be the correct conclusion to draw from the data.

First, the predictions from the PR and TC model agree in some cases, and there the evidence supports both. More importantly from our perspective, there are two cases where the data contradict transaction cost arguments. The first (Acemoglu *et. al.* 2005) finds that, whereas downstream product complexity leads to more integration, technological intensity upstream has the opposite effect, a finding that is very supportive of PR models. It is also interesting that the other significant evidence that is at odds with TC analysis comes from a study of integration between manufacturing and retailing (Woodruff 2002), which is the typical setting of tests of moral–hazard models. As we have shown, the predictions from PR and MH models have much in common. It might therefore be possible to learn something about the empirical relevance of PR models from the regularities that surfaced in studies of manufacturer/retailer or franchisor/franchisee relationships.

In particular, table 2 shows that, in all cases where the importance of the agent's effort is a significant determinant of integration, it leads to less, not more integration. With the MH model, effort is normally interpreted as the marginal productivity of the agent's work. However, the variables that have been used to measure effort, such as the need for personalized service or previous experience dummies, can equally be viewed as measures of the marginal productivity of the agent's human–capital investment. Moreover, as Lutz (1995) points out, a franchisee differs from a company manager not only because he is a residual claimant, but also because he owns the future profits of the outlet and the right to sell that outlet, at least as long as he satisfies the constraints that are imposed by the franchisor. Hence a franchisee's effort in making the business successful can be seen as an investment. To the extent that that investment is specific, TCE implies that it will be associated with more vertical integration. Thus, not only is the evidence in table 2 supportive of the moral–hazard model, but also it supports property–rights *vis–à–vis* transaction–cost arguments (see e.g. Woodruff (2002) for more on this).

We can also learn something about the relevance of PR models from an exam-

ination of the effect of risk on vertical integration. The empirical evidence that is summarized in table 1 appears to contradict a fundamental prediction of the moral–hazard model. Indeed, more downstream risk leads to less, not more, integration and therefore less insurance for the agent. This finding, which is puzzling in the context of moral–hazard theories, can be explained using property–rights arguments. Specifically, when downstream risk increases, the agent's ability to be flexible in the face of unforeseen contingencies becomes more important. For example, when faced with unpredictable fluctuations in sales, the retailer needs to be more skillful at managing inventories and employees; or, when customer tastes become more fickle and style changes more frequent, he needs a better understanding of customer needs. In other words, in the presence of increased uncertainty, his investments become more productive and PR models predict less integration as a consequence.[48]

Finally, a notion that is central to both TC and PR models but is more strongly emphasized by PR theory is that, in the absence of contractual incompleteness, problems should not surface. For example, there is no reason to believe that simply because a part is used by only one firm (a dedicated asset), it must also be noncontractible. Given the importance of the interaction between specificity and contractual incompleteness in the theory, it is surprising that it is rarely tested directly. Exceptions include Masten (1984), who assesses the interaction of specificity and complexity, and Anderson (1985) who considers the interaction of specificity and environmental uncertainty. It is surprising that the findings from the empirical TC literature are so robust given that the effect of each factor is considered independently. Still, in those cases where interaction effects were considered, authors found support for the theories.

Summary

Although property–rights models have been around for two decades, empirical testing of predictions derived from those models lags behind. Nevertheless, as argued above, we can glean some insights into the validity of PR theories through a reinterpretation of tests of MH and TC predictions. In particular, the evidence that comes from supplier/manufacturer relationships, which is the typical setting of TC tests, is not very supportive of PR arguments, at least when the two sets of predictions disagree. However, the evidence that comes from manufacturer/retailer or franchisor/franchisee relationships is much more positive. Not only is it consistent with many PR predictions, but also PR ideas provide insights into and suggest a so-

[48] This explanation relates also to Prendergast's (2002) argument.

lution to a puzzle that surfaces in the MH literature, namely the negative relationship between risk at the retail level and vertical integration. Still, much further work is needed before the relative lack of direct tests of PR predictions can be adequately addressed, and, perhaps more importantly, the potential for cross-fertilization among tests of different models can be fully realized.

2.3 Some Econometric Issues

We have thus far ignored econometric issues, most of which are not unique to the studies that we summarize. Nevertheless, many of the problems that arise in the literature that we survey are related to the discreteness of the choices that firms make. We therefore conclude this subsection with a brief discussion of some of those issues.

Many empirical studies that assess the incidence of vertical integration use transaction or outlet–level data. For example, one might have observations on franchise chains, each of which has many retail outlets, and be interested in modeling whether an outlet is operated by the franchisor or by a quasi–independent franchisee. In that case, the dependent variable in the estimating equation is discrete. Methods of dealing with discrete dependent variables are well known.[49] There are, however, a number of problems that are apt to surface in discrete–choice studies of the sort that we have in mind, problems whose solutions are more complex than when the dependent variable is continuous.

First, there is the ubiquitous endogeneity problem — this problem is endemic in empirical research in industrial organization and is compounded by the absence of valid instruments. To illustrate, firm age and size are to some extent the result of managerial decisions that can be based on an underlying factor that can also lead them to integrate a particular transaction. Similarly, in the studies that we discuss below, other outlet characteristics are included among the 'exogenous' explanatory variables that determine the method of transacting between manufacturer and retailer. However, when an upstream firm decides to change the nature of its relationship with a retailer, she might decide to change some of the outlet's characteristics and *vice versa*. For example, this is often the case with gasoline retailing — stations that are changed from full to self service are often changed from independent dealer to company operation at the same time. A simple method of overcoming the endogeneity problem is

[49] For example, see Wooldridge (2002, chapter 15).

to estimate a linear probability model by two–stage least squares.[50] Unfortunately, the linear–probability (LP) model has other undesirable features.[51] Other solutions to the endogeneity problem in the presence of limited dependent variables normally require strong assumptions (see, e.g., Wooldridge (2002, pp. 472–477)). Moreover, when the endogenous explanatory variable is itself binary, further complications arise. These issues make structural estimation methods particularly appealing. As we discuss in the next section, however, those methods suffer from several limitations of their own, especially in the context of empirical studies of vertical relationships.

Second, errors in a cross section, the type of data that one must often rely on in this type of study, are apt to be heteroskedastic. For example, outlets can be of very different sizes, which normally induces heteroskedasticity. OLS estimates in the presence of heteroskedasticity are inefficient. With a probit, in contrast, they are inconsistent. Indeed, heteroskedasticity changes the functional form for PROB($y = 1|x$), which is no longer normal. As before, the simplest remedy is to estimate a linear probability model with a correction for heteroskedasticity. However, if the true model is a probit, the LP estimates will still be inconsistent.

Finally, the errors in a discrete–choice model are apt to be spatially correlated in the sense that the off–diagonal entries in the variance/covariance matrix at a point in time are nonzero.[52] For example, outlets that are located in the city center might experience common shocks that are not experienced by ones that are located in the suburbs; or outlets that sell brands of a common manufacturer might have common private information. One possible remedy is to use the correction for spatial and time–series correlation of an unknown form that is developed in Pinkse, Slade, and Shen (2005).[53]

Many of the studies that we present suffer from one or more of these types of problems, as well as major measurement problems as described above, and the extent to which authors have tried to address these problems varies importantly across studies. In the end, however, we believe that a preponderance of evidence, garnered across numerous studies using different approaches in various institutional and industry contexts, is most apt to yield convincing evidence on the validity of various theories. We offer our summary of the evidence in that spirit.

[50] This method is simple provided that valid instruments can be found. However, the problem of finding instruments is just as acute here as in the continuous–choice situation.

[51] For example, it is usually not possible to constrain PROB($y = 1|x$) to lie between 0 and 1.

[52] We use the term spatial to denote either geographic or characteristic space.

[53] This correction is similar to the one developed by Newey and West (1987) in a time–series context. Note that the spatial procedure also corrects for heteroskedasticity.

3 The Consequences of Vertical Integration

Having considered factors that can lead firms to integrate vertically, we now turn to an examination of the empirical evidence concerning the effects of vertical integration on economic outcomes such as prices, profits, quantities, and costs. We do this because such evidence can shed light on two major questions: first, can we identify the purported benefits from vertical integration in a firm's profits or its choices of prices and quantities? Second, if we can find evidence of those benefits, can we identify the winners and losers? The first question is important because it can shed light on the circumstances under which vertical integration is apt to benefit the firms involved, and thus lead to better informed decision making. Furthermore, answers to both questions are important because they can help us distinguish cases where vertical integration is more apt to benefit consumers from those where it is more likely to be detrimental. Such an understanding is an important precondition for the design of sensible public policy towards vertical mergers and other forms of vertical–market transactions.

Why there should be any public–policy debate about vertical integration is unclear from our discussion so far. The motives for vertical integration that are associated with the theories that we have presented emphasize that, when firms choose vertical integration, it is efficient for them to do so. Moreover, by highlighting the importance of the different efficiency motives, the empirical evidence that we have reviewed suggests that vertical–merger policy should be *de minimus* if it exists at all. After all, both firms and consumers can benefit when firms realize efficiencies. Yet in reality, attitudes towards vertical relations and mergers have undergone important reversals in antitrust–policy circles, being sometimes restrictive and other times permissive.

Consider, for example, the history of the US Department of Justice's (the DOJ's) position towards vertical mergers. The first DOJ merger guidelines, which were published in 1968, were relatively hostile towards vertical integration. Indeed, they viewed with suspicion vertical mergers between firms that accounted for as little as 10% of their respective markets. The replacement guidelines, published in 1982 and 1984, in contrast, regarded non–horizontal mergers to be of interest for antitrust policy only if they had substantial horizontal consequences. Finally, the 1992 guidelines were renamed 'Horizontal Merger Guidelines,' as vertical mergers were essentially forgotten.[54]

[54] See Pitofsky (1997) on this, and Warren–Boulton (2003) for a more in–depth discussion of the history and interpretation of the DOJ Guidelines.

The existence of, at times stringent, controls on vertical mergers suggests that there must exist motives for vertical integration that are not as innocuous as the efficiency arguments we have discussed so far. And indeed this is the case. In this section, we briefly discuss this alternative set of motives for vertical integration, all of which have to do with the creation and exploitation of market power. As should become clear shortly, however, not all motives for vertical integration that are associated with the exercise of market power imply that vertical mergers are necessarily detrimental to consumers. In fact, efficiencies can be generated when firms integrate to, for example, eliminate double margins or input–choice distortions. The theories of vertical mergers that antitrust authorities are most concerned with in reality focus on the horizontal aspects of the merger, namely exclusion and collusion. In other words, when a manufacturer operates in an imperfectly competitive market, her interactions with her competitors – and in particular her capacity to collude with or exclude her rivals – can provide additional motives for vertical integration or separation.

Not surprisingly then, authors have looked for detrimental effects from vertical mergers mostly in concentrated markets (e.g., cable TV). As we will see below, however, even though authors typically choose markets where they expect to find evidence of exclusion, half of the studies find no sign of it. And where they find evidence of exclusion or foreclosure, they also at times document efficiencies that arise from the same merger. Thus, although foreclosure may occur some of the time, the end result is not necessarily detrimental to consumers. In fact, consistent with the large set of efficiency motives for vertical mergers that we have described so far, the evidence on the consequences of vertical mergers suggests that consumers mostly benefit from mergers that firms undertake voluntarily. On the other hand, divorcement requirements, which are separation requirements that are imposed by local authorities, often to protect local dealers, typically lead to higher prices and lower service levels for consumers. In other words, consumers are often worse off when governments require vertical separation in markets where firms would have chosen otherwise.

We begin by presenting the various arguments for vertical mergers that have given rise to public–policy concerns, which are the more traditional motives for vertical integration that arise in contexts where firms have market power. Since there are many such motives and many variants of each model, we review the arguments only briefly and do not present formal models. We then discuss some methodological issues that are particularly severe in this context and the methods that have been used to identify the effects of interest. Finally, we present the evidence on the effects of

vertical integration on economic outcomes, evidence that sheds light on the motives behind vertical integration and ultimately should inform public–policy decisions in the area of vertical mergers.

