VERTICAL INTEGRATION

Paul L. Joskow*

Abstract: This chapter reviews alternative economic theories of vertical integration and the empirical literature that examines the power of alternative theories to explain the incidence of vertical integration in practice. Neoclassical theories of vertical integration are discussed first. These theories turn on efforts of firms either to mitigate inefficiencies caused by market power at one or more levels of the vertical chain or to create or enhance market power at one of both levels. Organizational theories of vertical integration that focus on incomplete contracts, asset specificity, information imperfections, opportunistic and the costs of internal organization are discussed next. These theories focus on efforts by firms to mitigate transactions costs and various contractual hazards that may arise with bilateral contracts by choosing among alternative organizational and contractual governance arrangements that can reduce these costs. The chapter concludes with a discussion of the empirical literature that has examined theories of vertical integration that fall in both categories.

1. Introduction

Understanding the factors that determine which types of transactions are mediated through markets and which within firms through vertical integration has been an important subject of theoretical and empirical research in microeconomics for at least the last 25 years. Moreover, vertical integration and related vertical no-standard contractual arrangements (so-called “vertical restraints”) have historically attracted considerable attention under U.S. and E.U. antitrust laws. Surprisingly, however, most intermediate microeconomics textbooks pay little if any attention to the causes and consequences of vertical integration between suppliers of intermediate goods and

services (“upstream”) and the purchasers of those goods and services (“downstream”). This chapter reviews theoretical and empirical work in microeconomics that examines the causes and consequences of vertical integration and identifies relationships between vertical integration and substitute non-standard vertical contractual arrangements.

I want to emphasize at the outset that there is not and will never be one unified theory of vertical integration. Moreover, while some of the literature on vertical integration focuses on a simple dichotomy between the decision to “make” internally or “buy” through the market, in reality these two governance arrangements are polar cases. A comprehensive analysis of the underlying causes and consequences of vertical integration should not only examine the determinants of the boundaries between firms and markets but also the origins of various “non-standard” contractual arrangements or “hybrid forms” that lie between simple anonymous spot market transactions and internal organization. These hybrid forms include various types of long term contracts, franchise contracts, non-linear pricing arrangements, resale price maintenance agreements, requirements contracts, joint ventures, dual sourcing (partial vertical integration) and others. Depending on the circumstances, these alternative contractual arrangements may be perfect or imperfect substitutes for vertical integration for dealing with problems that may arise by relying only on simple repeated spot market relationships between upstream and downstream firms.

Virtually all theories of vertical integration turn in one way or another on the presence of market imperfections of some type. That is, deviations from the long list of explicit and implicit assumptions that are associated with textbook models of perfect competition and anonymous spot market transactions under perfect competition. Neoclassical approaches to vertical integration have tended to focus primarily (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, organizational economics based theories (transactions cost economics and related property rights based theories) of vertical integration start with the recognition that firms seeking to consummate transactions must confront a variety of potential transaction costs, contractual, and organizational hazards, which are related to the attributes of the transactions at issue. 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 may arise as a consequence of contractual hazards associated with transactions mediated through market arrangements and bureaucratic costs associated with internal organization. The transaction cost economics framework recognizes that vertical integration has both costs and benefits from the firms’ perspectives. The governance structures that are chosen, whether vertical integration, non-standard vertical contractual arrangements, or simple spot market transactions reflect efforts by the firms involved to reduce inefficiencies that might otherwise be associated with both ex ante investment decisions and the ex post performance of a trading relationship.

The first section of this chapter reviews what I call neoclassical theories of vertical integration. This section includes discussions of the incentives to vertically integrate and the consequences of vertical integration associated with vertical externalities, price discrimination, horizontal externalities, vertical foreclosure, and other theories. I then proceed to review organizational economics theories of vertical integration based on incomplete contract, transactions cost, asset specificity and property rights or rights of control considerations. The final substantive section discusses the empirical work that has examined the various neoclassical and transaction cost
economics theories of vertical integration and related non-standard vertical contractual arrangements.

2. Neoclassical 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 that takes 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 responses to market power that existed in upstream or downstream markets (or both) and/or reflected efforts to create or exploit market power. This is the fundamental basis for neoclassical theories of vertical integration.

