# VERTICAL INTEGRATION

Paul L. Joskow[1]

## 1. INTRODUCTION

Understanding the factors that determine which types of transactions are mediated through markets and which within hierarchical organizations called firms has been an important subject of theoretical and empirical work in microeconomics generally and central to work in New Institutional Economics (NIE) in particular for at least the last 25 years. This essay reviews research that examines the choice between governance of vertical relationships involving suppliers of intermediate goods and services ("upstream") and the purchasers of those goods and services ("downstream") through some form of market-based contractual arrangement versus governance within an organization through vertical integration. My primary emphasis is on the transaction cost economics (TCE) framework for understanding the choice of governance arrangements, though I will briefly discuss several other theories of vertical integration as well.

The essay proceeds in the following way. First, I discuss the general comparative governance framework that is a basic feature of transaction cost economics (TCE) analysis and within which theoretical and empirical analysis of vertical integration associated with the NIE has proceeded. Next, I discuss traditional "neoclassical" theories of vertical integration that rely on market imperfections associated with market power, free riding, uncertainty, and economies of scale

---

as explanations for vertical integration.  I then proceed to discuss theories of vertical integration from a TCE perspective, focusing on the role of incomplete contracts and relationship specific investment.  This section includes by a brief discussion of the "property rights" (PR) approach to vertical integration, emphasizing the similarities and differences between the TCE and PR frameworks.  Finally, I provide a brief discussion of the extensive empirical literature on vertical integration that has been stimulated by this theoretical research, focusing on key methodological issues (Peter Klein's chapter in this volume reviews empirical studies of the "make of buy" decision in more detail).

## 2.  COMPARATIVE GOVERNANCE ARRANGEMENTS FROM A TRANSACTION COST ECONOMICS (TCE) PERSPECTIVE: THE BASIC STORY

I want to emphasize at the outset that there is not and will never be one unified theory of vertical integration. While the literature on vertical integration tends to focus on a simple dichotomy between the decision to "make" internally or "buy" through the market, from a TCE or NIE perspective we must be sensitive to the fact that there are a wide array of market-based governance arrangements that represent alternatives to both simple anonymous repeated spot market transactions and vertical integration.  These two governance arrangements are polar cases.  Theoretical and empirical research in the NIE tradition examines not only the determinants of the boundaries between firms and markets but also the origins of various "hybrid forms" of governance structure that lie between simple anonymous spot market transactions and unified hierarchical organizations with varying expanses of vertical and horizontal control. These hybrid forms include various types of long term contracts, joint ventures, dual sourcing (partial vertical integration), holding companies, and public enterprises.  In addition, NIE research examines the attributes of internal organizations

with different internal structures, incentive arrangements, vertical, horizontal and multi-product

dimensions (Williamson, 1996, 2000). Accordingly, vertical integration is only one of many

potential vertical governance arrangements that transacting parties may choose from and represents

only one component of broader theories of the governance of contractual relationships and theories

of the firm. Indeed, except in cases where there are significant market imperfections "…market

mediation is generally to be preferred over internal supply …" (Williamson, 1971, p. 113)

Virtually all theories of vertical integration turn in one way or another on the presence of

market imperfections on some type. Traditional approaches to vertical integration have tended to

focus (though not exclusively as I discuss further below) on vertical integration as a response to pre-

existing market power problems or as a strategic move to create or enhance market power in

upstream or downstream markets. While not excluding these rationales for vertical integration, the

NIE approach to the analysis of alternative market and internal organizational governance

arrangements is much broader. It focuses on a well-defined array of attributes of individual

transactions between buyers and sellers of goods or services and how they affect the performance

(total cost) of alternative governance arrangements. It recognizes that there is a wide array of

governance structures through which transactions can be mediated --- from anonymous spot markets

to internal administrative procedures within hierarchical organizations. It recognizes further that the

task of consummating transactions must confront a variety of potential transaction costs, contractual,

and organizational hazards, which are related to the attributes of the transactions at issue and their

interplay with the attributes of alternative governance arrangements. These transactions costs

involve the direct costs of writing, monitoring and enforcing contingent contracts as well as the costs

associated with the _ex ante_ investment _and_ _ex post_ performance inefficiencies that arise as a

consequence of contractual hazards of various types and various bureaucratic costs associated with internal organization. The inefficiencies of particular interest are those that arise as a consequence of ex post bargaining, haggling, pricing and production decisions, especially those that arise as the relationship must adapt to changes in supply and demand conditions over time, though inefficiencies in ex ante investments are also relevant. (Williamson, 1975, 2000) The governance structures that are chosen, whether market or hierarchical, are those that are best adapted to the attributes of the transactions of interest in the sense that they economize on the total costs of the trading relationship. That is, governance mechanisms are chosen in an effort to reduce inefficiencies associated with both ex ante investment and ex post performance of a trading relationship.

Contractual incompleteness, and its interaction with the attributes of different types of transactional attributes including asset specificity, complexity, and uncertainty, plays a central role in the evaluation of the relative costs of governance through market-based bilateral contracts versus governance through internal organization. When transactions are mediated through market-based contracts, circumstances may arise where the buyer and seller have conflicting interests. Consider the situation where transacting parties are locked-in to a bilateral trading relationship, in the sense that the potential aggregate value of continuing the bilateral relationship is higher than terminating it and turning to alternative buyers and sellers. In this situation one or both parties may have the incentive and ability to behave "opportunistically" to serve their own interests. The resulting bargaining over the terms and conditions of trade will affect both the distribution of the rents associated with this particular bilateral trading relationship and potentially the efficiency (quantities and production cost distortions) of the trades that are consummated ex post as well (reducing the rents about which the parties can argue as well as the aggregate value of the trading relationship ex

ante and ex post).     The potential advantage of internal organization in this case is that internal

organizations are likely to better harmonize these conflicting interests and provide for a smoother

and less costly adaptation process under these circumstances, facilitating more efficient ex ante

investment in the relationship and more efficient adaptation to changing supply and demand

conditions over time.  As Williamson (1971, pp. 116-117) observed many years ago:


> "…The contractual dilemma is this: On the one hand, it may be prohibitively costly, if not infeasible, to specify contractually the full range of contingencies and stipulate appropriate responses between stages.  On the other hand, if the contract is seriously incomplete in these respects but, once the original negotiations are settled, the contracting parties are locked into a bilateral exchange, the divergent interests between the parties will predictably lead to individually opportunistic behavior and joint losses.  The advantages of integration thus are not that technological (flow process) economies are unavailable to non-integrated firms, but that integration harmonizes interests (or reconciles differences, often by fiat) and permits an efficient (adaptive, sequential) decision process to be utilized…."


Accordingly, there do not exist, except perhaps at very high cost, complete contingent

contracts that can specify at the time a contractual relationship is being contemplated how each of

the parties will perform under all possible contingencies that could arise as the trading relationship

proceeds over time.  Contracts may be incomplete because of the direct costs of specifying and

writing contracts that anticipate all contingencies, because of "bounded rationality" that makes it

unlikely that the transacting parties can foresee all possible contingencies, and/or because of high

monitoring, verification, and enforcement costs.

Incomplete contracts per se do not necessarily lead to market inefficiencies.  It is the

interaction between contractual incompleteness and certain attributes of transactions that can lead

the parties to a trading relationship to become "locked-in" to the relationship once the relationship is

consummated. This in term can lead to adaptation problems that can adversely affect ex ante investment incentives and the ex post efficiency of the trading relationship. These potential problems are especially acute when supply and demand conditions that are uncertain ex ante change over time and the bargaining threat points of the parties to the relationship move outside of a "self-enforcing range" (Klein and Leffler, 1981) or "off the contract curve" (Williamson, 2000) anticipated in the design of the initial market governance arrangements or contracts.

In this regard, as this literature has developed, relationship specific investments of various kinds, when they are required to support an efficient trading relationship, have come to play a central, though not exclusive, role in creating bilateral trading relationships that are susceptible to ex post bargaining and contractual performance problems. As I will discuss in more detail below, relationship-specific investments are investments which, once made, have a value in alternative uses that is less than the value in the use originally intended to support a specific trading relationship. Once specific investments have been made a potential "hold up" or "opportunism" situation is created if the parties can bargain over the appropriable ex post quasi rents (the difference in asset values between the intended and next best use --- Klein, Crawford and Alchian, 1978; Williamson 1979, 1996) created by specific investments or must bargain or "haggle" to adapt to changing circumstances as the relationship proceeds over time. If contractual arrangements cannot be fashioned ex ante to mitigate these ex post incentives to bargain opportunistically without full regard for the total surplus produced by the relationship, ex ante incentives to make specific investments in the first place will be adversely affected and ex post performance and adaptation may be inefficient as well. To protect against these potential problems that increase the costs of transacting parties will explore the availability of alternative governance arrangements that reduce the costs of these

contractual hazards, stimulating more efficient investment incentives <u>ex ante,</u> more efficient contract execution <u>ex post</u> and, more generally, that reduce the overall cost of the relationship. Vertical integration is favored when the benefits of mitigating opportunism problems by moving the transactions inside the firm, by reducing <u>ex ante</u> investment and <u>ex post</u> performance inefficiencies, are greater than other sources of static and dynamic inefficiency associated with resource allocation within bureaucratic organizations.