3.1 Market-Power Based Theories of Vertical Integration

Double Marginalization

Double marginalization occurs when there are successive stages of monopoly (or oligopoly) and the firms at each stage are not vertically integrated (Spengler (1950), Greenhut and Ohta (1979)). Unintegrated firms ignore the reduction in profits that they inflict on other stages of production when their prices increase, whereas vertically integrated firms capture that externality. As a result, prices are lower under integration.[55]

The name double marginalization refers to the fact that monopoly profits are extracted at each stage of production (e.g., there are multiple margins applied, each time to raise price above marginal cost).[56] Under vertical integration, in contrast, there is a single marginalization. Indeed, the vertically integrated monopolist maximizes the joint surplus, up and downstream, as this maximizes her profits. Furthermore, consumers also are better off under integration in this case, as they pay less for the product than under successive markups.

This situation is analogous to one in which two goods are perfect complements in downstream production or in ultimate consumption. With the latter situation, integration between the two producers also results in lower prices for consumers and higher joint profits as the pricing externality is internalized.

Variable Factor Proportions

When inputs are used in variable proportions, specifically, when they are substitutes, an upstream monopolist selling one of the inputs can have an incentive to integrate forward to prevent downstream firms from substituting away from her product. In particular, suppose firms in two upstream industries supply inputs to a competitive downstream industry. If the industry that supplies x_1 is monopolized, the monopolist will set a price for x_1 that exceeds marginal cost. Since the inputs

[55] We are assuming a fixed–proportion technology. See our discussion below on how ambiguities arise under variable proportions.

[56] As the number of marginalizations increases without bound, profits and sales go to zero.

are substitutes, downstream firms will use too much of the competitively supplied input, x_2, and too little of the monopolist's input, relative to the situation that would occur if both were sold at marginal cost. In other words, the downstream firms will substitute x_2 for x_1. This production inefficiency gives the upstream monopolist a motive for acquiring the downstream firms. Indeed, in so doing, x_1 can be transferred internally at marginal cost, and the inputs can be used in the correct proportions.

Superficially, it might seem that, in the variable–proportions situation, vertical integration makes everyone better off. However, that need not be the case. The problem is that the monopoly distortion persists; it has simply been moved from up to downstream. The effect of integration on the price of the final product is ambiguous and will depend on the parameters of the problem.[57]

Foreclosure and Raising Rival's Costs

Foreclosure occurs when practices are adopted that reduce buyers' access to suppliers (upstream foreclosure) or sellers access to buyers (downstream foreclosure). Foreclosure is an important concept. Indeed, the main worry of antitrust authorities when it comes to vertical relationships is the possibility that integration will foreclose entry by competitors at some level of the vertical chain, will cause competitors to exit, or will disadvantage them in some manner. For example, a manufacturer who acquires a large network involving most retailers might prevent competitors from gaining access to customers at reasonable cost, if at all. This in turn could prevent entry of potential competitors upstream, or perhaps even lead rivals to exit the upstream industry.

Early theories of the problems associated with foreclosure were not based on rigorous models. Furthermore, Chicago–School economists (e.g., Liebeler (1968), Bork (1969), and Peltzman(1969)) argued that those theories were spurious and that vertically integrated firms have no incentive to foreclose since they can achieve the same outcome whether or not they integrate. They thus concluded that vertical integration can have no pernicious effect.

Later, economists began to model the vertical–merger/foreclosure issue in a strategic setting (see, e.g., Salinger (1988), Hart and Tirole (1990), and Ordover, Saloner, and Salop (1990)). For example, Salinger (1988) shows that the effect of a vertical merger on prices in an industry with Cournot oligopolists at each stage is ambiguous.

[57] See, e.g., Schmalensee (1973) and Warren–Boulton (1974). See also Klein and Murphy (1997) for a different type of variable proportion problem that can give rise to vertical integration, namely one that operates through dealer services that are complementary to the manufacturer's product.

Indeed, there are two opposing forces at work: first, a merger can raise the costs of unintegrated downstream firms, a factor that can cause retail prices to rise; and second, a merger can eliminate double marginalization that existed in the pre–integrated situation, a factor that can cause retail prices to fall.

The Salinger (1988) model also demonstrates that vertical mergers can be beneficial to manufacturers even if the integrated manufacturer does not refuse to sell or completely foreclose access to facilities to unintegrated producers. In fact, it is often advantageous to simply raise rivals' costs. The incentive to raise the costs of unintegrated downstream competitors is easy to see. An increase in the wholesale price to a downstream competitor will cause that rival's retail price to rise, which will lead some of the rival's customers to switch to the integrated firm's retail facilities. This point is the focus of several papers on raising rivals' costs (see, e.g., Salop and Scheffman (1983) and Krattenmaker and Salop (1986)). In these models, in the absence of double marginalization in the unintegrated situation (e.g. if manufacturers use two–part tariffs), vertical mergers will result in increased prices to consumers.

Strategic Delegation and Collusion

When an industry is oligopolistic, vertical separation is often modeled as a two–stage game in which contracts are written in the first stage (wholesale prices w and fixed fees F are set), and retail prices (p) are chosen in the second. This setup implies that, if rival contracts can be observed, downstream agents will condition their retail–price choices on those contracts. Under vertical integration, in contrast, the product is transferred internally at transfer prices that normally cannot be observed by rivals. The integrated situation is therefore usually modeled as a one–shot game.

The idea that upstream firms can soften the intensity with which they compete by delegating the pricing decision to independent retailers is by now well understood. The models in this case focus on interbrand competition across vertical structures. When vertical structures compete directly with each other (i.e. when manufacturers set retail prices themselves as they do under vertical integration), the resulting Nash–equilibrium prices are lower than joint–profit–maximizing prices. However, when manufacturers sell to retailers who have some market power, if manufacturers delegate the pricing decision to those retailers (as they typically do under vertical separation), the equilibrium prices that result are higher than under integration (see, e.g., Rey and Stiglitz (1995)). A softening of competition occurs because prices are normally strategic complements (i.e., price reaction functions normally slope up). An increase in a manufacturer's wholesale price is therefore associated not only with higher own–

dealer prices, but also with higher competitor retail prices. Furthermore, with two–part tariffs, equilibrium prices will not exceed monopoly prices.[58]

The above argument is premised on the assumption that retailers or distributors have market power. Spatial separation is one — but not the only — factor that can lead to pricing power. The argument also relies on the assumption of price competition at the retail level, which is apt to be valid in most contexts. However, if downstream firms engage in quantity competition, delegation will not benefit the vertical chain.[59]

It is straightforward to show that, under agent risk neutrality, delegation of the pricing decision (vertical separation) is a dominant strategy. However, as risk or risk aversion increases, the advantages of delegation fall. This occurs because a higher retail price is accompanied by an increase in the proportion of the retailer's income that is variable, thereby increasing the risk that he must bear. At some level of risk and/or risk aversion, the retailer's need for compensation for bearing increased risk makes vertical separation unattractive, and the firm chooses to vertically integrate instead. On the other hand, the more substitutable are the products of rival retailers, the more the firms benefit from delegation (separation), and thus the more likely it will be chosen. Overall then, vertical integration will be preferred when there is substantial risk or risk aversion and products are not highly substitutable.[60]

Backward Integration by a Monopsonist

The above theories deal with forward integration by a monopolist. However, backward integration by a monopsonist can also occur, a situation that is considered by Perry (1978). In his model, the vertical chain consists of an upstream competitive industry that produces an input under conditions of increasing marginal cost due to the presence of a fixed factor. The competitive industry thus earns rent. The input is purchased by a monopsonist who would like to capture the upstream rent. However, in order for there to be a motive for backward integration, it must be the case that the scarce input is worth more to the monopsonist than to the competitive firms.

This motive is similar to the incentive for forward integration in the variable–proportions case. In particular, the monopsonist's incentive for backward integration can stem from the desire to internalize the efficiency loss that is due to under utilization of the input whose supply is upward sloping. There is, however, also a rent effect

[58] In the absence of fixed fees, delegation can lead to higher upstream profits but it is not guaranteed to do so.

[59] This is true because quantities are strategic substitutes.

[60] For a formal model that embodies all of these features, see Lafontaine and Slade (2001).

that enables the monopsonist to reduce the sum of rent payments to independent suppliers plus the costs of acquiring integrated suppliers.

With Perry's model, full backward integration eliminates efficiency losses due to monopsony behavior and lowers prices to consumers. With certain acquisition–cost functions, however, the monopsonist will not choose to integrate fully.

Price Discrimination

The last imperfectly competitive motive we discuss, price discrimination, can best be explained with the use of a simple stylized model. Consider an upstream monopolist that supplies an input to two competitive downstream industries with different elasticities of demand for the input.[61] If arbitrage is not possible, that is if the input cannot be purchased in one downstream market and sold in the other, the monopolist will be able to price discriminate. In this case, she will charge a higher price to the industry with the less elastic demand. However, if arbitrage is possible, a single price will prevail, and the monopolist's profit will be lower than in the no–arbitrage situation.

To remedy this problem, it suffices for the monopolist to acquire the buyers with more elastic demands and to suppress that market, say market one. Since customers in market one buy at a lower price, absent integration, arbitrageurs will purchase the input in that market and sell it to the buyers in market two, who are willing to pay more. Vertical integration suppresses the low–price market, which is the one that is causing the monopolist's problem, and enables her to engage in successful price discrimination.

As with most imperfectly competitive motives, the outcome for ultimate consumers under the price–discrimination motive is ambiguous. Indeed, relative to a uniform price for the input, consumers in market one pay lower prices under integration, whereas consumers in market two pay higher prices.

Summary

There are several general conclusions that can be drawn from our discussion of imperfectly competitive motives. First, there are few unambiguous results. Ambiguity in the theories makes an analysis of the data even more important. Second, even when the motive for a merger stems from imperfect competition in horizontal markets, vertical mergers can be unambiguously beneficial. Such is the case when the merger motive is to eliminate double marginalization under successive stages of

[61] See, e.g., Gould (1977).

monopoly. Third, it is always the link in the chain that has market power, whether it be monopoly or monopsony power, that benefits from integration. Thus absent market power at some stage in the chain, the above motives cannot be relied upon to explain the data. Finally, in many cases theory suggests that firms with market power are able to obtain the same results with various forms of vertical restraints rather than integration. For example, firms can eliminate double marginalization with two-part tariffs, maximum resale prices, or quantity forcing. They can also either achieve foreclosure or address variable–proportions issues via tying. To the extent that firms have these alternative mechanisms at their disposal to address the issues that are raised by the models, it is unclear whether researchers can expect to find evidence that firms opt for vertical integration to solve the problems that those models describe. Put differently, there would be no way to identify the consequences of vertical integration if non-integrated firms could achieve similar results with contractual restraints. On the other hand, if firms do use vertical mergers rather than contracts to foreclose rivals or facilitate collusion, it is important to recognize that public policy aimed only at preventing vertical mergers would prove ineffective as it would simply lead firms towards those alternative mechanisms.

3.2 Some Methodological Issues

We have already mentioned some of the important econometric problems that authors confront when conducting research into the factors that drive the vertical–integration decision. The same problems arise in studies of the effects or consequences of vertical integration. Furthermore, the issues are more serious in this literature as vertical integration decisions are clearly endogenous in analyses of the consequences of such decisions. This problem is compounded here as in other areas of empirical IO by the difficulty of obtaining valid instruments. These difficulties partly explain the appeal of structural estimation methods in the study of firm behavior. As we discuss below, however, those methods suffer from limitations of their own in the context of empirical studies of vertical relationships.