*Vertical externalities:* A classical explanation for vertical integration is as a response to inefficiencies that arise when there is market power in both the upstream and downstream markets.\(^1\) This in turn implies that market prices will be greater than the marginal cost of production in both upstream and downstream markets as firms exercise market power. The polar case is one where there is a pure monopoly upstream ("manufacturing") and another pure monopoly downstream ("retailing") and where the upstream monopoly has all of the bargaining power over the price that will be charged for goods or services sold by the upstream firm to the downstream firm. The monopoly at the upstream “manufacturing” level has a marginal production cost \(c\) and seeks to charge a monopoly input price \(P_M > c\) to the downstream firm. The downstream "retail" monopoly takes the price \(P_M\) as its input cost and exercises its monopoly power by charging a retail price \(P_{RM} > P_M\). The upstream monopoly has added a monopoly markup to its production costs and then the downstream monopoly has added another markup to the price it pays to the upstream firm for its

\(^1\) Market power as that term is used in economic theory. Firms face downward sloping demand curves and are not pure price takers. There need not be supra-normal profits in equilibrium, however.
inputs. This phenomenon is known as double-marginalization. When making its pricing decisions, the downstream monopoly ignores (because it does not see) the actual costs of production incurred by the upstream firm. This behavior of the independent monopolies results in aggregate profits that are smaller than they would be in the firms set prices so as to maximize their joint profits. It also leads to higher prices downstream than would maximize the joint profits if the upstream and downstream firms. Tirole refers to the failure of the downstream firm to take upstream production costs into account as a vertical externality (Tirole, 1988, Chapter 4).

If we assume that vertical integration is costless, the aggregate profits of the two monopolies will increase if they merge since the distortion from double marginalization will disappear. The integrated firm will set the profit maximizing downstream monopoly price properly taking into account the actual costs of production at the upstream level. Moreover, prices charged to consumers will fall and, as a result, consumers are made better off by vertical integration. This is the classic example of the maxim that a single monopoly is better than a chain of monopolies.

This approach treats vertical integration itself as being “costless” compared to alternative institutional arrangements. That is, no internal organization costs are recognized, but only any costs created by distortions in market prices, quantities, or the factor proportions used to produce output from a neoclassical production function. Accordingly, it identifies the potential benefits of vertical integration to the firms, but not any associated costs that firms may bear when they enter additional lines of business and expand the size and scope of activities undertaken by the firm. “Costless” vertical integration here is effectively being 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” as well as with
alternative (assumed costless) contractual arrangements --- simple linear transfer prices, two-part
tariffs, maximum retail price maintenance, quantity forcing contracts, requirements contracts, etc.
(Tirole, 1988, Chapter 4). Thus, vertical integration and a menu of non-standard vertical contractual
arrangements are all substitute mechanisms for solving the same double marginalization problem.
Absent transactions costs of some type, the alternative instruments are all equally attractive
mechanisms for responding to double marginalization. In situations where double marginalization is
a problem, however, it may be that relying on two-part tariffs or retail price maintenance or some
other non-standard contractual arrangement may give superior results once organizational and
contracting costs are taken into account.

Note that the double marginalization problem does not arise here if the downstream market
is perfectly competitive. In that case the downstream retail firms are price takers and simply pass
along the manufacturer’s monopoly price $P_M$. There is now only one monopoly distortion.
However, introducing uncertainty about the downstream firms’ costs and demand by the upstream
firm and risk aversion by downstream firms has the effect of making non-standard vertical
contractual arrangements more attractive than simple (linear) monopoly pricing and, ignoring
organization costs, makes vertical integration potentially profitable as well (Rey and Tirole 1986).

Regardless of the competitive structure of the downstream market, if the downstream firms
have production functions that allow for some substitution between an input they purchase from an
upstream monopoly and an input they can buy in a competitive market, then by charging a price
above its marginal cost an upstream monopoly will induce the downstream firm inefficiently to
substitute away from the monopoly priced input and use more of the competitively priced (Vernon
restore an efficient utilization of inputs used in the production of the downstream good or service and increase aggregate profits for the firms. However, a tying contract that ties sales of the competitively supplied input and the monopoly supplied input together in a way that reflects the inputs’ relatively marginal production costs can have the same effect. If the downstream market is characterized by monopoly, absent vertical integration resale price maintenance (RPM) specifying a maximum retail price would have to be added as well to the tying contract in order to mitigate the downstream price (as well as input utilization) effects of double marginalization. In that case, a franchise fee or two part tariff \( T = A + cq \) structured so that the marginal price is the manufacturer’s marginal cost of production will also restore the right input utilization incentives and correct the downstream price distortion resulting from double marginalization. The overall welfare effects of vertical integration or these substitute non-standard vertical contracts when there is market power upstream and input substitution possibilities downstream are positive although the effect on consumer prices is ambiguous when the downstream market is competitive (Schmalensee 1973, Warren-Boulton 1974).