## 3. TRADITIONAL APPROACHES TO EXPLAINING VERTICAL INTEGRATION

The explanation of the causes and consequences of vertical integration that emerged in the field of industrial organization during the post-World War II period was heavily influenced by the sharp distinction drawn by neoclassical economics between resource allocation mediated through markets and resource allocation taking place within private firms and related types of hierarchical organizations (e.g public enterprises). Microeconomics in general and applied price theory in particular were concerned with the way anonymous spot markets worked to allocate resources. The factors that determined the boundaries between firms and markets were largely ignored and issues associated with the internal organization of firms and the way firms allocated resources internally were, with a few exceptions (Simon, 1947; Cyert and March 1963; Arrow, 1974), viewed as outside of the domain of economics. Firms were conceptualized as production sets that defined the technologically most efficient opportunities to transform inputs into outputs. They relied on anonymous spot markets to buy and sell inputs and outputs. That is, what firms did and what markets did were *complementary* activities. Coase's (1937, 1972) view that firms and markets were *substitute* governance mechanisms was not an accepted part of received wisdom until relatively

recently. Precisely what was in a firm's production set and what was not was, at best, rather vague and there existed no meaningful economic theory to explain where to draw the line between firms and market transactions or to explain the diverse types of "non-standard" contractual arrangements observed in the real world.

Industrial organization theorists like Bain (1956, 1959) viewed the relevant firm production set rather narrowly as encompassing activities that were clearly physically related to one another. Both multi-plant economies and vertical integration downstream and upstream were generally viewed as being unnecessary for a firm to produce at minimum cost in the absence of technological relationships that physically joined production between plants. Instead, the presumption was that vertical integration, and non-standard vertical contractual arrangements, reflected the presence of market power somewhere in the system and/or efforts to create or exploit market power. Vertical integration could be a profitable response to costs of successive monopolies (e.g. double marginalization and related "vertical externalities" Tirole, 1988, Chapter 4), or it could facilitate price discrimination in a variety of different ways (Perry, 1978), or vertical integration (and long term contracts) could be used strategically to soften competition in the short run by raising rivals' costs or in the long run by increasing the costs of entry to foreclose rivals that might otherwise enter the market (Aghion and Bolton, 1987; Ordover, Salop and Saloner, 1990; Hart and Tirole, 1990).

Moreover, vertical integration itself was viewed theoretically as being "costless." That is, no internal organization costs were recognized, but only any costs realized through distortions in market prices, quantities, or the factor proportions used to produce output from a neoclassical production function. "Costless" vertical integration was also used as a benchmark against which alternative "costless" contractual arrangements could be compared. So, for example, distortions arising in one

way or another as a consequence of double marginalization can be "solved" through vertical integration as well as with alternative (costless) contractual arrangements --- two-part tariffs, maximum retail price maintenance, quantity forcing contracts, requirements contracts, service obligations, etc. (Tirole, 1988, Chapter 4). Absent transactions costs of some type, the alternative instruments are all equally attractive mechanisms for responding to double marginalization. Price discrimination could be accomplished with vertical integration or with contracts that prohibited resale and eliminated the associated arbitrage that could otherwise undermine price discrimination. The potential advantages and disadvantages of alternative contractual responses were discussed primarily in the context of their effects on market prices and quantities and their ability to allocate risk more or less efficiently between parties with different degrees of risk aversion. There was nothing embedded directly in the neoclassical theoretical approach to vertical relationships that allowed for attributes of the contractual and organization arrangements themselves to provide a basis to choose among alternative governance arrangements because there were no transactions costs associated with any of the institutional alternatives identified.

Another potential source of incentives for vertical integration is the free rider problem associated with the provision of pre-sale information and post-sale service by competing downstream retailers (Telser, 1960; Mathewson and Winter, 1986). If retailer's cannot fully appropriate for themselves the benefits of retail service expenditures but instead see some of the benefits accrue to their downstream competitors this "horizontal externality" (Tirole, 1998, Chapter 4) will lead downstream retailers to under-invest in retail service (at least from the perspective of the manufacturer). Vertical integration is one potential solution to this problem. So too are various combinations of exclusive territorial agreements, resale price maintenance, profit pass-over contracts

and other mechanisms. The unanswered question is how to choose among the alternative institutional arrangements in a systematic way?

Dennis Carlton (1979) has shown how the combination of uncertain demand for inputs and the failure of markets to be cleared by spot prices under some contingencies can create a private incentive for downstream firms to integrate backwards partially or fully for "supply security" reasons (See also Malmgren, 1961; Arrow, 1975; Green, 1986; and Bolton and Whinston, 1993 for related theoretical work). And there is abundant support in the business history literature for such a motivation for vertical integration (Chandler, 1964, p. 84). "The strong incentives for vertical integration arise because the vertically integrated firm is able to satisfy high probability demand by itself, and pass on the low probability demand to some other firm" (Carlton, 1979; p. 207). Williamson (1971, p. 117) points out that "…arguments favorable to vertical integration that turn on 'supply reliability' considerations commonly reduce to the contractual incompleteness issue (footnote omitted)." Moreover, it is not clear that the market imperfections that create the incentive to vertically integrate here could not be equally well (or even better) mitigated by downstream firms by arranging a portfolio of fixed price and spot market contracts.

Finally, George Stigler (1951) proposed a theory of vertical integration based upon Adam Smith's famous theorem that "the division of labor is limited by the extent of the market." Stigler advanced a dynamic or life-cycle theory of vertical integration. He argued that in an infant industry producing a new downstream product vertical integration would be more likely to occur because the demand for specialized inputs would be too small to support independent firms supplying intermediate goods. As the demand for the new product grows, intermediate good suppliers whose production is characterized by increasing returns would be spun off as independent firms supplying

inputs to multiple competing downstream suppliers. The empirical prediction is that as industries grow the extent of vertical integration should decline and as industries contract vertical integration should increase. Eberfeld (2002) argues that Stigler's theory is correct as long as there are no barriers to entry. Stigler's theory turns primarily on economies and diseconomies of scale and the implicit assumption that suppliers of new products require specialized inputs. It ignores transactions costs associated with both internal organization and market contracting. The theory has found little empirical support.

There is clearly no shortage of theories identifying potential incentives for vertical integration. This should not be surprising. As long as it is assumed that there are no additional costs associated with internal organization, almost any market imperfection necessarily becomes a candidate for creating private incentives for vertical integration. However, this approach ignores both the costs of internal organization and other costs of more complex contractual alternatives to either simple linear spot market contracts or vertical integration. In principle, the TCE framework should be able to encompass these traditional sources of market failure that have been identified as factors that increase incentives for vertical integration as well, and to do so in a richer and more systematic fashion. However, the emphasis of TCE to date has been on looking for other than traditional vertical and horizontal externality, foreclosure, uncertainty and risk allocation explanations for vertical integration and nonstandard vertical restraints rather than trying to incorporate these considerations into the analysis. This is a deficiency that can be and should be remedied (Joskow, 1991, 2002).

## 4. INCOMPLETE CONTRACTS AND ASSET SPECIFICITY

As noted earlier, one of the foundations of TCE is the recognition that contracts are incomplete and potentially lead to contractual hazards that adversely affect _ex ante_ investment incentives and the efficiency of _ex post_ performance, especially in response to adaptations required to respond to changing supply and demand conditions. These problems arise when the parties to a contractual relationship find themselves in a bilateral bargaining position _ex post_ as a consequence of lock-in as discussed earlier. While there are several potential sources of lock-in that lead to potential bilateral bargaining and adaptation problems, much of the theoretical and empirical work in the TCE tradition has focused on relationship specific investments (asset specificity) and/or the interaction between asset specificity and other transactional attributes such as uncertainty, product complexity, and information asymmetries. Accordingly, I will focus the discussion here on the role of asset specificity.

Before proceeding, it is useful to outline an idealized set of steps that leads to a well functioning vertical relationship to provide a framework for further analysis. The actual structural attributes of each of these steps and their implications for the costs of the feasible set of alternative governance arrangements then becomes the focus of comparative governance analysis. How would an idealized contractual relationship be structured? First, the responsibilities and authorities of the parties to the transaction (including the appropriate assignment of property rights) would be defined _ex ante_. Second, contractual provisions would need to be agreed to _ex ante_ that align the parties' incentives so that they have a mutual interest in performing in conformity with the intent of the contractual agreement. These contractual provisions include contractual formulas for adjusting prices and quantities and conditions change, cost and profit sharing provisions, assignment of investment responsibilities, financial guarantees and collateral requirements, etc., (Joskow, 1988b).

Under normal conditions these provisions are incentive compatible and self-enforcing. Finally, the parties may agree to a process through which the terms of the contract can be adjusted (a renegotiation process) to facilitate smooth adaptation to changing circumstances that were not fully anticipated in the original terms of the contract. By facilitating realignment the costs of haggling over the consequences of changed circumstances can be reduced. These governance design challenges are confronted in a broader context that includes, among other things, the effects of parties' behavior on the value of reputational capital, and the external financial constraints on their behavior that reputational considerations imply.