The most straightforward way to evaluate the effects of vertical mergers or divestitures is to present some persuasive descriptive statistics. For example, one can compile information on retail prices before and after such mergers. Descriptive statistics are useful in so far as they convince the reader that there is an empirical regularity that needs to be explained. The obvious problem, however, is that there can be many explanations for that regularity. For this reason, most researchers combine descriptive

statistics with some form of econometric analysis. In what follows, we describe some of the main methods that have been used and their principal limitations.

3.2.1 Cross–Section, Time–Series, and Panel Methods

For the purpose of presentation, assume that the data consist of a set of firms in a given industry, where a subset of the firms is vertically integrated (v) and another subset transacts with independent suppliers or retailers in a market (m). We are interested in the consequences of that difference for some measure of average performance y (e.g., profits, sales, prices, or costs). We call the first set of observations the treatment group and the second the control group. In other words, we think of vertical integration as a treatment that the firms undergo. This is a classic example of policy evaluation, and the techniques that we describe are used to assess many different policy issues.

Let Δ be the difference in average performance that we wish to measure. Our ideal measure would be

$$\Delta^* = y_{v,T} - y_{vm,T}, \tag{19}$$

where T is some time period, $y_{v,T}$ is the average performance of v firms in that period, and $y_{vm,T}$ is the hypothetical average performance that the treated observations would have experienced had they not been treated (had they been m firms). Unfortunately, we cannot observe $y_{vm,T}$ and must use some proxy to measure Δ. Ultimately, the most useful data set will be a panel that includes both cross–sectional and time–series variation in firm organization. In other words, v and m firms can coexist at a point in time, and some of those firms may change their method of transacting with suppliers or retailers over time.

There are many panel–data methods for estimating treatment effects.[62] We focus on one here — a difference–in–difference (DD) estimator. Suppose that in period T_1 all firms are m types. However, a subset of those firms (m_1) undergoes an organizational change between periods (i.e., they become v firms), whereas the remaining firms (m_2) remain untreated. With a DD estimator, Δ is approximated by

$$\Delta_{DD} = (y_{v,T_2} - y_{m_1,T_1}) - (y_{m_2,T_2} - y_{m_2,T_1}). \tag{20}$$

The first difference, $y_{v,T_2} - y_{m_1,T_1}$ in equation (20) measures the change in the average performance of the treated firms, whereas the second difference, $y_{m_2,T_2} - y_{m_2,T_1}$, measures the change in average performance among the untreated. Finally, the difference

[62] See, e.g., Wooldridge (2002, chapter 18) for a general discussion.

in difference measures the relative change — the amount by which the performance changes differ across the two groups. Since the DD estimator removes both firm and time–period fixed effects, it is common to attribute the final difference to the treatment. Of course, average performance is usually measured conditional on a vector of explanatory variables. With a DD estimator, however, only explanatory variables that differ over both firms and time are relevant.

More generally, an advantage of panel data is that it is possible to include both firm and time–period fixed effects in the estimating equation. The firm dummies remove the influence of firm characteristics that are time invariant, whereas the time dummies remove the influence of factors that are common to all firms at a point in time. Furthermore, if a potential endogeneity problem arises due to unobserved characteristics (common causal factors) that differ by firm (over time) but not over time (by firm), the firm (time) fixed effects will purge the equation of that problem.

Unfortunately, in many of the empirical contributions discussed below, researchers did not have access to panel data. They then either exploited a time-series data set and performed a before–and–after study, or more times than not, they used a cross–sectional data set to exploit the variation in organizational form across firms. The obvious problem with using only time-series variation is that many other things can change between time periods in addition to the vertical structure of the firms. Similarly, in purely cross–sectional data, the set of firms that are vertically integrated and those that are not are not random draws from an underlying population. Indeed, the method of organizing transactions is usually an endogenous choice. In both time-series and cross-sectional data, the endogeneity problem can be partially overcome by including a vector of control variables (e.g., variables that measure demand and supply conditions in a time series, and variables that measure firm characteristics in a cross section).[63] Unfortunately, it is rarely possible to obtain data on all relevant control variables.

In some cases, however, the organizational choice comes from outside the vertical structure, and thus can be considered exogenous. This might be the case, for example, when certain vertical arrangements are prohibited by law (see our discussion of divorcement in subsection 2.1.2) and there is time–series or cross–sectional variation in legal practice. Although the fact that the firms did not choose their organizational form in such cases mitigates the endogeneity problem, it does not resolve the issue entirely. Indeed, the firms might be targeted by the law exactly because the problem

[63] When control variables are included in time–series and other models, the researcher compares conditional means.

that the government agency was trying to remedy was thought to be more acute for them.

3.2.2 Event Studies

In our discussion of the evaluation of the effects of vertical–structure decisions so far, we have been concerned with the realized consequences of changes. One can also estimate market forecasts of the effects of those changes (e.g., mergers and divestitures) on firm value. The tool that is commonly used to perform such evaluation is an event study, which requires that the firms whose data are used in the analyses be publicly traded.

An event study is based on the assumption that stock markets are efficient and that share prices reflect all currently available information at every point in time. In other words, it is assumed that the current stock price equals the expected value that accrues to the holder of the share — the expected discounted stream of capital gains and dividends — where expectations are formed efficiently and rationally. With efficient markets, when a 'surprise' occurs, the associated change in the share price is an estimate of the expected value of the change in that flow.[64]

It is common to base an event study on the Sharpe (1963) market model that relates the return on asset i in period t, R_{it},[65] to the market return, R_{Mt}, where the market return is the return on a broad portfolio of traded assets,[66]

$$R_{it} = \alpha_i + \beta_i R_{Mt} + u_{it}, \qquad i = 1, \ldots, n, \qquad t = 1, \ldots, T. \tag{21}$$

When using the market model to assess an event such as a merger, however, it is important that the event be a 'surprise.' Unfortunately, it is often the case that news of an impending merger leaks out prior to the event. In addition, the market might not react instantaneously to the news. For this reason, it is common to focus on a window that surrounds the event (e.g., the merger). The goal of the analysis is to compare what would have occurred in the event window had the event not taken place to what actually took place, in other words to assess the abnormal returns that are due to the merger.

[64] See MacKinlay (1997) for a general discussion of the use of event studies in economics and finance.

[65] The return on asset i in period t is the capital gain earned plus dividends issued between $t-1$ and t divided by the share price in $t-1$.

[66] The market model can be augmented to include other financial and nonfinancial assets, as in the APT model of Ross (1967).

With this in mind, let $t = 0$ denote the period in which the event occurs,[67] so that $t < 0$ ($t > 0$) denotes time before (after) the event, and let $t = t_1, \ldots, t_2 < 0$ be periods before the event — periods in the estimation window. Periods $t = t_3 \ldots, t_4 \geq 0, t_3 > t_2$, are then chosen to be in the event window.

The estimation proceeds as follows: first, the market model is estimated using observations in the estimation window, i.e. for $t = t_1, \ldots, t_2$. The estimated equation is then used to forecast returns inside the event window, $\hat{R}_{it} = \hat{\alpha}_i + \hat{\beta}_i R_{Mt}, t = t_3 \ldots, t_4$. Abnormal returns are then calculated for observations in the event window, where abnormal returns are realized minus forecast returns, $AR_{it} = R_{it} - \hat{R}_{it}$. One can plot and perform statistical tests on individual abnormal returns. However, more commonly abnormal returns are summed over the observations in the window to find the overall effect of the merger for each firm, and averaged across firms to find the average effect. Obviously, the choice of event window, which can be one or several periods, is crucial here. Indeed, if the window is too narrow, the event can be missed, and if it is too broad, averaging can remove the effect. Finally, standard errors of each estimate can be calculated using well-known formulas.[68]

Clearly, positive (negative) abnormal returns imply that the market values the news as profitable (unprofitable) for a firm that is involved in the merger. However, notice that there are three firms involved, the acquiring firm, the acquired firm, and the merged firm, and it is possible for news to be good for the acquired but bad for the acquiring firms, or *vice versa*.

Event studies can be used to evaluate the consequences of vertical mergers. They can also be used to distinguish between efficiency and anticompetitive motives for mergers.[69] In the case of horizontal mergers, the procedure is straightforward — a merger for market power is good for rivals, whereas one for efficiency is bad. One therefore looks at the effect of the event on rival share prices (see Eckbo (1983)). However, with vertical mergers, things are more complex. Indeed, a vertical merger can harm downstream rivals either because it lowers the integrated firm's costs (an efficient merger) or because it raises unintegrated costs due to foreclosure (an anti-

[67] Note that t in this case does not represent calendar time, and $t = 0$ is often a different calendar date for each event.

[68] See, e.g., Campbell, Lo, and MacKinley (1997, chapter 4).

[69] Since this technique assesses how the market evaluates a particular vertical merger, it addresses the question of incidence only in the sense that, if abnormal returns are positive, we can conclude that the managers made the right decision when they chose to bring this supplier or retailer within the firm. It does not, however, consider what characteristics of the transaction or firms made integration desirable.

competitive merger).[70] One remedy is to look at share–price effects for buyers of the downstream product (see Mullin and Mullin (1997)). However, in many contexts, this effect can be far removed and is apt to be quite weak.

3.2.3 Computer Simulations and Structural Models

The econometric methods that we have discussed thus far involve estimating reduced–form equations. In particular, there is no way to recover the structural parameters that characterize tastes and technology from such models. This is not a criticism in itself, but it does mean that certain types of analyses cannot be performed. In particular, it is not possible to use reduced-form equations for *ex ante* forecasts of the consequences of changes in policy.[71] There are many circumstances, however, both academic and practical, in which it is desirable to assess the consequences of changes in vertical structures that have not yet occurred. A merger simulation is a tool that could be used for that purpose, and this tool requires a structural model.

The goal of a merger simulation is to predict the equilibrium prices charged and quantities sold under the new, post–merger, market structure using only the information available pre merger. Of course, the advantage of such an approach is that, if the simulation can forecast accurately, performing an *ex ante* evaluation is much less wasteful than waiting for an *ex post* assessment. In particular, the possibility of costly divestitures is lessened by methods that accurately forecast merger effects before the fact.

To illustrate the horizontal–merger technique, consider the case of K firms that produce n branded products with $K \leq n$. The brands are assumed to be substitutes, but the strength of substitutability can vary by brand pair. It is standard to assume that the firms are engaged in a static pricing game. A market structure in that game consists of a partition of the product space into K subsets, where each subset is controlled by one firm or decision maker. Specifically, each firm can choose the prices of the products that are in its subset. A merger then involves combining two or more of the subsets and allowing one player to choose the prices that were formerly chosen by two or more players.

Consider a typical player's choice. When the price of product i increases, the demand for brand j shifts out. If both brands are owned by the same firm, that firm will capture the pricing externality. However, if they are owned by different

[70] On the other hand, foreclosure could benefit integrated and unintegrated upstream rivals.
[71] This is just another example of the Lucas (1976) critique.

firms, the externality will be ignored. After a merger involving substitute products, therefore, prices should increase, or at least not fall. The question that horizontal–merger simulations aim to answer is by how much. Clearly the answer depends on the matrix of cross–price elasticities. Merger simulations have therefore focused on modeling and estimating demand.

Whereas it is becoming increasingly common to supplement traditional horizontal–merger analysis with a merger simulation along the lines just described, this method of evaluation is not yet common for vertical mergers. We are aware only of work by Hendricks and McAfee (1999) and McAfee *et. al.* (2001), who focus on homogeneous, intermediate–goods markets in which both buyers and sellers have market power.[72] In their work, mergers occur between firms (refiners) that are already partially vertically integrated. Their mergers thus have both horizontal and vertical components. Unfortunately, many of the firms that we are concerned with do not fit the assumptions of the McAfee *et. al.* model. In particular, many produce differentiated retail products, not homogenous intermediate goods.