Of course monopoly upstream and downstream or monopoly upstream and perfect competition downstream are both extreme cases. More realistic cases will involve imperfect competition both upstream and downstream. The double marginal problem will still arise if firms at both levels of the production chain have some market power as they will mark up their prices above their actual or perceived marginal cost in qualitatively (though not quantitatively) the same way as in the standard chain of monopolies case. The opportunity to eliminate or reduce this double marginalization problem still creates incentive for vertical integration or reliance on the types of alternative non-standard vertical contractual arrangements that I have discussed. However, when
there is imperfect competition upstream and downstream rather than pure monopolies at both levels ex ante, vertical integration may also have effects on the intensity of competition upstream and or downstream and associated price-cost margins and profits. This will in turn affect the incentives to vertically integrate, the distribution of profits between firms resulting from vertical integration, and consumer prices. Moreover, the social welfare effects of vertical integration are now more likely to be ambiguous, depending on the assumptions made about the nature of competition upstream and downstream and how the intensity of competition is affected by vertical integration (Kuhn and Vives 1999, Riordan 1998).

Introducing service and product quality dimensions further enriches the analysis. Let us return to the chain of monopolies case but assume that the downstream firm can provide service that is valuable to retail consumers. The more service the retailer provides the greater is the demand for the upstream manufacturer’s product. However, the downstream retailer incurs costs to provide this service. In making the decision about how much service to provide the retailer will compare the incremental cost of service against the retail margin \((P_{RM} - P_M)\) earned from each unit of additional sales. However, this underestimates the true aggregate incremental profit resulting from incremental sales for the entire vertical chain by \((P_M - c)\). It also underestimates the incremental consumer welfare from increased service. Vertical integration would internalize this externality, reduce retail prices, and lead to greater expenditures on retail service. Here, neither franchise fees nor maximum resale price maintenance (RPM) are perfect substitutes for vertical integration since they cannot remedy the externality associated with the mispricing of the incremental value of expenditures on retail service. A quantity forcing contract in which the retailer must sell a minimum quantity of the manufacturer’s goods could effectively force it to promote the product efficiently (Tirole 1988,
Chapter 4).

This chain of monopolies model with downstream service provision by a retail monopoly can be extended to the case where there are identical (perfectly) competing downstream firms selling a monopoly manufacturer’s goods and where the provision of retail service by each downstream firm expands the demand for the manufacturer’s product. This is a case of *intra-brand competition*. Similar distortions in expenditures on retail service are realized in the unintegrated situation as discussed above. The vertically integrated solution fully internalizes the retail service expenditure and demand effects. However, as is well known, a downstream monopoly (the result of vertical integration here) the makes choices about both prices and service quality will consider marginal consumer surplus while the competitive (and efficient) result maximizes average consumer surplus (Spence 1970). Vertical integration may be profitable but the welfare consequences are therefore ambiguous because of the replacement of competition with monopoly in the downstream market (Tirole 1988, p. 182). The analysis can be enhanced further by allowing the downstream market to be monopolistically competitive rather than perfectly competitive. The manufacturer now may seek to vertically integrate to control retail prices, the provision of retail services *and* the number of outlets (or product variety). The welfare effects of vertical integration are now even more ambiguous. Moreover, none of the non-standard vertical contractual arrangements (e.g. franchise fees, RPM, etc.) individually is sufficient to restore a fully efficient allocation even assuming perfect information, though combinations of these instruments may be (Mathewson and Winter 1986, Winter 1993). These models also raise horizontal externality issues that I will turn to presently.

*Price discrimination:* Opportunities to engage in price discrimination in the sale of an intermediate good or service to downstream firms in different industries arises when the elasticity of
the derived demand for the intermediate goods varies from one industry to the other. Differences in the elasticity of derived demand creates the opportunity for the upstream monopoly profitably to engage in third-degree price discrimination by charging a higher price to firms in the downstream industry with the less elastic derived demand and a lower price to downstream firms in the industry with the more elastic derived demand. However, the upstream monopoly must confront the challenge of blocking the firms in the industry that are paying lower prices from profitably reselling the intermediate good to firms in the downstream industry that are being charged higher prices, effectively defeating the third-degree price discrimination strategy.

Blocking resale is always a problem faced by a firm that seeks to engage in third degree price discrimination. One way effectively to block resale is for the upstream monopoly to vertically integrate forward into the industry with the more elastic derived demand for the intermediate good (Perry 1978). The vertically integrated firm then only sells the intermediate good to “external” buyers at the higher profit maximizing price that reflects the lower demand elasticity in the other industry. In this case, the vertically integrated firm effectively charges itself a lower price (effectively because it knows its own marginal cost of producing the intermediate good and sets the profit maximizing price for the product it now sells in the downstream industry taking account of that industry’s (higher) demand elasticity) for the downstream output it now produces as a vertically integrated firm than it charges to other “external” buyers.