What departures from these idealized contractual attributes might emerge in practice and increase the cost of consummating the transactions covered by the contract? The assignment of responsibilities, authorities or property rights may be incorrect or fail to cover all contingencies. The performance incentives may not work as intended under all contingencies and create ex post rent extraction opportunities and bargaining inefficiencies. The terms of the contract may be difficult to adapt to significant changes in economic circumstances leading to further rent extraction opportunities and performance inefficiencies. Imperfections in each step of the institutional design process can lead to distortions in ex ante investment incentives and ex post performance and adaptation inefficiencies. These imperfections in turn will vary based on the attributes of the transactions covered and the relative strengths and weaknesses of alternative governance arrangements in supporting a smooth trading relationship over time. As I will discuss in more detail below, the TCE approach takes into account all aspects of the trading relationship and compares their performance attributes under alternative governance arrangements, including internal organization. It emphasizes ex post adaptation problems, but also recognizes potential

inefficiencies in ex ante investment.

### *Asset Specificity in Detail*

As noted earlier, a relationship specific investment is an investment which once made (sunk) by one or both parties to an ongoing trading relationship has a lower value in alternative uses than it has in the intended use supporting this specific bilateral trading relationship. In the extreme, an investment made by a supplier in anticipation of supplying a product to a particular customer could be worthless if used to serve any other customer. More generally, we can think of specific investments as having significantly lower values, or producing lower gains from trade, when employed other than in supporting the intended relationship with a particular customer or supplier. Such investments create a bilateral dependency after they have been sunk defined by the difference between the value of the investment in its intended use and its next best alternative use. The parties to the transaction may then have an incentive to haggle over the distribution of the ex post quasi rents created by the specific investments. It is the problem of economically protecting the aggregate value of the specific investments from being reduced through this potential haggling that drives the choice of governance structure.

Asset specificity that is directly relevant to vertical integration is thought to arise in a number of different contexts (Williamson, 1983; 1996, Chapter 4):[2]

1. *site specificity*: The buyer and the seller are in a "cheek-by-jowl" relationship with one another, reflecting ex ante decisions to minimize inventory and transportation expenses. Once sited the assets in question are highly immobile. A mine mouth coal plant (Joskow, 1985, 1987) or a

---

[2]  Masten, Meehan and Snyder (1991) identify "temporal asset specificity" as a sixth category. Williamson (1996, p. 106) argues that this is a form of site specificity and I agree with his assessment.

bauxite processing plant and the associated mines (Stuckey, 1983) are examples of site specificity.

2. *physical asset specificity*:  When one or both parties to the transaction make investments in equipment and machinery that involve design characteristics specific to the transaction which have lower values in alternative uses.  A boiler that has been designed to burn a particular type of coal (Joskow, 1985) and investments in tools and dies to produce parts that can be used in a specific downstream manufacturer's products (Klein, Crawford and Alchian, 1978;  Klein, 1988) have this characteristic.

3. *human asset specificity*:  When, as a consequence of learning by doing, workers accumulate relationship specific human capital that makes it possible for them to produce goods and services more efficiently than can otherwise equivalent workers who do not have this firm specific human capital.  Such human capital is of particular value to the suppliers and customers that benefit from it, and is of lower value to the workers (or the firms they work for) if not utilized to support the specific relationship within which it accumulated.  Design engineers who have developed special skills in designing a particular type of aircraft or automotive components are examples of human asset specificity (Monteverde and Teece, 1982; Masten, Meehan and Snyder, 1989).

4. *dedicated assets*:  General investment by a supplier that would not otherwise be made but for the prospect of selling a significant amount of product to a particular customer.  If the relationship is terminated prematurely, it would leave the supplier with significant excess capacity and a lower price to support the investment would be realized ex post than had been ancticipated ex ante.  The development of a large natural resource deposit in a remote location to supply a large upstream user is an example of dedicated assets (Joskow, 1985).

5.  *intangible assets*:  Although specific investments are most frequently conceptualized as

either physical investments or relationship specific human capital, intangible capital such as brand name loyalty can have relationship specific attributes. For example, McDonalds has significant brand name value which has accumulated over time through investments in product quality, advertising and promotion. The value of these investments is tied completely to the McDonalds brand name. In order to sell its products, however, McDonalds must convey the use of its valuable brand name to its distribution outlets, some of which it owns (vertical integration) and some of which are independent franchisees.

### *Effects on ex ante investment and ex post adaptation and performance*

The combination of incomplete contracts and relationship specific investments can have adverse effects on both ex ante investment and the efficiency of ex post performance by creating a bilateral trading situation in which the parties' bargaining in their own individual self-interest leads to a reduction in the overall size of the pie. The property rights literature (Grossman and Hart, 1986; Hart and Moore, 1990; Hart 1995) discussed below focuses on ex ante investment incentives and assumes that transactions are consummated efficiently ex post. The only inefficiencies that arise from improper alignment of property rights are reflected in distortions in ex ante investments, and these investment distortions are the primary focus of the analysis. Under these and other assumptions (Maskin and Tirole, 1999) the proper assignment of property rights or decisionmaking authority can mitigate hold-ups over the ex post division of appropriable quasi rents and the associated adverse effects on ex ante investments. On the other hand, TCE is concerned with both ex ante and ex post inefficiencies that arise in bilateral trading relationships, though Williamson has generally placed more emphasis on ex post haggling and associated inefficiencies (Williamson, 1971, 1975, 2001; Klein, 2002) than on ex ante investment distortions. This emphasis, in turn, apparently reflects the

view that internal organization is most likely to be superior to market contracting in circumstances where ex post bargaining/haggling costs associated with market contracting are high since what internal organization is good at is harmonizing the otherwise conflicting interest of the parties to the transaction and facilitating a smooth and efficient adaptation to changing supply and demand conditions..

For ease of exposition, I focus first on a simple model in which contractual incompleteness leads to an ex post expropriation of quasi-rents but not to inefficiencies in ex post trade. That is, while there is bargaining over the appropriable ex post quasi rents, and this affects ex ante investment incentives, given a particular level of ex ante investment the efficient quantity is produced and traded ex post. As we shall see, contractual incompleteness plus asset specificity leads to underinvestment ex ante.

Before relationship specific investments are made to support a trading relationship, let us assume buyers and sellers may have numerous potential suppliers and customers among whom they can choose to enter into an ongoing relationship (though there is no reason why there cannot be small numbers bargaining ex ante and "smaller numbers" bargaining ex post to incorporate pre-existing market power considerations as reflected in the traditional vertical externality and foreclosure literatures discussed above). That is, the markets are reasonably competitive ex ante. However, once a relationship is developed and relationship specific investments are made to support it, a competitive bargaining situation ex ante is transformed into a small numbers or bilateral monopoly bargaining position ex post. The small numbers bargaining situation results from the investment in relationship specific assets. Once these investments are made, the fact that they have lower values in alternative uses creates a stream of potentially appropriable quasi rents (the

difference between the value in the intended uses and in alternative uses).  It is these quasi rents over which buyers and sellers may haggle and bargain in the absence of a complete and easily enforceable contract that defines clearly the rights and obligations of each party when various contingencies arise.  The potential for ex post bargaining over these quasi rents in turn affects the expected returns from the initial investment in specific assets and the associated incentives to invest ex ante.  Accordingly, the level of investment in relationship specific assets will differ from what would maximize the aggregate gains from trade that potentially could be achieved absent opportunistic behavior.  This increases the total costs of providing the goods and services being traded and is a social cost of opportunistic behavior.

The nature of the problem that arises when specific investments must be supported by incomplete contracts can be articulated more precisely with the help of a set of very simple analytical examples which are drawn from Tirole (1988, pp. 25-28).  Assume that we have a buyer that is interested in acquiring one unit (for simplicity) of a product with particular characteristics.  In order to meet the buyer's requirements efficiently, a supplier must make investments that are specific to the product that the buyer desires.  Once these investments are sunk, their value in producing products for other buyers is lower than it is when the investments are employed to produce output for the intended buyer.  The buyer and its supplier agree to enter into a supply relationship in period 0, the supplier must make its investment in period 1, and production and trade take place in period 2.  (In this sense we have a "long term" relationship.)

Let I be the amount of the supplier's investment and $C(I)$ the ex post costs of producing the product to this specific buyer's satisfaction once the investment has been made, where $C'(I) < 0$ and $C''(I) > 0$.  Assume initially that the value of the investment is zero if it is employed in alternative

uses. The value of the output to the buyer is $v > C(0)$. Finally, assume that the parties do not enter into a contract prior to the supplier's investment, but instead agree to bargain over the price when the product is finally delivered.