For a number of reasons, we are not optimistic that a 'generic' vertical–merger model that could be used in a wide variety of contexts can be designed. Some aspects of the problem are shared with horizontal–merger analysis and some are unique to vertical mergers. The former include the difficulty of capturing changes in efficiency (cost–lowering effects) and coordinated effects (changes in the ability to collude). Cost savings that arise from a vertical merger are particularly hard to handle in a simulation because they are often based on motivational factors (i.e., better alignment of incentives) rather than arising from technological considerations. In addition, one must consider strategic interactions among horizontal rivals in a vertical context.[73] Equations for rival brands can be included; however a complete model would be very complex. Indeed, it would require assumptions concerning the horizontal games that are played both up and downstream as well as the bargaining games between members of the vertical structures. Moreover, in common with horizontal–merger simulations, if the assumptions that underlie the simulation model are inaccurate, the forecasts will also be inaccurate. We therefore feel that, although this is a fruitful area for future research, routine use of simulation methods to assess vertical mergers is unlikely in the near future. In fact, none of the evidence on the effect of vertical

[72] See also Asker (2004) and Brenkers and Verboven (2006), who adopt a structural approach to assessing the effect of vertical restraints.

[73] It is common to ignore vertical considerations when modeling horizontal mergers, but that does not justify the practice.

mergers that we present below is derived using a structural or simulation approach.

3.3 Evidence on the Consequences of Vertical Mergers

The research reported in tables 1–14 is devoted to an assessment of the incidence of vertical integration. In other words, the variable that is explained in most studies is a measure of whether a transaction takes place (or has a tendency to take place in more aggregate studies) inside a firm or in a market. The research that is reported in tables 15 – 17, in contrast, assesses consequences. In other words, the latter group of studies evaluates the effect of vertical integration on own or rival price, cost, profits, product offerings, survival, or some other economic variable. The information reported for each study is generally the same as that included in the earlier tables. Notice however, that since the dependent variable pertains to one of several outcomes that are not comparable across studies, the penultimate column in tables 15 and 17 indicates the outcome of interest. The final column, in this case, indicates the estimated effect of vertical integration as interpreted by the author. In table 16, we add one more column at the end, to summarize conclusions relative to consumer well being.

Foreclosure and Raising Rival Costs

Competition authorities have focused most attention on foreclosure and raising–rival–cost motives for mergers.[74] It is therefore not surprising that empiricists have also devoted considerable attention to testing whether vertical mergers give rise to foreclosure.

The industries that have been examined tend to be those that have received the most scrutiny from authorities; for example, cement and concrete, cable TV programming and distribution, and oil refining and distribution. In all these cases, the industries, which are natural monopolies or oligopolies, have little in common with the fast–food and other franchise chains that have typically been studied in the empirical moral–hazard literature. In particular, the chances of uncovering anticompetitive behavior is much higher with the former than with the latter.

Table 15 lists articles that test for foreclosure effects. In the table, we do not distinguish between foreclosure and raising rivals' costs. Instead, we include studies that consider imperfectly competitive industries in which some firms are vertically integrated and some are not and where the authors attempt to assess the consequences of that difference. Some of the studies look for tendencies to exclude the

[74] See Rosengren and Meehan (1991) for a list of challenged mergers.

products of unintegrated rivals (e.g., rival programs in the case of cable TV), others assess whether unintegrated rivals pay higher prices for the upstream product (e.g., wholesale prices for gasoline), whereas still others evaluate stock–market reactions to vertical–merger announcements (e.g., changes in returns to holding shares in either rival or downstream consumer firms).

It is clear from the table that some authors have uncovered evidence of foreclosure. However, the existence of foreclosure is, by itself, insufficient to conclude that vertical integration is pernicious. Indeed, recall that Salinger's (1988) model shows that there are two countervailing factors associated with vertical mergers: an increase in foreclosure or other practices that disadvantage rivals and a lessening of double marginalization or other practices that are inefficient. One must therefore balance the two.

Two of the papers in the table attempt to assess that tradeoff (i.e., Mullin and Mullin (1997) and Chipty (2001)), and both conclude that efficiency gains outweigh foreclosure costs. The evidence in favor of anticompetitive foreclosure is therefore, at best weak, particularly when one considers that the industries studied were chosen because their vertical practices have been the subject of antitrust investigations.

Strategic Delegation

Next, we examine evidence concerning the principal's incentive to delegate the pricing decision (vertical separation) in a strategic setting. Recall that under retailer risk neutrality, principals prefer the separated situation. However, the strategic agency model of price competition predicts that vertical integration will gain advantage as risk or risk aversion gains importance and as products become less substitutable. One can test those hypotheses individually but, to our knowledge, this has not been done. Alternatively, a joint test can be constructed from the observation that integration is less apt to occur when rival reaction functions are steep, since the slope of the reaction function determines the strength of rival response to own price increases. This sort of test requires information about each unit's competitors. Slade (1998b), who has such data, finds that, in the context of retail–gasoline sales, integration is indeed less likely when rival-reaction functions are steep. This finding is consistent with the idea that prices should be higher under separation, which means that, although firms might prefer that arrangement, consumers will prefer integration.

Other Consequences

Table 16 summarizes the evidence from several other articles devoted to the conse-

quences of vertical integration. This research is more heterogeneous and more difficult to put in neat pigeonholes. In particular, the consequences of vertical integration for price, cost, investment, profit, profit stability, stock ratings, and, for the work that is based on the capital–asset pricing model, abnormal returns and systematic risk, have all been the subject of investigation. We do not attempt to discuss each article in the table. However, one can get a fairly good idea of the bare bones of each study from the table itself. As mentioned above, this table again includes information on the variable that is assessed (denoted y in the table to indicate that it is the dependent variable), and the effect that integration has on y. For example, if y were cost, and if cost were found to fall with vertical integration, the penultimate column would contain a minus. If in addition, the negative effect were significant at 5%, using a two–tailed test, that column would contain a *. The final column in table 16 shows the effect that integration has on consumer well being (W). When more than one consequence is examined (e.g., profit and profit variability), the effect on well being is determined by the combination of the consequences. For this reason, we indicate the overall effect of vertical integration rather than the effect on each consequence. Not surprisingly, some of the well being results are ambiguous (denoted by ? in the table). For example, if profits increase after integration, we cannot say if consumers are better or worse off, since the change could be due either to higher prices or to lower costs.

To give an idea of the variety of the work, we discuss two somewhat arbitrarily chosen studies. The first, which is by Anderson (1988) examines the effect of vertical integration on opportunism. The setting is one of industrial sales forces, which can be either direct (vertically integrated) or manufacturers' reps (vertically separated). In order to get a measure of opportunism, Anderson asked sales managers to answer questions that reflected the behavior and attitudes of their sales forces. For example, one question asked managers to rank the validity of the statement that sales people distort information to the company in order to protect their own interests. The 'opportunism' variable, which was constructed as an index of the answers to several such questions, was then regressed against a vertical–integration dummy, as well as variables that capture asset specificity, environmental uncertainty, and other relevant factors. Anderson found that integration significantly reduced opportunistic behavior. In addition, she found that opportunism was positively related to specificity and uncertainty.

The second study, which is by Kerkvliet (1991), involves mine–mouth electric–

generating plants. In other words, the plants studied are located in close proximity to coal mines, which is a classic example of site specificity (see e.g. Joskow, 1985). Kerkvliet estimated a neoclassical cost function. However, instead of using market prices of inputs, as would be common in competitive environments, he allowed input shadow prices to differ systematically from market prices. The factors that he considered might cause distortions, or wedges between the two sets of prices, are regulatory variables, monopsony power, and vertical arrangement (integration or separation). He found that integration led to increased allocative and technical efficiency. Furthermore, although site specificity endowed all plants with monopsony power, the tendency to exercise that power was significantly reduced by vertical integration.

Like the studies just discussed, the body of research that is reported in table 16 is highly supportive of the efficiency of vertical integration and mergers. In particular, there are no minus signs in the final column of the table, which indicates that integration benefits consumers, or at least does not harm them. In addition, almost all of the positive findings are statistically significant. Finally, even in this table we find that many of the horizontal markets examined (e.g., ready–mix concrete) are highly concentrated. Since these are exactly the type of markets where one might expect to find negative welfare effects from vertical mergers, it is particularly informative that the set of results in this table shows no such negative effects.

Divorcement

The mergers and divestitures whose effects we have considered so far were voluntarily undertaken by the parties to the transaction.[75] Not all changes in vertical structures, however, come from within the upstream/downstream relationship. Indeed, it is not uncommon for government agencies to mandate structural changes, usually divestitures. This is most apt to occur when the agency believes that the vertical structure is exacerbating horizontal market power.

For example, gasoline stations can be owned and operated by the oil company (CC contracts), owned by the company but operated by the dealer (CD contracts), or owned and operated by the dealer (DD contracts). In other words, transactions can occur within a vertically integrated firm (CC), in an arm's length market (DD), or under an intermediate arrangement (CD), and there are many efficiency considerations that motivate the choice among those possibilities. Nevertheless, in a number of instances, competition authorities or regional governing bodies have alleged that anti-

[75] We do not mean that the table excludes hostile takeovers. Instead, it excludes mergers or divestitures that have been mandated by public authorities.

competitive motives outweigh efficiency considerations when firms make that choice. In particular, a number of US states have outlawed CC contracts on the grounds that integrated oil companies would attempt to disadvantage unintegrated downstream competitors.

Table 17 contains details of five studies that relate to the issue of divorcement. The three studies that assess gasoline divorcement directly (Barron and Umbeck (1984), Vita (2000) and Blass and Carlton (2001)) conclude that retail prices and costs were higher and hours were shorter after it occurred. In other words, they are unanimous in concluding that the policy was misguided. The fourth study of the gasoline market, Hastings (2004), looks at a slightly different issue. She finds that, although retail prices are higher at vertically integrated stations than at unintegrated independents, there is no difference between prices at CC and CD stations. Given that the rationale behind divorcement was that CC arrangements gave oil companies incentives to charge higher wholesale prices to CD stations, her finding is unsupportive of that motive.

Finally, the contracts that are written between brewers and publicans in the UK beer market are almost identical to those between oil companies and service stations in the US. Moreover, those contracts have also been the subject of investigations that eventually led to divorcement. However, in that market, divorcement involved changes in ownership not mode of operation. In other words CD contracts were changed to DD. Slade (1998a) finds that that change also led to higher retail prices, probably as a result of double marginalization.

The logic that led to divorcement regulations thus seems to have been flawed. In particular, the forced move from CC to CD contracts for gasoline appears to have ignored the fact that integrated oil companies owned the affected stations and chose whether to operate them under CC or CD arrangements. Having made a profit–maximizing decision to operate some of their owned stations internally and allow dealers to operate the others under rental contracts (presumably based on efficiency considerations), it would be perverse for those companies to turn around and attempt to disadvantage their affiliated CD retailers and drive them out of the market. After all, the oil company could have chosen closure or self operation for those outlets in the first place.

The thinking that led to the move from CD to DD contracts in the beer market, in contrast, appears to have ignored the fact that divestiture is associated with countervailing factors — the introduction of double marginalization and the elimination

of foreclosure — and that the former costs can outweigh the latter benefits.

Summary

The literature on the consequences of vertical integration appears to be much more fragmented than that on incidence. This is perhaps due to the fact that it is less well integrated with the theory. In particular, there is a need for a simple model that can encompass the various predictions that have been tested in the empirical literature. Instead, we have a set of theoretical models, each one concentrating on a single aspect of the problem. This is clearly an area where future research could help us understand and better interpret the body of evidence.

In spite of the lack of unified theory, overall a fairly clear empirical picture emerges. The data appear to be telling us that efficiency considerations overwhelm anticompetitive motives in most contexts. Furthermore, even when we limit attention to natural monopolies or tight oligopolies, the evidence of anticompetitive harm is not strong.

4 Conclusions

In our attempt to organize what is now a very large empirical literature on the vertical boundaries of the firm, we have covered a lot of ground. Of course, to keep the paper manageable, we have also made a number of choices. First, we have decided not to cover the neoclassical approach to integration in which a firm is seen as a production function — a set of feasible input/output relationships. With that class, integration is motivated by technological considerations of economies of scale and scope, including vertical economies. Although the empirical literature in that area is vast, we chose to not cover it as we do not feel that technological factors are especially complementary to the incentive–based motives that have been our focus.