To solve for the transaction price in period 2 and the supplier's expected profit maximizing investment in period 1 we must specify how the transactions price is determined ex post. Assume that the buyer and seller negotiate a price that splits the ex post gains from trade evenly between them once the investment has been made (i.e. Nash bargaining) and that the product is (efficiently) supplied given the level of specific investment that is forthcoming ex ante. Then the ex post price is given by:

$$P(I) - C(I) = v - P(I) \quad or \qquad\qquad (1)$$
$$P(I) = [C(I) + v]/2$$

and the supplier's profit is given by:

$$max_I \ [P(I) - C(I) - I] \ = max \ [v/2 - C(I)/2 - I] \qquad\qquad (2)$$

Equation (2) shows that from the supplier's perspective a $1 ex ante investment in cost reduction yields only a $0.5 return to the supplier. The rest of the ex post surplus is captured or "held up" by the buyer. Accordingly, the privately optimal investment in cost reduction is lower than the socially optimal investment in cost reduction. Solving equation (2) yields the result that the supplier invests up to the point where:

$$-C'(I) = 2 \qquad\qquad\qquad (3)$$

while the optimal investment is defined by the maximization of the total surplus *[v - C(I) - I]*, not

just the supplier's slice of the surplus.  The socially optimal investment is then given by:

$$max\ [\ v - C(I) - I]\ \ or \qquad\qquad\qquad (4)$$

$$-C'(I) = 1 \qquad\qquad\qquad (5)$$

Since *C(I)* is convex, the supplier chooses a level of investment in cost reduction (equation (3)) that

is too low compared to the level of investment that would maximize the gains from trade (equation

(5)).  The difference in production costs associated with these different investment levels is the cost

of contractual hazards arising from the combination of the need for specific investments to support

cost minimizing trade between the buyer and the seller and the absence of a contract that defines

what the terms of trade will be ex post.  The example's assumption that a specific investment has no

value in alternative uses is, of course, extreme.  However, the problem continues to emerge as long

as the value of the specific investment in alternative uses is less than its value in the intended

relationship (Tirole, 1988, p. 25.)

If the investments in specific assets could somehow be protected from this type of ex post

hold-up problem, the supplier would be willing to increase investment in specific assets and reduce

the overall cost of the trading relationship.  The challenge in this case is to find a set of institutional

arrangements that mitigate the hold-up problem that leads to underinvestment <u>ex ante</u> without incurring other costs that more than exceed the increased surplus resulting from additional investment in cost reducing specific assets.

Of course, in our example, we assumed that there was no contract negotiated by the parties in period 0, before the specific investments were made, which defined the terms and conditions of trade in period 2. We have assumed the ultimate in incomplete contracts --- no <u>ex ante</u> contract at all. The most obvious way to solve the problem that is identified in the example above is to allow the parties to enter into a binding contract prior to the time that the supplier makes its investment decisions. In particular, the contract could specify that the buyer would pay a price $p = P^*$ in period 2 where $P^*$ is any price that lies between $v$ and $C(I) + I$. The critical assumption is that $P^*$ does not depend on $I$, but is fixed <u>ex ante</u>. For example, $P^*$ could be arrived at in period 0 through competitive bidding by suppliers competing to gain the right to supply the goods required by the buyer in period 2. If such a contract could be written and enforced costlessly, our problem would be solved. There would be no <u>ex post</u> bargaining opportunities and the supplier would have precisely the correct investment incentives since its payoff is given by $(P^* - C(I) - I)$ whose maximum satisfies equation (5) above. This is the ultimate complete contract and eliminates the contractual hazard that was previously identified with no contract.

### *Ex post contractual adaptation and performance problems*

The departure of TCE is to recognize that this type of complete contract will be impossible to write in many interesting situations. For example, even in this simple case, let's say that the seller refuses to deliver the product unless a price higher than the agreed price is paid, perhaps reflecting the quasi rents that could be expropriated if there were no contract at all. The buyer could go to court

to enforce delivery at the contract price, but court enforcement can be costly, uncertain and time consuming. In addition, if it turns out that the supplier is unable to deliver in a timely way (specific performance), damages must be calculated which gets the courts involved in potentially complicated valuation deliberations with an uncertain outcome. These "haggling costs" both reduce ex post surplus and can affect its division among the parties reducing both ex ante incentive to invest efficiently and the ex post value of the relationship given any particular level of investment.

The latter observations suggest that the assumption that ex post bargaining is efficient and, as a result, that the only inefficiencies arise with regard to ex ante investments is not a particularly realistic assumption. More generally with uncertainty and asymmetric information (Tirole, 1988, pp. 21-24; Myerson and Satterthwaite, 1983; Williamson, 1975, 2000) in the presence of bilateral monopoly ex post, bargaining will generally be inefficient. These inefficiencies in turn reduce the quasi-rents available to cover investment costs which further distorts ex ante investment incentives. The overall costs of the relationship increase do to the combination of ex ante and ex post inefficiencies. As already noted, Williamson (2000) argues that it is with regard to the ex post bargaining and adaptation inefficiencies where the action lies in terms of the opportunities for internal organization to reduce transactions costs and become a lower cost alternative than even the best available market contracting alternative.

Of course, real world transactions are often more complicated than what is captured in these simple examples, but the basic problems created by ex post lock-in and incomplete contracts are similar. For example, the nature of the technology for producing the goods or services at issue may lead to a situation where cost minimizing exchange requires complementary relationship specific investments by both the buyer and the seller (there may be other effective governance-related

reasons, rather than technological reasons, for a contract to require both the buyer and the seller to make specific investments to mitigate incentives to behave opportunistically --- Williamson, 1983). This not only changes the attributes of the ex post bargaining space, but also complicates the effects of outside opportunities on ex ante investment decisions. When we introduce uncertainty about future production and investment costs, uncertainty about the buyer's ex post valuation and the quantities of the product required, the need for bilateral investments in specific assets to be made by both parties in order to support an efficient trading relationship, and a product quality dimension, we face a much more significant contracting and enforcement problem than in the simple model presented above. Correspondingly, it becomes more and more likely that it will be extremely costly or even impossible to write credible complete contracts that specify ex ante how the buyer and seller will behave when any contingency arises; or to design an associated enforcement mechanism that will require the performance promised or assess damages for non-performance without distorting behavior and increasing the total costs of the transactions at issue. Moreover, complex long term contracts aimed at tying the hands of the parties so that they cannot behave opportunistically when foreseeable contingencies arise may also embody costly rigidities and have poor adaptive properties when contingencies not specifically provided for in the contract arise (Joskow, 1988, 1990; Williamson 1996, Chapter 4). Accordingly, while complex long term contracts carry potential benefits by better protecting against the opportunistic behavior associated with specific investments than would simpler but more incomplete contracts, they also incur potential adaptation costs when unanticipated contingencies arise. Inefficiencies associated with ex ante investment distortions and ex post contract performance problems will increase and internal organization will become a relatively more attractive governance structure.

On the other hand, there may be other constraints operating to affect <u>ex post</u> bargaining power that may mitigate the hold up problem and the effects of opportunism on <u>ex ante</u> investments. In particular, the presence of reputational capital and the potential erosion of its value by opportunistic behavior may operate to mitigate such behavior (Williamson, 1975, chapter 9; Klein and Leffler, 1981; Klein, 2002) and facilitate efficient relational contracting in the presence of specific investments and incomplete formal contracts.

## 5. VERTICAL INTEGRATION AS AN ALTERNATIVE GOVERNANCE STRUCTURE

These considerations help to explain why we observe a wide array of contractual arrangements in the real world that sometimes look very different from the "standard" anonymous spot market transaction that is featured in elementary and intermediate micro economics textbooks. However, these more complex contractual arrangements are unlikely to protect completely against the opportunistic behavior associated with specific investments and other sources of <u>ex post</u> lock-in, and necessarily incur negotiating, monitoring, enforcement and adaptation costs when changed circumstances push the threat points of the parties outside of the "self-enforcing range." Vertical integration represents an alternative governance structure to bilateral contracts for mediating the supply of a product that requires specific investments to support cost minimizing exchange. Rather than fiddling with contractual protections to mitigate the inherent conflicts of interest that may arise between independent buyers and sellers in the presence of specific investments, and dealing with other distortions and rigidities that such contracts may entail, the buyer may choose instead to integrate backward (or the seller integrate forward) into the supply of the input at issue (or sale of the downstream good). By so doing the parties to the transaction would substitute internal hierarchical

governance mechanisms within a firm for governance by market-based bilateral contracts. The vertically integrated firm's managers (arguably) will pursue a common objective to maximize the firm's value. They are not expected to engage in the kinds of opportunistic behavior associated with specific investments and an incomplete contract between an individual buyer and seller each pursuing its individual firm's objectives.

If we were to apply the traditional methodology for analyzing incentives for vertical integration in response to market power and other standard market imperfections as discussed above (e.g. Tirole 1988, Chapter 4), the vertically integrated firm would simply be modeled as a costless entity having the objective function specified in equation (4) above. That is, the vertically integrated firm would have the incentive to maximize the aggregate gains from trade associated with the transactions at issue without any offsetting costs of internal organization. The contractual hazard associated with the specific investment is removed because there is no longer an economic agent in the picture that has the incentive or ability to haggle over the distribution of the ex post quasi rents. This approach assumes that both pieces of the ex post quasi rent pie are internalized into the unified objective function of the vertically integrated firm that can execute the associated transactions costlessly and efficiently.

This analytical approach is not fully satisfactory. As Coase has observed (1937), this approach implies that vertical integration will always either have superior efficiency properties or equivalent efficiency properties to decentralized trade between independent buyers and sellers in a market. Why then doesn't all economic activity take place within vertically integrated hierarchies with a single objective function? The problem is that this basic approach to analyzing vertical integration doesn't tell us what it is about internal organization that makes it a superior governance

structure to imperfect market transactions in some circumstances circumstances but not in others. The attributes and associated costs of allocating resources within internal organization are missing.

An important difference between internal organization and market contracting is the nature of the delegation of authority to make decisions when contingencies arise which could not otherwise be contracted on effectively through bilateral contracts. The property rights approach focuses on ownership of physical and intangible assets (but not human capital that accrues to individual workers) where ownership carries with it the authority to determine how these assets will be used (Grossman and Hart, 1986; Hart and Moore, 1990; Hart 1995). Ownership of specific investments (e.g. through vertical integration) gives the owner the residual authority to use the assets to further the owner's objective function. While negotiations between managers within a firm may arise, the firm establishes clear lines of authority to resolve them. Ownership and the rights of control that go along with it change the status quo bargaining point within the firm and the ultimate allocation of the rents over which the bargaining takes place. That is, when specific investments are involved, ownership of the specific assets allocates the residual rights of control to the party that makes the specific investment. The owner then has the authority over the ex post trading decision and any internal transfer prices that may be relevant. Hart (1995) shows how various combinations of specific physical and human capital can affect the allocation of resources under alternative ownership arrangements. The residual rights of control that are conveyed by ownership affect the ex post distribution of surplus which in turn affects the ex ante incentives to invest.

The property rights approach strips the firm of most of its organizational features and focuses on how ownership and the associated residual rights of control affect the bargaining power of otherwise self-interested economic agents engaged in bilateral trade. This approach does not allow

for any other changes in incentives and behavior of the transacting parties when the relationship is brought from the market inside of the firm (vertical integration). Thus, it largely ignores important differences between market transactions and internal organization other than simply a change in relative bargaining power between self-interested managers (Williamson, 1996, Chapter 4). However, the objective functions possessed by managers and the incentive and payoff structure that they face are different for managers within a firm as compared to managers in separate firms. One of the key tasks of management is to develop monitoring and financial incentive arrangements within the firm that induce the managers and employees to pursue the interests of the firm rather than the interests of a hypothetical independent division of the firm producing for its own account (Williamson, 1985, Chapter 6; Holmstrom and Milgrom 1990; Williamson, Wachter and Harris, 1975). These incentive arrangements include compensation contracts that partially tie compensation to overall firm performance and the effects of employee behavior on promotion opportunities and continued employment. In short, other things equal, the incentive and ability of a manager within a firm to exploit specific investments to hold up another division is different from what it would be if the managers were managing two independent firms.

Monitoring behavior and the costs and distribution of information are also likely to be different within a firm than between independent firms (Williamson, 1975; Arrow, 1975; Hart, 1995, p. 72). Employees within a firm have different incentives and obligations to reveal information to senior management from those of employees of an independent firm and can be subject to swifter and different penalties for hiding information (e.g. termination). Moreover, senior management has authority to use a variety of monitoring and information gathering mechanisms that can be matched quickly to problems as they arise without adhering to formal (incomplete) auditing contracts. The

internal auditing departments of large firms have substantial authority to range far and wide in identifying behavior that is inconsistent with the firm's objectives. Accordingly, internal organization is likely to be better at obtaining those types of information that are directly relevant to monitoring, and controlling the opportunistic behavior by the firm's managers that would otherwise arise from the combination of asset specificity and incomplete contracts if the transactions took place between independent firms.

Internal organization can also rely on more informal and less time consuming procedures to resolve conflicts inside the firm than would independent agents bound by formal contracts. As Williamson (1996, Chapter 4) observes, the internal "contract law" within a firm is quite different from the arbitration and litigation procedures to which independent economic agents would have to turn if they could not resolve disputes. The latter can be a costly and time consuming process that typically involves a third party decision maker that must become informed about the issues de novo. The internal decision maker, whether the CEO or the relevant division manager, can utilize simpler and faster internal procedures for resolving conflicts between divisions and also is likely to come to the problem with much more information of relevance than would a third party arbitrator.

If hierarchical organizations have these attractive properties, why don't we see more economic activity taking place within very large organizations rather than through markets? The answer is that internal organization is good at some things, but not at others. Williamson (1996, Chapter 4) observes that when we look at the bigger dynamic picture, internal organization is a last resort that we turn to only in the presence of significant contracting hazards and associated transactions costs. This is because, opportunistic behavior associated with specific investments aside, decentralized market arrangements have superior adaptive properties to internal organization

in many <u>other</u> important dimensions. Differences in the information structure between market and hierarchical governance structures which help to mitigate opportunism problems associated with specific investments may lead to inefficiencies in other dimensions (Hart, 1995, pp. 71-72). For example, employees may be less willing to reveal information that adversely affects their promotion possibilities or continuing employment. The kinds of low-powered incentives that characterize internal compensation arrangements may also mute incentives to exert the optimal amount of worker effort (Williamson, 1985, Chapter 6; Holmstrom and Milgrom, 1990). In addition, while internal organization is likely to be better at removing certain kinds of internal information asymmetries in the short run, it may be an inferior structure for obtaining, processing and using external information about prices, costs, quality, and technological change in the long run compared to repeated market transactions. For example, when a firm vertically integrates (or enters into a very long term full requirements contract) it is likely to lose some of the benefits associated with continually examining and accessing outside opportunities through repeated contracting. These opportunities include information about the "least cost" prices of the goods and services that the firm is producing internally and the availability of new technologies and production methods. While there is nothing that prohibits a vertically integrated firm continuing to look to outside opportunities to benchmark its performance, an internal division in competition with outside sources may have strong incentives to hide or misrepresent outside opportunities in order to protect itself from external competition. This type of organizational opportunism is different from the kinds of hold-up problems created by specific investments, but may be even more costly in the long run.

For these reasons, even in the face of significant contractual hazards resulting from specific investments and incomplete contracts, firms may still find it advantageous to continue to rely on

arms-length market transactions for all or a fraction of their input or distribution requirements (dual sourcing) involving specific investments rather than turning to complete vertical integration. This choice may be made to provide management with external information that it can use to assess the performance of its internal divisions and to counteract the costs of various types of internal organizational inefficiencies. Competitive market prices convey a tremendous amount of information that is difficult to reproduce using internal accounting cost and auditing information. Moreover, this information is updated very quickly as supply and demand conditions change if a firm is in the market repeatedly. As organizations get larger the volume of auditing information that must be processed by management grows non-linearly with the size and scope of the firm (Williamson, 1975) and becomes more difficult to use to control costs and quality effectively and to adapt to changing market conditions. The potential shirking problems resulting from low power internal compensation incentives are also likely to become more significant as monitoring becomes more difficult in large organizations.

There are other dynamic considerations that may make the relative attractiveness of alternative governance arrangements in a particular industry or transactional setting change over time (Langlois and Robertson, 1989). As revealed by the extensive analysis of the GM-Fisher Body relationship (Klein, Crawford and Alchian, 1978; Klein, 1988, 2000, 2002) a detailed long term contract involving a transaction with significant relationship specific investment may work satisfactorily for some period of time. However, when external circumstances change, the existing contractual arrangements can lead to significant adaptation problems and increase the costs of that lead to significant adaptation problems and associated performance inefficiencies. These adaptation problems are less likely to have emerged if the production of car bodies had been governed through

internal organization rather than a rigid contract that did not anticipate a big increase in demand and the conflicts that emerged over plant location decisions. Changes in technology or government regulations may also change the relative attractiveness of alternative governance arrangements. For example, with specific reference to my work on coal contracts, changes in environmental laws in the U.S. have independently made it attractive to invest in fuel-flexibility capabilities in coal burning power plants. This flexibility in turn makes coal users less dependent on specific coal suppliers or coal supply locations, reducing the value of long term contracts. Accordingly, one would anticipate seeing a shift to shorter term more flexible contracts as the net costs of fuel-flexibility to mitigate potential hold-up problems declines as a consequence of environmental regulations that increase the value of such flexibility.

The bottom line is that there are benefits and costs of internal organization. Market transactions incur transactions costs associated with writing and enforcing contingent contracts and the inefficiencies ex ante and ex post resulting from opportunistic behavior that exploits specific investments. Internal bureaucratic allocation mechanisms can help to mitigate these types of transactions costs but incur other types of transactions or organization costs. The costs of internal organization are associated with the relatively inferior adaptive properties of bureaucratic hierarchies to rapidly changing outside opportunities over the longer term and the difficulty of designing compensation mechanisms to give managers and employees appropriate incentives to control costs and product quality. No governance structure is free from at least some transactions costs. The decision whether or not to vertically integrate then becomes a tradeoff between the costs of alternative governance arrangements. Governance arrangements are selected which represent the best that can be accomplished from a set of imperfect governance alternatives. Understanding the

tradeoffs between governance alternatives and how they are affect by the attributes of products, production processes, inputs, legal, political and regulatory institutions is what the comparative governance approach is all about.