Second, we have made no attempt to review the large body of work that deals with vertical integration in the management literature. Much of that work relies on TCE arguments, but significant portions also rely on alternative views, focusing on ideas like organizational capabilities or resources, where firms hire people or develop processes that make them good at certain things but not good at others. Considerations in that vein imply that firms will integrate into areas that are consistent with their capabilities and will shy away from areas that are not. We view those arguments as similar to the traditional arguments described above and, in that sense, outside the scope of this survey.

Third, we have not discussed models that are too new to have produced a body of

empirical evidence. In particular, we have not touched upon dynamic models such as those that are based on the notion of relational contracts (e.g., Baker, Gibbons, and Murphy (2002)) or the earlier notion of self–enforcing contracts (Klein and Leffler (1981), Bull (1987), Klein and Murphy (1997)).[76] With those dynamic models, interaction, whether it be in a firm or a market, is modeled as a repeated game with spot transactions (again in a firm or market) as the punishment for reneging. This is clearly a fruitful area for future empirical research that has remained relatively unexplored.[77]

Other important areas that are ripe for empirical assessment include the relationship between vertical integration and development, particularly the development of the institutions that make contracting a feasible alternative (see, e.g., McMillan and Woodruff 1999 and Acemoglu, Johnson, and Mitton 2007), and the relationship between vertical integration and trade (see, e.g., Antras 2003).

Third, we have tried to provide simple versions of the theoretical models that underlie the material that we cover in order to derive predictions around which we could organize the findings. Of course, by definition, those simple models neglect many important issues and extensions. For example, our moral–hazard model does not consider the rich set of tools that can be used to provide incentives inside firms, the transaction–cost model glosses over the hold–up problems that can occur within firms, and the property–rights model does not explain why firms, rather than individuals, own assets.[78] A further limitation is that the theories, at least as we have presented them, are more applicable to the entrepreneurial firm. In particular, we have modeled the manufacturer as both decision maker and asset owner. This means that we have not considered the important conflicts that can occur between managers and shareholders in the modern corporation — problems that result from divorce of ownership and control.[79] It is therefore perhaps surprising that, as we have seen, the models' predictive powers are so high.

Finally, we have partitioned the theories that we discuss into distinct groups,

[76] Other models, such as the PR model of Grossman and Hart (1986) and Wernerfelt's (1997) adjustment–cost model are dynamic in that they involve sequential actions. However, they do not involve the threat of punishment to sustain cooperation.

[77] Most of the empirical literature in this area takes the form of case studies, as in the many articles on the GM-Fisher Body case, Kenney and Klein (1983) and the related literature on the Paramount case, which brought studio ownership of theaters under fire, and Kaufmann and Lafontaine (1994), on McDonald's, for example. But see Gil and Hartmann (2007) for an example of more quantitative analysis of the effect of reputation on vertical integration decisions.

[78] This last point is made by Holmstrom (1999) who provides a good discussion of the relationship between MH and PR models.

[79] The issue of divorce of ownership and control was emphasized by Berle and Means (1933).

which we have called moral–hazard, transaction–cost, property–rights, and market–power arguments, and we have used that partitioning to organize the evidence. The partition that we chose, however, is somewhat arbitrary, and it is often difficult to fit empirical studies into neat non–overlapping classes. As a result, there are a number of instances where we have included a single study in more than one pigeonhole, and, in other instances, we have not included studies in pigeonholes that the authors might find appropriate. We use the study by Baker and Hubbard (2003) to illustrate the difficulties involved in categorizing. In their setting — for–hire trucking — relationship specific assets are not particularly important, whereas incentives and job design are. In that sense, therefore, their study fits into the moral–hazard, particularly the multitasking paradigm. However, the issue at stake is residual decision rights —the ability to determine asset use in contingencies not specified in contracts — and not residual claimancy. In that sense, therefore, their study fits into the property–rights paradigm.

The problem with any attempt to categorize the evidence is that the world is more complex than the simple models might lead one to believe. The advantage, however, is that the possibilities for cross fertilization are abundant. To illustrate, although there are few direct tests of the property–rights model, we have been able to gain insights into its predictive power by considering evidence that comes from the moral–hazard and transaction–cost literature. Moreover, we have found that, at least when TC and PR model predictions are at variance with one another, there is little support for the PR theory in evidence that can be gleaned from make–or–buy decisions. However, there is much more support for the PR model in evidence that can be obtained from manufacturer/retailer and franchisor/franchisee integration decisions. The relationship among predictions and evidence from these two sets of models, and the opportunities for further cross fertilization, clearly deserves further thought.

We have emphasized – and perhaps overemphasized – some of the differences between TC and PR models. We did this to underline the fact that one is not merely a formalization of the other, as is sometimes claimed. Nevertheless, the similarities are strong. Indeed, both are based on the idea that relationship–specific assets and noncontractability lead to *ex post* problems when quasi rents must be allocated.

The links between MH and PR models have received much less attention in the literature. Yet the fact that their predictions are similar should not be surprising. Indeed, although, the first looks at the division of revenues and the second at the

division of control, if these two important aspects of the firm are divorced from one another, new agency costs will arise. Similarities also stem from the fact that both are concerned with mechanisms that can alleviate the problems associated with double–sided moral–hazard, or more generally, double–sided incentive issues.

We have discussed some of the econometric problems that are involved in identifying the effects of interest. We have also stressed that many of the tests that are summarized in the tables are incomplete in the sense that they look at one factor that is predicted to affect vertical integration, holding the other factors constant, whereas it is often a combination of factors that ought to be assessed. For example, asset specificity does not create problems on its own but only in conjunction with noncontractibility. Yet only a few studies have considered such interactions. Moreover, as Novak and Stern (2003) note, vertical integration decisions for parts of a system — in their application systems within cars — are apt to be related to one another rather than taken separately. To our knowledge, theirs is the only study to have considered such interactions. Empirical work that accounts for complementarities among factors and decisions is an under exploited area that is ripe for future research.

We have described measurement difficulties throughout, starting with the difficulty of measuring vertical integration itself as distinct from various types of contractual arrangements. A related problem, that we have thus far ignored, arises from the difference between marginal and average. The empirical studies are mostly concerned with averages (e.g., the importance of the agent's effort). Moreover, the bare–bones models that we have presented make no distinction between the two. However, in so far as the two differ in real–world contexts, the data often measure the wrong one.

One must bear these and other caveats in mind when evaluating the evidence. Nevertheless, the empirical regularities are both consistent and strong. In other words, when we compare the evidence concerning the effect of a particular factor on vertical–integration taken from studies of different industries, time periods, and geographic regions, we find that the sign of the effect is almost always consistent across studies, at least in cases where it is significant. For example, upstream asset specificity encourages vertical integration, no matter what its source. The evidence is therefore stronger than one might expect given the difficulties involved. In fact, the degree of consistency suggests that perhaps some form of publication bias exists. Specifically, it might be easier to publish papers that confirm theories. Yet it is our view that much can be learned from studies that find evidence contradicting certain assumptions or predictions from theory.

Of course, we showed that not all of the evidence is consistent with the theory that motivates a test. A striking example of this is the negative relationship between downstream risk and vertical integration, which is inconsistent with the tradeoff between incentive and insurance concerns that is fundamental to the basic moral–hazard model. We have noted that the regularity could result from endogenous risk (i.e., agents with higher powered incentives respond more strongly to shocks), from self selection (i.e., agents with lower risk aversion choose riskier activities), or from increased agent importance (i.e., when conditions are more volatile, agent investments are more valuable, as suggested by PR theories). Nevertheless, the strength of the finding is puzzling, and we are encouraged by the fact that theorists have responded to this contrary evidence and provided new models to explain it.

As to what the data reveal in relation to public policy, we did not have a particular conclusion in mind when we began to collect the evidence, and we have tried to be fair in presenting the empirical regularities. We are therefore somewhat surprised at what the weight of the evidence is telling us. It says that, under most circumstances, profit–maximizing vertical–integration and merger decisions are efficient, not just from the firms' but also from the consumers' points of view. Although there are isolated studies that contradict this claim, the vast majority support it. Moreover, even in industries that are highly concentrated so that horizontal considerations assume substantial importance, the net effect of vertical integration appears to be positive in many instances. We therefore conclude that, faced with a vertical arrangement, the burden of evidence should be placed on competition authorities to demonstrate that that arrangement is harmful before the practice is attacked. Furthermore, we have found clear evidence that restrictions on vertical integration that are imposed, often by by local authorities, on owners of retail networks are usually detrimental to consumers. Given the weight of the evidence, it behooves government agencies to reconsider the validity of such restrictions.

References Cited

Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti (2005) "Vertical Integration and Technology: Theory and Evidence," mimeo, MIT.

Acemoglu, Daron, Simon Johnson, and Todd Mitton (2007) "Determinants of Vertical Integration: Financial Development and Contracting Costs," mimeo, MIT.

Ackerberg, Daniel A. and Maristella Botticini (2002) "Endogenous Matching and the Empirical Determinants of Contractual Form," *Journal of Political Economy,* 110: 564–591.

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

Allen, B. (1971) "Vertical Integration and Market Foreclosure: The Case of Cement and Concrete," *Journal of Law and Economics*, 14: 251–274.

Allen, Douglas W. and Dean Lueck (1995) "Risk Preferences and the Economics of Contracts." *American Economic Review*, 85: 447–451.

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

Anderson, Erin (1988) "Transaction Costs as Determinants of Opportunism in Integrated and Independent Sales Forces," *Journal of Economic Behavior and Organization*, 9: 247–264.

Anderson, Erin, and David C. Schmittlein (1984) "Integration of the Sales Force: An Empirical Examination," *Rand Journal of Economics*, 15: 385–395.

Antras, Pol (2003) "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics,* 118: 1375–1418.

Arruñada, Benito, González-Daz, M. and A. Fernández (2004) "Determinants of organizational form: transaction costs and institutions in the European trucking industry," *Industrial and Corporate Change*, 13: 867–882.

Asker, John (2004) "Measuring Cost Advantages from Exclusive Dealing An Empirical Study of Beer Distribution," Harvard University mimeo.

Azoulay, Pierre (2004) "Capturing Knowledge Within and Across Firm Boundaries: Evidence from Clinical Development," *American Economic Review*, 94: 1591–1612.

Bai, C. and Tao, Z. (1999) "Contract Mixing in Franchising as a Mechanism for Public Good Provision," *Journal of Economics and Management Strategy*, 9: 85–113.

Baker, George P. , Gibbons, Robert and Murphy, Kevin J. (1999) "Informal authority in organizations," *The Journal of Law, Economics, & Organizations,* 15: 56–73.

Baker, George P. , Gibbons, Robert and Murphy, Kevin J. (2002) "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics*, 117: 39–84.

Baker, George P., and Hubbard, Thomas N. (2003) "Make Versus Buy in Trucking: Asset Ownership, Job Design, and Information," *American Economic Review*, 93: 551–572.

Baker, George P., and Hubbard, T.N. (2004) "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking," *Quarterly Journal of Economics*, 119: 1443–1479.

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

Bhattacharyya, Sugato and Lafontaine, Francine (1995) "Double-Sided Moral Hazard and the Nature of Share Contracts," *Rand Journal of Economics*, 26: 761-81.

Berle, A. and Means, G. (1933) *The Modern Corporation and Private Capital*, New York: Macmillan.

Blass, A.A. and Carlton, Dennis W. (2001) "The Choice of Organizational Form in Gasoline Retailing and the Cost of Laws that Limit that Choice," *Journal of Law and Economics*, 44: 511–524.

Bork, Robert H. (1969) "Vertical Integration and Competitive Processes," *Public Policy Towards Mergers*, F. Weston and S. Peltzman (eds.), Pacific Palisades: Goodyear Publishing Co.