Despite these observations, however, I think that it is fair to say that the TCE literature on vertical integration, especially the empirical literature, has focused much more on the inefficiencies of market transactions than it has on the strengths and weaknesses of internal organization. Indeed, this may be one of the reasons why Gibbons (2003) argues that there is a lot of confusion about the similarities and differences between the TCE approach attributed to Williamson, Klein and others and the property rights or rights of control approach attributed to Grossman and Hart (1986) and Hart and Moore (1990). TCE emphasizes (verbally) ex post adaptation issues and the associated bargaining and performance costs, recognizing that these costs also affect ex ante investment incentives. The property rights literature assumes that ex post bargaining is efficient and emphasizes the effects of ex post rent expropriation on ex ante investment. However, both literatures have emphasized specific investments and, as we shall see, much of the empirical literature relates variations in the costs of governance structure to variables measuring asset specificity of various kinds, rather than on direct measures of ex post adaptation costs (which could be argued would be off the equilibrium path anyway), good proxies for their expected magnitude, or variables measuring variations in the costs of internal organization. The full implementation of a comparative governance framework requires that the costs of alternative market governance arrangements and the costs of internal organizations with different attributes be given equal treatment. The situation has not changed all that much since 1971 when Williamson (1971, p. 113) observed that:

> "A complete treatment of vertical integration requires that the limits as well as the powers of internal organization be assessed. As the frictions associated with the powers of administrative coordination become

progressively more severe, recourse to market exchange becomes more attractive, *ceteris paribus*… it is simply asserted [in this essay] that, mainly on account of bounded rationality and greater confidence in the objectivity of market exchange in comparison with bureaucratic process market intermediation is generally to be preferred over internal supply in circumstances in which markets may be said to 'work well" (footnote omitted)

"The properties of the firm that commend internal organization as a market substitute … fall into three categories: incentives, controls, and what may be referred to broadly as 'inherent structural advantages.'  In an incentive sense, internal organization attenuates the aggressive advocacy that epitomizes arm's length bargaining.  Interests, if not perfectly harmonized, are at least free of representations of narrowly opportunistic sort… In circumstances…where protracted bargaining between independent parties to a transaction can be reasonably anticipated, internalization becomes attractive (footnote omitted)."

"… when conflicts develop, the firm possesses a comparatively efficient conflict resolution machinery… fiat is frequently a more efficient way to settle minor conflicts (say differences in interpretation) than is haggling or litigations. "


## 6.  EMPIRICAL EVIDENCE

The choice of governance structure and how this choice is affected by the kinds of transaction cost considerations that have been discussed here have attracted considerable empirical study.  This empirical work has focused on decisions to vertically integrate, the design of non-standard contractual arrangements and the performance of both vertical integration and non-standard contractual arrangements over time as supply and demand conditions change. This work has included both detailed case studies of particular firms or types of contractual and organizational arrangements as well as econometric analyses based on large numbers of observations on the governance arrangements chosen for transactions with different attributes.  Interestingly, the TCE framework has stimulated much more empirical work than either the traditional theories of vertical integration outlined above or 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. It also reflects the continuing interaction between theory, empirical analysis and public policy that has characterized developments in TCE over the last 25 years, a productive interaction that is largely absent from many other areas of industrial organization. Moreover, in the case of TCE related research, the empirical results are much more supportive of the relevant theory than is the case with the other theories of vertical integration (see for example Waterman and Weiss, 1996; Chipty, 2001; Mason and Phillips, 2000). Whinston (2003) argues that the empirical work stimulated by TCE does not do a good job distinguishing between TCE-based theories and property rights based theories of vertical integration and provides suggestions for how the predictions of property rights theories might be tested and its distinct predictions, including those that are contrary to the implications of TCE-based theories, tested empirically. Empirical work to date has not focused on trying to distinguish between TCE and property rights theories of vertical integration and there has been little effort to test property rights theories directly. This is probably the case because those who have led the development of property rights theories of vertical integration have made little effort to specify clear testable hypotheses or to relate them to variables that are likely to be measurable and available for empirical analysis. That is, they have not provided adequate guidance to empirical researchers and, as a result, the similarities and differences between TCE-based theories and property rights theories have been of much more interest to theorists than to empiricists. Perhaps the guidance provided by Whinston will lead empirical researchers to focus more attention on property rights theories.

There have been at least 500 papers published that have examined various aspects of comparative institutional choice from a TCE perspective. A significant fraction of these studies

have examined the vertical integration or "make or buy" decision. There have also been several survey articles that have reviewed the empirical literature stimulated by TCE theories, including many related to vertical integration and non-standard vertical contracting arrangements (Joskow, 1988; Shelanski and Klein, 1995; Crocker and Masten, 1996; Coeurderoy and Quélin, 1997; Vannoni 2002). In addition, Peter Klein has written a chapter in this volume that reviews the empirical work on the "make or buy" decision. Accordingly, my discussion here will be relatively brief and focus on methodological issues rather than providing a complete catalogue of empirical studies and their results.

Masten, Mehan and Snyder (1991) present a very useful empirical model that captures the essence of the comparative governance approach, the associated TCE predictions regarding the choice between governing vertical relationships with market contracts or through vertical integration (internal organization), and the issues raised for empirical analysis. I will use that model here since it helps to organize alternative empirical approaches and to identify important empirical issues. While it is most relevant to econometric studies of the choice between "market" and "organization," it also provides useful guidance for related case study research.

Masten, Mehan and Snyder focus on a model of the choice between market contracting (m) and internal organization --- vertical integration --- (o). Following the comparative governance arrangements approach they define the costs of the two alternative governance arrangements as:

$C_o$ = cost of organizing transactions inside a firm (e.g. VI)

$C_m$ = cost of organizing transactions through a (least cost) market contracting mechanism

Then the choice between market contracting and vertical integration depends on the relative costs of the too alternative governance arrangements. That is, the theory implies that transacting parties will choose internal organization if $C_o < C_m$ and vice versa. The total costs of transacting associated with the alternative governance arrangements then depend on those attributes of the transactions that affect the costs of market contracting (measured by the elements of a vector $Z$) and the costs of vertical integration (measured by the elements of a vector $X$). $X$ and $Z$ may or may not include common elements. Accordingly, the costs of vertical integration or internal organization ($o$) and of market contracting ($m$) can be modeled as:

$$C_o = \alpha X + e \tag{6}$$

$$C_m = \beta Z + u \tag{7}$$

where $\alpha$ and $\beta$ are the coefficients that measure the marginal governance cost associated with each relevant transactional attribute for internal organization and market governance structures respectively and e and u are random disturbance terms which may or may not be correlated with one another. It follows immediately that the probability of choosing vertical integration depends on the probability that the costs of internal organization are less than the costs of market transactions.

Probability of choosing internal organization = $Pr(C_o < C_m) = Pr(e - u < \beta Z - \alpha X)$ (8)

Unfortunately, it is rarely possible to measure the costs of internal organization ($C_o$) and the

costs of market contracting ($C_M$) directly. As a result, we cannot estimate the parameters of the structural model (1) and (2) directly. Nor do we typically have good cardinal measures of the transactional attributes that enter $X$ and $Z$. Instead, as we shall see, researchers frequently must rely on ordinal proxy variables to measure variations in the elements of $X$ and $Z$. So, empirical studies often rely on observations indicating whether or not a relationship is governed by internal organization or market contract, creating a 0,1 limited dependant variable, and various proxies for variations in transaction related variables that are elements of $X$ and/or $Z$, such as asset specificity, complexity, uncertainty, and frequency of transactions or repeated interaction. Hypotheses regarding organization form can then be based on the estimated coefficients of $\alpha$ and $\beta$. These coefficients can be estimated using limited dependent variable techniques like probit or logit (though only up to a proportionality factor without independent information about the variance of ($e$ - $u$)). Moreover, if $X$ and $Z$ share common variables (e.g. asset specificity) then one can test whether $(\beta_k - \alpha_k) > 0$, but not whether both coefficients are non-negative, which weakens the power of the associated hypothesis tests. For example, if one believed that when cost minimizing exchange requires more resources devoted to specific investments the costs of both market contracts <u>and</u> internal organization increased, we would only be able to identify the relative costs (or net costs) of one governance arrangement compared to the other. If it were the case that asset specificity has little effect on market transactions, but reduced the costs of internal organization, then $\alpha_k$ would be negative and $(\beta_k - \alpha_k) > 0$ even though asset specificity has no effect on the costs of transacting through bilateral contracts. These possibilities reduce the power of the reduced form approach to hypothesis testing, though how important this is depends on whether there is good reason to believe that asset specificity affects internal organization costs in any systematic way. I have not been

convinced that there is.        If $X$ and $Z$ do not share common elements (e.g. variations in asset specificity affect the cost of market transactions but do not affect the costs of internal organization directly), we can identify the signs of the relevant coefficients independently.  In many applications, the implicit assumption is that $X$ and $Z$ are orthogonal (e.g. variations in asset specificity affect the costs of market organization but not the costs of internal organization), the focus is on the measurement of the elements of $Z$ ( i.e. when $\beta_k > 0$ then $\alpha_k = 0$ and vice versa) and on the signs and magnitudes the estimated coefficients of the "market failure" variables included in $Z$.   If the assumption that $X$ and $Z$ are orthogonal is not correct, not only can we measure only the sign and magnitude of $(\beta_k - \alpha_k)$, but if there are left out variables in $X$ that are correlated with the included elements of $Z$ it may not even be possible to get unbiased estimated of $(\beta_k - \alpha_k)$.

Masten, Meehan and Snyder go on to show that <u>if</u> we can measure $C_o$ <u>or</u> $C_m$, we can measure the individual structural coefficients (not just their   difference) <u>and</u> the cost of the other organizational form using  switching regression techniques (if $X$ and $Z$ are orthogonal <u>or</u> $e$ and $u$ are uncorrelated). They then apply this technique to measure the costs of governing the production of several components used in naval shipbuilding with varying transactional attributes using either vertical integration or contracting with third parties. I am familiar with only one other study that measures the costs of alternative governance arrangements directly (Kerkvliet 1991).

With this basic empirical model in mind, let us now turn to the methods used in the empirical literature that seeks to test whether variations in transaction attributes such as asset specificity affect the choice between vertical integration and market contracting as TCE theory predicts.   These studies tend to follow a similar empirical methodology.  They generally focus on a particular good or service that is used as an input to produce or distribute a specific class of products:  automobile

components, (Klein, Crawford and Alchian, 1978; Klein, 2000, 2002; Monteverde and Teece, 1982; Walker and Weber, 1984; Langois and Robertson, 1989); coal (Joskow, 1985, 1987, 1988b, 1990; Kerkvliet 1991); aerospace systems (Masten, 1984); aluminum (Stukey, 1983); chemicals (Lieberman, 1991); forestry (Globerman and Schwindt, 1986)); carbonated beverages (Muris, Scheffman and Spiller 1992); pulp and paper (Ohanian, 1994); property-liability insurance (Regan, 1997).  Other studies focus on a set of products that can be distributed through a similar set of alternative distribution modes (Anderson and Schmittlein (1984)).  The sale of these goods and services is mediated by several different governance structures (e.g. vertical integration, franchise agreements, long term contracts, spot market sales) and the governance choices are observable.

The empirical challenge is then to determine whether the incidence of the alternative governance structures observed in practice can be explained by variations in the transactional characteristics of the goods and services whose governance structures are being investigated, in particular by the importance of asset specificity, holding other transactional attributes constant (or assuming that any associated missing variables are uncorrelated with the measures of asset specificity).  That is, to measure the coefficient of $\beta$ and $\alpha$.  In light of the empirical model discussed above this challenge is more daunting than might first meet the eye given the limitations on measuring all of the dependent and independent variables that we would measure under ideal conditions.

To respond to this empirical challenge in the absence of direct measures of the costs of transacting under different governance arrangements, empirical work typically relies instead on information about the actual utilization of alternative governance structures to mediate specific groups of transactions within the (narrow) class of products being studied.  For example, in my work

on coal supply arrangements for electric utilities I identified specific generating stations that relied on coal supplies from affiliated mines as well as generating stations that contracted for coal with unaffiliated suppliers. (My research program involved an examination of the choice between vertical integration and market contracting (1985), the duration of contracts for the bulk of the coal supply relationships that did not involve vertical integration (Joskow, 1987), the adjustment features built into the contracts to allow prices and quantities to adapt over time (Joskow, 1988), and the ex post performance of the contractual arrangements, including the use of the courts to enforce contractual commitments (Joskow, 1990). As another example, Monteverde and Teece identify the proportions of various automobile components utilized by U.S. automobile manufacturers which are produced internally compared to the proportion procured from third parties. In both sets of studies, there is substantial variation across the sample in the reliance on vertical integration. Many other studies take a similar approach.

The next step is to develop measures of the exogenous characteristics of the underlying transactions, with particular attention devoted to measuring the importance of specific investments and other variables that may interact with asset specificity to affect the costs of market contracting and the incidence of opportunistic behavior that could make market contracting more costly in terms of ex ante investment distortions and ex post performance inefficiencies. Measuring variations in the importance of specific investments to support cost-minimizing exchange is difficult. In a number of cases survey data have been used to characterize the importance of specific investments in supporting different groups of transactions within the set that is being studied (Monteverde and Teece, 1982; Anderson and Schmittlein, 1984). In other cases efforts are made to develop ordinal characterizations of the different types of asset specificity associated with specific sets of

transactions within the groups (e.g. Joskow, 1985) focuses on mine-mouth plants as classical examples of site specificity).

Ideally, it would be desirable to identify attributes of the firms in the sample that are expected to affect the costs of internal organization as well. Then the comparative costs of alternative governance arrangements could be captured directly. As a practical matter, most of these studies rely on samples of firms that are reasonably assumed to have identical "internal organizational" attributes. The implicit assumption is that the variables that affect the costs of market contracting are orthogonal to the variables that affect the costs of internal organization. For example, in my work on coal contracts, the firms involved all produced electricity (a homogeneous product), were regulated monopolies and where subject to similar types of economic regulation. I had no reason to believe that variations in the importance of asset specificity affected the costs of internal organization significantly, while they were expected to affect the costs of market contracting significantly. Monteverde and Teece (1982) look at transactions involving a small number of automobile firms with widely varying demands for components. They implicitly assumed that some variables (e.g. firm size) affected the costs of internal organization and others affected the costs of market contracting. That is, the implicit assumption again is that $X$ and $Z$ are orthogonal. If this assumption is not correct it reduces the power of the tests of the TCE theory and may lead to biased estimates of the coefficients being estimated.

With these data in hand, and recognizing the potential problems associated with the measurement issues that I have discussed, the analysis then proceeds to determine whether the variations in the governance structures observed (vertical integration or bilateral contracting, extent of vertical integration, extent of contractual pre-commitment) can be associated with variations in

the importance of specific investments and other characteristics of transactions that the theory suggests will lead to opportunistic behavior and related contractual hazards. As noted earlier, in econometric application, the dependent variable is typically a limited dependent variable that indicates whether vertical integration is utilized (y = 1) or not (y = 0) or the intensity of vertical integration (y = % of purchases by a buyer from internal sources). For the reasons discussed above, and nicely illuminated by Masten, Mehan and Snyder (1991), the results and power of the tests must be interpreted with great care.

In the end, we become convinced about the validity of a theory based on the accumulation of evidence from many studies examining a wide range of transactions and governance arrangements. The accumulation of results provides the most important power of the hypothesis tests. The empirical studies of vertical integration and how the choice of this governance structure is influenced by the importance of specific investment and other variables that could lead to ex ante and ex post contractual inefficiencies overwhelmingly show that the importance of specific investments is both a statistically and economically important causal factor influencing the decision to vertically integrate. Indeed, it is hard to find many other areas in industrial organization where there is such an abundance of empirical work supporting a theory of firm or market structure. And it is the combination of compelling theoretical analysis combined with a large body of supporting evidence that makes the TCE approach to understand vertical integration and alternative vertical governance arrangements so important.

## 7. CONCLUSIONS

Let me conclude where I began. There is no single unified theory of vertical integration that

exists today or is likely to exist in the future. There are many types of market imperfection that could lead transacting parties to turn to vertical integration as an alternative governance arrangement, recognizing that vertical integration is one of many governance alternatives to relying on anonymous spot market contracting. However, the NIE/TCE approach provides a framework that can encompass and enrich all leading theories of vertical integration. It does so by taking a comparative governance approach, recognizing that there are unique but systematic costs associated with alternative market contracting structures, with vertical integration and with various hybrid forms, and by building on a synergistic relationship between theoretical and empirical analysis.

This being said, there is still much to learn about vertical integration, alternative market contracting structures and various hybrid forms. In my view, we have made more progress in understanding and measuring the hazards and associated costs of market contracting in the presence of alternative transactional attributes than we have about the costs of internal organization and how these costs are affected by different internal organizational and incentive structures. This observation is especially relevant to the state of empirical analysis where measurement issues remain a serious challenge. In addition, I believe that it would be very productive to focus more attention on the dynamic properties of both contractual relationships and internal organization. The "make or buy" decision is not a once and for all decision. Firms may choose to vertically integrate and then decide that it is less costly to rely on market contracting. As in the case of GM-Fisher body, firms may choose to govern a relationship by contract and then decide to take production in house. Better understanding how idiosyncratic contractual relationships adapt to changing supply and demand conditions over time, how organizations respond to changing circumstances and ageing over time, and why governance arrangements change over time will provide deeper insights into both the

market imperfection and organization imperfection considerations that affect the choice of governance arrangements.

Paul. L. Joskow

Elizabeth and James Killian Professor of Economics

Massachusetts Institute of Technology

Cambridge, MA, USA

December, 2003

### *References*

Aghion, Philippe and Patrick Bolton. 1987. "Contracts as a Barrier to Entry," *American Economic Review*, 77:388-401.

Anderson, Erin and David Schmittlein. 1984. "Integration of the Sales Force: An Empirical Examination," *Rand Journal of Economics*. 15: 385-95.

Arrow, Kenneth J. 1974. *The Limits of Organization*. New York: W.W. Norton.

Arrow, Kennth, J. 1975. "Vertical Integration and Communication," *The Bell Journal of Economics*, 173-83.

Bain, Joe. 1956. *Barriers to New Competition*. Cambridge, MA: Harvard University Press.

Bain, Joe. 1959. *Industrial Organization*. New York: John Wiley & Sons.

Bolton, Patrick and Michael Whinston. 1993. "Incomplete Contracts, Vertical Integration, and Supply Assurance," *Review of Economic Studies*, 60: 121-48.