Brenkers, R. and Verboven, F. (2006) "Liberalizing a Distribution System: The European Car Market," *Journal of the European Economic Association*, 4: 216–251.

Brickley, James A. (1999) "Incentive Conflicts and Contractual Restraints: Evidence from Franchising," *Journal of Law and Economics*, 42: 745–774.

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

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

Brickley, J.A., Linck, J.L., and Smith, C.W. (2003) "Boundaries of the Firm: Evidence from the Banking Industry," *Journal of Financial Economics*, 70: 351–383.

Bull, C. (1987) "The Existence of Self–Enforcing Implicit Contracts," *Quarterly Journal of Economics*, 102: 147–159.

Campbell, J.Y., Lo, A., and MacKinley, C. (1997) *The Econometrics of Financial Markets*, Princeton University Press.

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

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

Chipty, Tasneem (2001) " Vertical Integration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry," *American Economic Review*, 91: 428–453.

Ciliberto, Federico (2006) "Does Organizational Form Affect Investment Decisions?" *Journal of Industrial Economics*, 54: 63–93.

Coase, Ronald (1937) "The Nature of the Firm," *Economica*, 4: 386–405.

Cooper, J., Froeb, Lueck M., O'Brien, Daniel, and Vita, M. G. (2004) "Inferring the Effects of Vertical Integration and Restraints," *International Journal of Industrial Organization*, forthcoming.

Corts, Kenneth S. (2001) "The Strategic Effects of Vertical Market Structure: Common Agency and Divisionalization in the US Motion Picture Industry," *Journal of Economics and Management Strategy*, 10: 509–528.

Eckbo, B.E. (1983) "Horizontal Mergers, Collusion, and Stockholder Wealth," *Journal of Financial Economics*, 11: 241–273.

Edwards, K., Jackson, J.D., and Thompson, H.L. (2000) "A Note on Vertical Integration and Stock Ratings of Oil Companies in the US," *The Energy Journal*, 21: 145–151.

Forbes, Silke J. and Mara Lederman (2006) "Control Rights, Network Structure and Vertical Integration: Evidence from Regional Airlines," mimeo, University of Toronto.

Forbes, Silke J and Mara Lederman (2007) "The Performance Implications of Vertical Integration: Evidence from Regional Airlines," mimeo, University of Toronto.

Ford, G.S. and Jackson, J.D. (1997) "Horizontal Concentration and Vertical Integration in the Cable TV Industry,"*Review of Industrial Organization*, 12: 501–518.

Gibbons, Robert (2005) "Four Formal(izable) Theories of the Firm?" *Journal of Economic Behavior and Organization*, 58: 200–245.

Gil, Ricard (2007a) "Make-or-Buy in Movies: Integration and Ex-post Renegotiation" *International Journal of Industrial Organization*, forthcoming.

Gil, Ricard (2007b) "Revenue Sharing Distortions and Vertical Integration in the Movie Industry," mimeo, HBS and UC-Santa Cruz.

Gil, Ricard and Wesley R. Hartmann (2007) "Airing Your Dirty Laundry: Vertical Integration, Reputational Capital and Social Networks," mimeo, Stanford Graduate School of Business.

González-Diaz, M., Benito Arruñada, and A. Fernández. (2000) . "Causes of Subcontracting: Evidence from Panel Data on Construction Firms," *Journal of Economic Behavior and Organization*, 42: 167–187.

Gould, J.R. (1977) "Price Discrimination and Vertical Control," *Journal of Political Economy*, 85: 1063-1071.

Greenhut, M.L. and Ohta, H. (1979) "Vertical integration of Successive Oligopolists," *American Economic Review*, 69: 137–141.

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

Grout, P. (1984) "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," *Econometrica*, 52: 449–460.

Hanson, Gordon H. (1995) "Incomplete Contracts, Risk, and Ownership," *International Economic Review*, 36: 341–363.

Hart, Oliver (1995) *Firms, Contracts, and Financial Structure*, Oxford: Oxford University Press.

Hart, Oliver and Moore, J. (1990) "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98: 1119-1158.

Hart, Oliver and Tirole, Jean (1990) "Vertical Integration and Market Foreclosure," *Brookings Papers in Economic Activity: Microeconomics*, 205–276.

Hastings, Justine (2004) "Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California," *American Economic Review*, 94: 317–328.

Hastings, Justine and Gilbert, Richard (2005) "Market Power, Vertical Integration, and the Wholesale Price of Gasoline," *Journal of Industrial Economics*, 469–492.

Helfat, Constance E. and Teece, David J. (1987) "Vertical Integration and Risk Reduction," *Journal of Law, Economics, and Organization*, 3: 47–67.

Hendricks, Kenneth and McAfee, R.Preston (1999) "Measuring Industry Concentration in Intermediate Goods," University of Texas mimeo.

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

Holmstrom, Bengt (1999) "The Firm as a Subeconomy," *Journal of Law, Economics, and Organization*, 15: 74–102.

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

Holmstrom, Bengt and Milgrom, Paul (1994) "The Firm as an Incentive System," *American Economic Review*, 84: 972–991.

Holmstrom, Bengt and Roberts, J. (1998) "The Boundaries of the Firm Revisited," *Journal of Economic Perspectives*, 12: 73–94.

Hortaçsu, Ali and Syverson, Chad (2007a) "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices," *Journal of Political Economy,* forthcoming.

Hortaçsu, Ali and Syverson, Chad (2007b) "Vertical Integration and Production: Some Plant-Level Evidence," University of Chicago mimeo.

Jin, Ginger and Philip Leslie (2005) "Reputation Incentives for Restaurant Hygiene," mimeo, University of Maryland.

John, George, and Weitz, Barton A. (1988) "Forward Integration into Distribution: An Empirical Test of Transaction Cost Analysis," *Journal of Law, Economics, and Organization*, 4: 337–356.

Johnson, Ronald N. and Allen M. Parkman (1987) "Spatial Competition and Vertical Integration: Cement and Concrete Revisited: Comment," *American Economic Review*, 77: 750–53.

Joskow, Paul L. (1985) "Vertical Integration and Long–Term Contracts: The Case of Coal–Burning Electric Generation Plants," *Journal of Law, Economics, and Organization*, 1: 33–80.

Joskow, Paul L. (2005) "Vertical Integration" in Ménard, C. and M. Shirley, eds. *Handbook of New Institutional Economics*, Dordrecht: Springer.

Kalnins, Arturs T. and Francine Lafontaine. (2004) "Multi-Unit Ownership in Franchising: Evidence from the Fast-Food Industry in Texas." *RAND Journal of Economics*, 35: 747-761.

Kaufmann, Patrick J. and Francine Lafontaine (1994) "Costs of Control: The Source of Economic Rents for McDonald's Franchisees, *Journal of Law and Economics*, October 1994, 417–453.

Kehoe, M.R. (1996) "Franchising, Agency Problems, and the Cost of Capital," *Applied Economics*, 28: 1485–1493.

Kenney, Roy and Klein, Benjamin (1983) "The Economics of Block Booking," *Journal of Law and Economics*, 26: 497–540.

Kerkvliet, J. (1991) "Efficiency and Vertical Integration: The Case of Mine–Mouth Electric Generating Plants," *Journal of Industrial Economics*, 34: 467–482.

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

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

Klein, B. and Leffler, K. (1981) "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 89: 615–641.

Klein, Benjamin and Kevin M. Murphy (1997) "Vertical Integration as a Self-Enforcing Contractual Arrangement" *American Economic Review* 87: 415–420.

Klein, Peter G. (2005) "The Make-or-Buy Decision: Lessons from Empirical Studies" in Ménard, C. and M. Shirley, eds. *Handbook of New Institutional Economics*, Dordrecht: Springer.

Knight, Frank (1921) *Risk, Uncertainty, and Profit*, Chicago: Houghton Mifflin.

Krattenmaker, T. and Salop, Steven C. (1986) "Anti-competitive Exclusion: Raising Rival's Costs to Achieve Power over Price," *Yale Law Journal*, 96: 209–293.

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

Lafontaine, Francine and Bhattacharyya, Sugato (1995) "The Role of Risk in Franchising," *Journal of Corporate Finance*, 2: 39–74.

Lafontaine, Francine and Emmanuel Raynaud (2002) "Residual claims and self enforcement as incentive mechanisms in franchise contracts: substitute or complements," in Brousseau, E. and J.M. Glachant Ed., *The Economics of Contract in Prospect and Retrospect* Cambridge University Press.

Lafontaine, F. and Shaw, Kathryn (1999) "The Dynamics of Franchise Contracting: Evidence from Panel Data," *Journal of Political Economy*, 107: 1041–1080.

Lafontaine, F. and Shaw, Kathryn (2005) "Targeting Managerial Control: Evidence from Franchising," *RAND Journal of Economics*, 36: 131–150.

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

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

Lafontaine, F. and Margaret E. Slade (2001) "Incentive Contracting and the Franchising Decision," *Game Theory and Business Applications*, K. Chatterjee and W. Samuelson (eds.) Kluwer Academic Press, 133-188.

Lafontaine, F. and Margaret E. Slade(2006) "Exclusive Contracts and Vertical Restraints: Empirical Evidence and Public Policy," *Handbook of Antitrust Economics*, P. Buccirossi (ed.) Cambridge: MIT Press, forthcoming.

Lafontaine, Francine and Margaret E. Slade (2008) "Inter-Firm Contracts: The Evidence." In *Handbook of Organizational Economics*, R. Gibbons and J. Roberts (eds.), forthcoming.

Levin, R.C. (1981) "Vertical Integration and Profitability in the Oil Industry," *Journal of Economic Behavior and Organization*, 2: 215–235.

Liebeler, W.J. (1968) "Towards a Consumer's Antitrust Law: The Federal Trade Commission and Vertical Mergers in the Cement Industry," *UCLA Law Review*, 15: 1153–1202.

Lieberman, Marvin B. (1991) "Determinants of Vertical Integration: An Empirical Test," *The Journal of Industrial Economics*, 34: 451–466.

Lucas, Robert (1976) "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy*, 1: 1946.

Lyons, Bruce (1995) "Specific Investment, Economies of Scale, and the Make–or–Buy Decision," *Journal of Economic Behavior and Organization*, 26: 431–443.

MacKinlay, A. G. (1997) "Event Studies in Economics and Finance," *Journal of Economic Literature*, 35: 13–39.

Maness, Robert (1996) "Incomplete Contracts and the Choice between Vertical Integration and Franchising," *Journal of Economic Behavior and Organization*, 31: 101–115.

Martin, Robert E. (1988) "Franchising and Risk Management" *American Economic Review* 78: 954–968.

Masten, Scott E. (1984) "The Organization of Production: Evidence from the Aerospace Industry," *Journal of Law and Economics*, 27: 403–417.

Masten, Scott E. (1988) "A Legal Basis for the Firm," *Journal of Law, Economics and Organization*, 4: 181

Masten, Scott E., Meehan, J.W., and Snyder, Edward A. (1989) "Vertical Integration in the US Auto Industry: A Note on the Influence of Transaction Specific Assets," *Journal of Economic Behavior and Organization*, 12: 265–273.

Masten, Scott E., Meehan, J.W., and Snyder, Edward A. (1991) "The Costs of Organization," *Journal of Law, Economics and Organization*, 7: 1–25.

McAfee, R.Preston, Hendricks, Kenneth, Fried, J.M., and Williams, M.A. (2001) "Measuring Anticompetitive Effects of Mergers when Buyer Power is Concentrated," *Texas Law Review*, 79: 1621–1632.

McBride, M.E. (1983) "Spatial Competition and Vertical Integration: Cement and Concrete Revisited," *American Economic Review*, 73:1011–1022.

McMillan, John and Christopher Woodruff (1999) "Dispute Prevention without courts in Vietnam," *Journal of Law, Economics, and Organization,* 15: 637–658.