Carlton, Dennis. 1979. "Vertical Integration in Competitive Markets Under Uncertainty," J*ournal of Industrial Economics*, 27: 189-209.

Chandler, Alfred. 1964. *Strategy and Structure: Chapters in the History of American Industrial Experience*, MIT Press, Cambridge, MA.

Chipty, Tasneem. 2001. "Vertical Integration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry," *American Economic Review*, 91: 428-53.

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

Coase, Ronald. 1972. "Industrial Organization: A Proposal for Research." In. V.R. Fuchs, ed. *Policy Issues and Research Opportunities in Industrial Organization*. New York: National Bureau

of Economics Research, pp. 59-73.

Coeurderoy, Régis and Bertrand Quélin. 1997. "Transaction Cost Theory: A Survey on Empirical Studies on Vertical Integration*,"* [in French], *Revue Economie Politique*, 107: 146-181.

Crocker, Keith and Scott Masten. 1996. "Regulation and Administered Contracts Revisited: Lessons Fram Transaction-Cost Economics for Public Utility Regulation," *Journal of Regulatory Economics,* 9: 5-40.

Cyert, Richard and James March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice Hall.

Elberfeld, Wallter. 2002. "Market Size and Vertical Integration: Stigler's Hypothesis Reconsidered," *Journal of Industrial Economics*, 50: 23-43.

Gibbons, Robert. 2003, "Four Formal(izable) Theories of the Firm?," preliminary and incomplete draft, August 6, 2003.

Globerman, Steven and Richard Schwindt. 1986. "The Organization of Vertically Integrated Transactions in the Canadian Forest Products Industry," *Journal of Economic Behavior and Organization*. 7:199-212.

Green, Jerry. 1986. "Vertical Integration and Assurance of Markets," in *New Developments in the Analysis of Market Structure*, ed. F.G. Mathewson and J.E. Stiglitz, MIT Press, Cambridge, MA.

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

Hart, Oliver. 1995. *Firms Contracts and Financial Structure*. Oxford: Clarendon Press.

Hart, Oliver and John Moore. 1990. "Property Rights and the Nature of the Firm," *Journal of*

*Political Economy*, 98:1119-58.

Hart, Oliver and Jean Tirole. 1990. "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity*, Special Issue: 205-276.

Holmstrom, Bengt and Paul Milgrom. 1990. "Regulating Trade Among Agents," *Journal of Theoretical and Institutional Economics*, 146: 85-105.

Joskow, Paul L. 1985. "Vertical Integration and Long Term Contracts: The Case of Coal-burning Electric Generating Stations," *Journal of Law, Economics and Organization*. 1:33-80.

Joskow, Paul L. 1987. "Contract Duration and Relationship Specific Investments," *American Economic Review*, 77:168-75.

Joskow, Paul L. 1988a. "Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence," *Journal of Law, Economics and Organization*, 4: 95-117.

Joskow, Paul L. 1988b. "Price Adjustment in Long Term Contracts:  The Case of Coal," *Journal of Law and Economics*. 31:47-83.

Joskow, Paul L. 1990. "Price Adjustment in Long Term Contracts: Further Evidence from Coal Markets," *Rand Journal of Economics*, 21:251-74.

Joskow, Paul L. 1991. "The Role of Transactions Cost Economics in Antitrust and Public Utility Regulatory Policies," *Journal of Law, Economics and Organization*, 7 (Special Issue):497-540.

Joskow, Paul L. 1998. "Asset Specificity and Vertical Integration," *The New Palgrave Dictionary of Economics and Law*, Peter Neuman, Editor, MacMillan, 1998.

Joskow, Paul L. 2002. "Transaction Cost Economics, Antitrust Rules and Remedies," *Journal of Law,  Economics and Organization,*  18:. 95-116

Kerkvliet, Joe. 1991, "Efficiency and Vertical Integration: The Case of Mine-mouth Electric Generating Plants," *The Journal of Industrial Economics* 39: 467-482.

Klein, Benjamin. 1988. "Vertical Integration as Organized Ownership: The Fisher Body-General Motors Relationship Revisited," *Journal of Law, Economics and Organization*, 4:199-213.

Klein, Benjamin. 2000. "Fisher-General Motors and the Nature of the Firm," *Journal of Law and Economics*, 43:105-141.

Klein, Benjamin, 2002. "Fisher Body-General Motors Once Again: How do We Know When A Holdup Occurs?" mimeo.

Klein, Benjamin, Robert Crawford and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21:297-326.

Klein, Benjamin and Kieth Leffler. 1981. "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy* 89: 615-41.

Langlois, Richard and Paul L. Robertson, 1989. "Explaining Vertical Integration: Lessons from the American Automobile Industry," *The Journal of Economic History*, 69: 361-375.

Levy, D.T. 1984, "Testing Stigler's Interpretation of 'Division of Labor is Limited by the Extent of the Market'," *Journal of Industrial Economics,* 32: 377-389.

Lieberman, Marvin. 1991. "Determinants of Vertical Integration: An Empirical Test," *Journal of Industrial Economics*. 39:451-66.

Malmgren, H.B. 1961. "Information, Expectations and the Theory of the Firm," *Quarterly Joutnal of Economics*, 75: 399-421.

Maskin, Eric and Jean Tirole. 1999. "Two Remarks on the Property-Rights Literature," *Review of Economic Studies*, 66: 139-149.

Masten, Scott. 1984. "The Organization of Production:  Evidence From the Aerospace Industry," *Journal of Law and Economics.*  27:403-17.

Masten, Scott, James Meehan and Keith Crocker. 1989. "Vertical Integration in the U.S. Auto Industry: A Note on Specific Assets," *Journal of Economic Behavior and Organization*, 12:265-73.

Masten, Scott, James Meehan and Edward Snyder. 1991. " The Costs or Organization," *Journal of Law, Economics and Organization*, 7: 1-25.

Mason, Charles F and Owen R. Phullips. 2000. "Vertical Integration and Collusive Incentives: An Experimental Analysis," *International Journal of Industrial Organization*, 18:471-93.

Mathewson, Frank and Ralph Winter. 1986. **"**The Economics of Vertical Restraints in Distribution," In *New Developments in the Analysis of Market Structures,* ed. Frank Mathewson and Joseph Stiglitz, Cambridge, MA. MIT Press.

Monteverde Kirk and David Teece. 1982. "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *Bell Journal of Economics*, 13:206-213.

Muris, Timothy, J. David Scheffman, and Pablo Spiller. 1992. "Strategy and Transactions Costs: The Organization of Distribution in the Carbonated Soft Drink Industry," *Journal of Economics and Management Strategy*.  1:83-128.

Myerson, Roger and M. Satterthwaite. 1983. "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*,  28:265-281.

Ohanian, Nancy Kane. 1994. "Vertical Integration in the U.S. Pulp and Paper Industry, 1900-1940," *Review of Economics and Statistics*, 76: 202-207.

Ordover, Janusz, Steven Salop and Garth Saloner. 1990. "Equilibrium Vertical Foreclosure,"

*American Economic Review.* 80:127-42.

Perry, Martin. 1978. "Price Discrimination and Vertical Integration," *Bell Journal of Economics*, 9, Spring, 209-17.

Regan, Laureen. 1997. "Vertical Integration in the Property-Liability Insurance Industry: A Transaction Cost Approach," *Journal of Risk and Insurance*, 64: 41-62.

Shelanski, Howard and Peter Klein. 1995. "Empirical Research in Transaction Cost Economics: A Review and Assessment," *Journal of Law, Economics and Organization,* 11:335-361.

Simon, Herbert A. 1947. *Administrative Behavior*. New York: Macmillan.

Stigler, George. 1951. "The Division of Labor is Limited by the Extent of the Market," *Journal of Political Economy*, 59: 185-193.

Stukey, John. 1983. *Vertical Integration and Joint Ventures in the Aluminum Industry*. Cambridge, MA: Harvard University Press.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Vannoni, Davide. 2002., "Empirical Studies of Vertical Integration: The Transaction Cost Othodoxy," *International Review of Economics and Business,* 49:113-141.

Vita, Michael. 2000. "Regulatory Restrictions on Vertical Integration and Control: The Competitive Impact of Gasoline Divorcement Policies," *Journal of Regulatory Economics*, 18: 217-33.

Walker, Gordon and David Weber. 1984. "A Transactions Cost Approach to Make or Buy Decisions," *Administrative Science Quarterly.* 29:373-91.

Waterman, David and Andrew Weiss. 1996. "The Effects of Vertical Integration Between Cable Television Systems and Pay Cable Networks," *Journal of Econometrics,* 72:357-395.

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

Williamson, Oliver. 1971. "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 61: 112-123.

Williamson, Oliver. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

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

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

Williamson, Oliver. 1996. *The Mechanisms of Governance*. New York: Oxford University Press.

Williamson, Oliver. 2000. "The New Institutional Economics: Taking Stock, Looking Ahead," *Journal of Economic Literature* 38: 595-613.

Williamson, Oliver, Michael Wachter and Jeffrey Harris. 1975. "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange," *Bell Journal of Economics,* 6: 250-280.