Minkler, A (1990) "An Empirical Analysis of a Firm's Decisions to Franchise," *Economics Letters*, 34: 77–82.

Minkler, A. and Park, T.A. (1994) "Asset Specifity and Vertical Integration in Franchising," *Review of Industrial Organization*, 9: 409–423.

Monteverde, K. and Teece, David J. (1982) "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *Bell Journal of Economics*, 13: 206–213.

Mullainathan Sendhil and David Scharfstein (2001) "Do Firm Boundaries Matter?" *The American Economic Review Papers and Proceedings*, 91: 195–199.

Mullin, J.C. and Mullin, Wallace P. (1997) "United States Steel's Acquisition of the Great Northern Properties: Vertical Foreclosure or Efficient Contractual Governance?" *Journal of Law, Economics, and Organization*, 13: 74–100.

Muris, Timothy J., Scheffman, D.T., and Spiller, Pablo T. (1992) "Strategy and Transaction Costs: The Organization of Distribution in the Carbonated Soft Drink Industry," *Journal of Economics and Management Strategy*, 1: 83–128.

Newey, W.K. and Kenneth D. West (1987) "A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55: 703–708.

Nickerson, Jackson A. and Brian S. Silverman (2003) "Why Arent All Truck Drivers Owner-Operators? Asset Ownership and the Employment Relation in Interstate For-Hire Trucking," *Journal of Economics and Management Strategy* 12: 91–118.

Norton, Seth W. (1988) "An Empirical Look at Franchising as an Organizational Form," *Journal of Business*, 61: 197–217.

Novak, Sharon and Scott Stern (2003) "Complementarity Among Vertical Integration Decisions: Evidence from Automobile Product Development," mimeo, Northwestern University.

Ohanian, Nancy K., (1994) "Vertical Integration in the US Pulp and Paper Industry: 1900–1940, *Review of Economics and Statistics*, 76: 202–207.

Ordover, J.A., Saloner, Garth, and Salop, Steven (1990) "Equilibrium Vertical Foreclosure," *American Economic Review*, 127–142.

Peltzman, Samuel (9169) "Issues in Vertical Integration Policy," *Public Policy Towards Mergers*, Weston, F. and Peltzman, S. (eds.), Pacific Palisades: Goodyear Publishing Co.

Pénard Thierry, Emmanuel Raynaud and Stphane Saussier (2003) "Dual Distribution and Royalty Rates: An Empirical Analysis using French Data," *Journal of Marketing Channels* 10: 5–31.

Perry, Martin K. (1978) "Vertical Integration: The Monopsony Case," *American Economic Review*, 68: 561–570.

Pinkse, J., Margaret E. Slade, and Brett, C. (2002) "Spatial Price Competition: A Semiparametric Approach," *Econometrica*, 70: 1111–1155.

Pinkse, J., Margaret E. Slade and Shen, L. (2005) "Dynamic Spatial Discrete Choice Using One Step GMM: An Application to Mine Operating Decisions," *Spatial Economic Analysis*, 1: 53–99.

Pirrong, S. Craig (1993) "Contracting Practices in Bulk Shipping Markets: A Transactions Cost Explanation," *Journal of Law and Economics* 36: 937–76.

Pitofsky, R. (1997) "Vertical Restraints and Vertical Aspects of Mergers — A US Perspective," speech delivered to the Fordham Corporate Law Institute.

Prendergast, Canice (2002) "The Tenuous Trade-Off Between Risk and Incentives." *Journal of Political Economy*, 110: 1071–1102.

Reiffen, D. and Kleit, A. (1990) "Terminal Railroad Revisited: Foreclosure of an Essential Facility or Simple Horizontal Monopoly?" *Journal of Law and Economics*, 38: 419–438.

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

Rosengren, E.S. and Meehan, J.W. (1994) "Empirical Evidence on Vertical Foreclosure," *Economic Inquiry*, 32: 303–317.

Ross, S.A. (1967) "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13: 41–360.

Salinger, Michael A. (1988) "Vertical Mergers and Market Foreclosure," *Quarterly Journal of Economics*, 103: 345–356.

Salop, Steven C. and Scheffman, David T. (1983) "Cost–Raising Strategies," *Journal of Industrial Economics*, 36: 19–34.

Schmalensee, Richard (1973) "A Note on the Theory of Vertical Integration," *Journal of Political Economy*, 81: 442–449.

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

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

Sharpe, W.F. (1963) "A Simplified Model for Portfolio Analysis," *Management Science*, 9: 277–293.

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

Shepard, Andrea (1993) "Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing," *RAND Journal of Economics*, 24: 58–77.

Slade, Margaret E. (1996) "Multitask Agency and Organizational Form: An Empirical Exploration," *International Economic Review*, 37: 465-486.

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

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

Snyder, Christopher M. (1996) "Vertical Integration for Efficiency or Market Power? Event Studies ofthe US Oil Industry," George Washington University mimeo.

Spengler, J.J. (1950) "Vertical Integration and Antitrust Policy," *Journal of Political Economy*, 58: 347-352.

Spiller, Pablo T. (1985) "On Vertical Mergers," *Journal of Law, Economics, and Organization*, 1: 285-312.

Tadelis, Steven (2002) "Complexity, Flexibility, and the Make-or-Buy Decision," *American Economic Review,* 92: 433-437.

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

Vita, Michael G. (2000) "Regulatory Restrictions on Vertical Integration and Control: The Competitive Impact of Gasoline Divorcement Policies," *Journal of Regulatory Economics*, 18: 217–233.

Walker, G. and Weber, D. (1984) "A Transaction Cost Approach to Make–or–Buy Decisions," *Administrative Science Quarterly*, 29: 373–391.

Warren-Boulton, Frederick R. (1974) "Vertical Control With Variable Proportions," *Journal of Political Economy*, 82: 783-802.

Warren–Boulton, Frederick R. (2003) "The Contribution of the Merger Guidelines to the Analysis of Non–Horizontal Mergers," MiCRA mimeo.

Waterman, D. and Weiss, A.A. (1996) "The Effects of Vertical Integration Between Cable Television Systems and Pay Cable Networks," *Journal of Econometrics*, 72: 357–395.

Weiss, A. (1992) "The Role of Firm–Specific Capital in Vertical Mergers," *Journal of Law and Economics,*, 35: 71–88.

Wernerfelt, Birger (1997) "On the Nature and Scope of the Firm: An Adjustment-Cost Theory," *Journal of Business,* 70: 489–514.

Whinston, Michael D. (2003) "On the Transaction Cost Determinants of Vertical Integration," *Journal of Law, Economics, and Organization*, 19: 1–23.

Williamson, Oliver E. (1971) "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 61: 112–23.

Williamson, Oliver E. (1975) *Markets and Hierarchies*, New York: Free Press.

Williamson, Oliver E. (1979) " Transaction Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 22: 233-261.

Williamson, Oliver E. (1983) "Credible Commitments: Using Hostages to Support Exchange," *American Economic Review*, 73: 519–540.

Williamson, Oliver E. (1985) *The Economic Institutions of Capitalism*, New York: Free Press.

Williamson, Oliver E. (1991) "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives," *Administrative Science Quarterly*, 36: 269–96.

Woodruff, C. (2002) "Non–Contractible Investment and Vertical Integration in the Mexican Footwear Industry," *International Journal of Industrial Organization*, 20: 1197–1224.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Yeap, Clarissa (2006) "Residual Claims and Incentives in Restaurant Chains," University of Minnesota and Census Bureau Center for Economic Studies Working paper, CES 06-18.

Table 1: The Effect of Risk on Forward Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Anderson & Schmittlein	1984	Electronic components & sales	Cross section Logit	% Forecast error, sales	-
John & Weitz	1988	Industrial Goods & sales	Cross section OLS, Logit	Environmental uncertainty index	+
Martin	1988	Retail & services	Panel Weighted Least Sq.	Dispersion in detrended sales	-*
Norton	1988	Restaurants & motels	Cross Section OLS, 2SLS	Dispersion in detrended sales - Restaurants - Motels - Refreshment Places	 + - -*
Lafontaine	1992	Retail & services	Cross section Tobit	Proportion of outlets discontinued	-*
Lafontaine & Bhattacharyya	1995	Retail & services	Descriptive	Sales dispersion Rate of Outlet discontinuation	-*
Woodruff	2002	Footwear & sales	Cross section regressions	Frequent fashion change	-*

* denotes significance at 5% using a two–tailed test.

Table 2: The Effect of the Importance of Agent Effort on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Caves & Murphy	1976	Retail & services	Cross section Regressions	Personalized service dummy	-*
Norton	1988	Restaurants & motels	Cross Section OLS, 2SLS	Employee to sales ratio - Restaurants - Motels - Refreshment Places	 -* + -*
Lafontaine	1992	Retail & services	Cross section Tobit	Sales minus franchisor inputs Franchisee experience required	- +
Shepard	1993	Gasoline refining & sales	Cross section Regressions	Full service dummy	-
Scott	1995	Retail & services	Cross Section Regressions	Capital to labor ratio	(+*)
Maness	1996	Various Chains	Descriptive	Control over costs	-
Slade	1996	Gasoline refining & sales	Cross section Probit	Full service dummy	-*
Woodruff	2002	Footwear& sales	Cross section Probit	Frequent fashion change	-*
Brickley, Linck & Smith	2003	Banks & offices	Cross section Logit	Rural location	-*

* denotes significance at 5% using a two–tailed test.
Parentheses in the last column indicate that the variable examined is an inverse measure of the construct and is therefore expected to have the opposite effect on the extent of vertical integration.

Table 3: The Effect of the Importance of Upstream Effort on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Lafontaine	1992	Retail & services	Cross section Tobit	Weeks of training Lagged chain size Years before franchising	+* +* +*
Muris, Scheffman & Spiller	1992	Soft-Drink Bottling	Descriptive	National Accounts	+
Minkler & Park	1994	Retail & services	Panel Grouped Logit	Market minus book value	+*
Thompson	1994	Retail & services	Cross Section Regression	Years before franchising	+*
Scott	1995	Retail & services	Cross Section Regression	Days of training	+
Nickerson & Silverman	2003	Trucking Services	Cross Section Tobit	Advertising Expenditures	+*
Pénard, Raynaud, & Saussier	2003	Retail & services	Panel Tobit	Years before franchising	+*
Lafontaine & Shaw	2005	Retail & services	Panel Tobit	Advertising expenditures Advertising fee Years before franchising	+* +* +*

* denotes significance at 5% using a two–tailed test.

Table 4: The Effect of Outlet Size on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Brickley & Dark	1987	Retail & services	Cross section Regressions	Initial investment by franchisee	+*
Martin	1988	Retail & services	Panel Weighted LS	Average sales	+*
Norton	1988	Restaurants & motels	Cross section OLS, 2SLS	Average sales - Restaurants - Motels - Refreshment Places	 - -* -*
Brickley, Dark & Weisbach	1991	Retail & services	Cross section Tobit	Initial investment by franchisee	+*
Lafontaine	1992	Retail & services	Cross section Tobit	Initial investment by franchisee Average sales	+* +*
Ohanian	1994	Pulp and paper mills	Cross section Logit and Tobit	Paper Capacity	+*
Thompson	1994	Retail & services	Panel Regressions	Initial investment by franchisee	+*
Scott	1995	Retail & services	Cross section Regressions	Initial investment by franchisee	+
Kehoe	1996	Hotels	Cross section Tobit, Logit	Number of rooms	+*
Brickley	1999	Retail & services	Cross section Logit	Initial investment by franchisee	+*
Lafontaine & Shaw	2005	Retail & services	Panel Analyses Tobit	Employees per outlet	+*
Hortaçsu & Syverson[a]	2007b	Manufacturing plants	Panel Analyses	Value of plant shipments	+*

* denotes significance at 5% using a two–tailed test.

[a] Results relate to plants that are brought into vertical structures, whether they are upstream or downstream components of this structure.

Table 5: The Effect of Monitoring Cost on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Part 1		**Outcome Monitoring**			
Anderson & Schmittlein	1984	Electronic components & sales	Cross section Logit	Index of measurement difficulties	+*
Anderson	1985	Electronic components & sales	Cross section Logistic	Index of evaluation difficulties Importance of non-selling activities	+* +*
John & Weitz	1988	Industrial Goods & sales	Cross section OLS, Logit	Length of selling cycle	+*
Part 2		**Behavior Monitoring**			
Brickley & Dark	1987	Retail & services	Cross section Regressions	Distance from headquarters	-*
Norton	1988	Restaurants & motels	Cross section OLS, 2SLS	Rural population in state (%)	-*
Minkler	1990	Taco Bell restaurants	Cross section Logit, Probit, Linear Prob.	Distance from headquarters Outlet density	-* (-)
Brickley, Dark & Weisbach	1991	Retail & services	Cross section Tobit	Outlet density	(+*)
Carney & Gedajlovic	1991	Retail & services	Cross section Descriptive	Outlet density	(+*)
Lafontaine	1992	Retail & services	Cross section Tobit	Number of states in which operates	-*
Scott	1995	Retail & services	Cross section Regressions	Number of states in which operates	-*
Kehoe	1996	Hotels	Cross section Tobit, Logit	Number of same–chain hotels in city	(+*)
Baker & Hubbard	2003	Shipping & trucking	Panel First Dif, IV	Presence of on–board computer	(+*)
Baker & Hubbard	2004	Trucking & Trucks	Panel First Dif, IV	Adoption of on–board computer	(+*)
Brickley, Linck & Smith	2003	Banks & offices	Cross section Logit	Rural location	-*
Lafontaine & Shaw	2005	Retail & services	Panel Tobit	Number of states in which operates	-*

* denotes significance at 5% using a two–tailed test.
Parentheses in the last column indicate that the variable examined is an inverse measure of the construct and is therefore expected to have the opposite effect on the extent of vertical integration.

Table 6: The Effect of Non-Repeat Business on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Brickley & Dark	1987	Retail & services	Cross section Regressions	Non-repeat sector dummy Highway dummy (outlet)	+* -*
Norton	1988	Restaurants & motels	Cross section OLS, 2SLS	Household trips in the state - Restaurants - Motels - Refreshment Places	 + -* +
Brickley, Dark & Weisbach	1991	Retail & services	Cross section Tobit	Non-repeat industry dummy	-
Minkler	1990	Taco Bell restaurants	Cross section Logit, Probit Linear Prob.	Highway dummy	+
Brickley	1999	Retail & services	Cross section Logit	Non-repeat industry dummy Index: how local are your customers?	- -

* denotes significance at 5% using a two–tailed test.

Table 7: The Effect of Physical Capital Specificity on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Masten	1984	Parts & aerospace	Cross section Probit	Highly specialized dummy	+*
Masten, Meehan, & Snyder	1989	Parts & automobiles	Cross section regressions	1-10 scale of specificity	+
Lieberman	1991	Inputs to chemical products	Cross section Logit	Input is gas	+*
Masten, Meehan, & Snyder	1991	Naval Shipbuilding	Cross section regressions	Index of Specificity	+
Lyons	1995	Inputs to engineering firms	Cross section Logit	Survey index of specificity	+*

* denotes significance at 5% using a two–tailed test.

Table 8: The Effect of Human Capital Specificity on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Monteverde & Teece	1982	Parts & automobiles	Cross section Probit	Engineering design effort	+*
Anderson & Schmittlein	1984	Electronic components & sales	Cross section Logit	Index of specialized knowledge	+*
John & Weitz	1988	Industrial goods & distribution	Cross section regressions	Firm–specific training	+*
Masten, Meehan, & Snyder	1989	Parts & automobiles	Cross section regressions	1-10 scale of know how	+*
Masten, Meehan, & Snyder	1991	Naval Shipbuilding	Cross section regressions	Index of skill & knowledge specificity	+*
Hanson	1995	Apparel	Cross section OLS and Tobit	Degree of Standardization	+*
Woodruff	2002	Footwear & sales	Cross section Probit	Frequent fashion change	-*

* denotes significance at 5% using a two–tailed test.

Table 9: The Effect of Dedicated Assets on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Monteverde & Teece	1982	Parts & automobiles	Cross section Probit	Part specific to firm	+*
Lieberman	1991	Inputs to chemical products	Cross section Logit	Firm share of purchases	+
Acemoglu Aghion, Griffith & Zilibotti	2005	Manufacturing plants	Cross section Discrete Choice	Plant share of purchases	+*

* denotes significance at 5% using a two–tailed test.

Table 10: The Effect of Site Specificity on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Masten	1984	Parts & aerospace	Cross section Probit	Importance of colocation	+
Joskow	1985	Coal & electricity	Descriptive	Mine–mouth plant	+*
Masten, Meehan, & Snyder	1989	Parts & automobiles	Cross section regressions	Importance of colocation	-

* denotes significance at 5% using a two–tailed test.

Table 11: The Effect of Temporal Specificity on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Masten, Meehan, & Snyder	1991	Naval Shipbuilding	Cross section regressions	Importance of on–time availability index	+*
Pirrong	1993	Bulk Ocean Shipping	Descriptive	Market Thinness Potential	+
Nickerson & Silverman	2003	Trucking and Subcontractors	Cross Section regressions	Atypical haul weight measures	+*
Arruñada, González-Diaz, & Fernández	2004	Trucking	Cross Section OLS, Logit	Specialized Freight Dummy	+*

* denotes significance at 5% using a two–tailed test.

Table 12: The Effect of General Specificity on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Weiss	1992	Many vertical mergers	Descriptive	Residual correlation of returns	+*
González-Diaz, Arruñada, & Fernández	2000	Construction Firms and Subcontractors	Panel regressions	Index Capturing how many firms offer same product	+*
Ciliberto	2005	Physicians & hospitals	Panel Multinomial logit	Percent of local patients in HMO	+*

* denotes significance at 5% using a two–tailed test.

Table 13: The Effect of Complexity on Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Monteverde & Teece	1982	Parts & automobiles	Cross section Probit	Engineering design effort	+*
Masten	1984	Parts & aerospace	Cross section Probit	Dummy based on firm's classification	+*
Masten, Meehan, & Snyder	1991	Naval Shipbuilding	Cross section regressions	Index of complexity	U-shaped*
Woodruff	2002	Footwear & sales	Cross section Probit	Product heterogeneity	+*
Forbes & Lederman	2005	Major & Regional Airlines	Cross Section Logit	Major hub Weather	+* +*
Acemoglu Aghion, Griffith & Zilibotti	2005	Manufacturing plants	Cross section Discrete Choice	R&D intensity Upstream Downstream	 -* +*
Gil	2007	Movie Distribution	Cross Section Linear Prob. Model	Renegotiation Frequency	+*
Hortaçsu & Syverson	2007b	Manufacturing plants	Panel	Complex Inputs (Upstream Value Added)	+*

* denotes significance at 5% using a two–tailed test.

Table 14: The Effect of Uncertainty on Backward Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect on VI*
Walker & Weber	1984	Parts & automobiles	Cross section regressions	Index of volume & specification uncertainty (U)[a]	+*
Lieberman	1991	Inputs to chemical products	Cross section Logit	Variance of detrended sales	
				Upstream	+*
				Downstream	-
				Uncorrelated[c]	+*
Hanson	1995	Apparel manufacturers & suppliers	Cross section OLS and Tobit	Frequent style change (D)[b]	+*
González-Diaz, Arruñada, & Fernández	2000	Construction Firms and subcontractors	Panel regressions	Variation in Number of Workers (U)[a]	-

* denotes significance at 5% using a two–tailed test.
[a] (U) denotes upstream uncertainty.
[b] (D) denotes downstream uncertainty.
[c] Uncorrelated denotes upstream uncertainty that is uncorrelated with downstream sales.

Table 15: Assessment of Foreclosure and Raising Rivals Costs

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Finding*
Allen	1971	Cement & concrete	Descriptive	Acquisitions	Foreclosure
Reiffen & Kleit	1990	Railroads & terminals	Descriptive	Access to railroad terminals	No foreclosure
Rosengren & Meehan	1994	Challenged mergers	Event study	Returns, unintegrated downstream rivals	No foreclosure
Waterman & Weiss	1996	Cable TV programming & distribution	Cross sectional regressions	Program offerings	Fewer rival programs carried Foreclosure
Snyder	1996	Crude oil & refining	Event study	Returns, integrated rivals	Foreclosure
Mullin & Mullin	1997	Iron ore & steel	Event study	Returns, downstream consumers	No foreclosure Efficiency gains
Ford & Jackson	1997	Cable TV programming & distribution	Cross sectional IV regressions	Subscription price Program cost	Foreclosure Lower program cost No welfare change
Chipty	2001	Cable TV programming & distribution	Cross sectional IV regressions	Program offerings, price, & subscriptions	Fewer rival programs carried Foreclosure Efficiency gains outweigh losses
Hastings & Gilbert	2005	Gasoline refining & sales	Difference in difference	Wholesale price to unintegrated rivals	Foreclosure
Hortacsu & Syverson	2007a	Cement & concrete	Panel Difference in difference Probit	Concrete price Concrete production Plant survival	No foreclosure Efficiency gains

* denotes significance at 5% using a two–tailed test.

Table 16: The Consequences of Vertical Integration

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined (y)*	*Effect on y*	*Effect on W*
Shelton	1967	Restaurant	Panel Data Description	Costs Profit	+ +	+
Levin	1981	Crude oil & refining	Panel regressions	Profit Stability of profit	- -*	+
McBride[a]	1983	Cement & concrete	Regional panel	Delivered price	- *	+
Spiller	1985	Various	Cross section regressions	Financial gains Systematic risk	+* -	+
Helfat & Teece	1987	Various	Paired samples Difference in difference	Systematic risk	-*	+
Anderson	1988	Electronic Component sales	Cross section regressions	Index of opportunism	-*	+
Kerkvliet	1991	Coal & electricity	Panel regressions	Cost efficiency Exercise of monopsony power	+* -*	+
Muris, Scheffman & Spiller	1992	Soft drinks & bottlers	Panel regressions	Retail price	-*	+
Shepard[b]	1993	Gasoline refining & sales	Cross section Regressions	Retail price	-*	+
Ford & Jackson	1997	Cable TV programming & distribution	Cross section regressions	Program cost Price	-* +*	?
Edwards, Jackson & Thompson	2000	Crude oil & refining & pipelines	Panel Ordered probit	Stock rating	+*	?
Corts	2001	Film production & distribution	Cross section Tobit	Release date clustering	- *	+
Mullainathan & Scharfstein	2001	Chemical	Panel Regressions	Investment responsiveness	- *	?
Ciliberto	2005	Physicians & hospitals	Panel Regressions	Investment in health care services	+*	+
Jin & Leslie	2005	Restaurant Chains	Panel Regressions	Quality (health scores)	+*	+
Gil	2007b	Movie Distribution	OLS Duration Analysis	Movie run Length	+*	+

* denotes significance at 5% using a two–tailed test.

[a] Johnson & Parkman (1987) note that the introduction of time trends in the regressions renders effects documented here insignificant.

[b] results significant for unleaded sold full service only.

Table 17: Empirical Assessment of Divorcement

Author	*Year*	*Industry*	*Data/Technique*	*Variable Examined*	*Effect of Divorcement*
Barron, & Umbeck	1984	Gasoline refining & sales	Difference in difference	Retail price Station hours	Price higher Hours shorter
Slade	1998	Beer brewing & sales	Difference in difference	Retail price	Price higher
Vita	2000	Gasoline refining & sales	Panel	Retail price	Price higher
Blass & Carlton	2001	Gasoline refining & sales	Cross section	Retail cost	Cost higher
Hastings	2004	Gasoline refining & sales	Difference in difference	Retail price	No difference between CC & CD

* denotes significance at 5% using a two–tailed test.