Unintegrated downstream firms competing in the industry into which the upstream monopoly has integrated forward now can only buy the intermediate good at the higher “external” price designed to capture monopoly rents from firms in the other industry with the smaller demand elasticity. Unless these firms are otherwise more efficient as producers of products in the
downstream market into which the upstream firm has integrated forward, the consequence of this pricing strategy will be that the incumbent downstream firms in this market will be unable to compete with the vertically integrated firm because their costs will be too high due to the high price they must now pay for the intermediate good. This is a classical \textit{price squeeze}. Despite the fact that the upstream firm has extended its monopoly into one of the downstream markets, downstream prices may fall compared to what they would have been if the upstream firm had stayed independent and charged a uniform monopoly price to downstream firms in both industries. Similarly, the downstream price in the low demand elasticity industry into which the upstream firm has not integrated will rise. As with third degree price discrimination generally, the welfare effects of this price discrimination strategy are ambiguous.

Of course, the upstream monopoly could accomplish the same price discrimination result if it could sign contracts that could credibly restrict resale. With such contracts there would also be no price squeeze as the firms in each downstream industry would now face the same industry-specific price for the input sold by the upstream monopoly.

\textit{Horizontal Externalities:} Another potential source of incentives for vertical integration is the \textit{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). Here upstream firms manufacture branded or differentiated products so that they face a downward sloping demand for each product. Moreover, the demand for the upstream firm’s product is affected by downstream retail sales and service activity and associated expenditures that are made by downstream retailers of their products. If retailers cannot fully appropriate for themselves the benefits of retail service expenditures but instead see some of the benefits accrue to their downstream retail competitors, this
“horizontal externality” (Tirole, 1998, Chapter 4) will lead downstream retailers to under-invest in retail service from the perspective of the manufacturer.

Vertical integration is one potential solution to this problem. It would allow the upstream manufacturer to internalize the value of expenditures on sales and service at the retail/downstream level. So too are various combinations of exclusive territorial agreements, minimum resale price maintenance, profit pass-over contracts and other vertical contractual mechanisms. As usual, absent the consideration of the transactional and organizational costs, the unanswered question is how to choose among the alternative institutional arrangements in a systematic way.

Vertical Foreclosure: Vertical integration (and long term vertical contracts) can 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 (Salop and Scheffman, 1983; Aghion and Bolton, 1987; Ordover, Salop and Saloner, 1990; Hart and Tirole, 1990; Riordan 1998). Here it is important to distinguish between a naive view of “market foreclosure” that is sometimes associated with vertical integration and the issues that arise as a result of the strategic use of vertical integration to soften competition to raise prices in the upstream market, the downstream market, or both. Whenever a firm vertically integrates and self-supplies itself with some input, other potential suppliers are in some sense “foreclosed” from providing those input supplies to the vertically integrated firm. By this definition all vertical integration “forecloses competition.” This is not a useful or sensible notion of anticompetitive vertical foreclosure.

The classic case of potentially anticompetitive vertical foreclosure arises when there is a monopoly over the supply of some “essential facility” or “bottleneck resource” input that competing or potentially competing firms need access to at comparable terms and conditions in order to
compete in a downstream market. A high voltage electric power transmission network is an example of an essentially facility in this sense because electricity generating firms must have access to the transmission network to produce and sell their output efficiently. A firm that controls the essential facility and also competes in the downstream market may find it profitable to deny access to that facility or charge a high price to third parties seeking to use the facility in order to sustain its monopoly in the downstream market. The classic response to this argument is that there is “only one monopoly profit” to be had here and that by charging a monopoly price for access to the essential facility the firm that controls it can extract all of the monopoly rents that are potentially available. It does not need to restrict access to the facility nor can it extend its monopoly into a downstream market in a way that further disadvantages consumers. This argument breaks down in a number of cases. When the price for access to the essential facility is regulated, the firm that controls it may find it attractive to restrict access to it in order to restrict entry into unregulated markets in which the owner of the essential facility is also a competing supplier (Beard, Kaserman and Mayo 2001). Vertical foreclosure arguments often arise as regulated vertically integrated monopolies are subject to public policies that open up opportunities for competitors to enter one or more lines of business served by the vertically integrated firm. So, for example, the development of competitive wholesale markets for electric power requires competing generators and their customers to have access to an electric transmission network which has natural monopoly characteristics (Joskow 1997). A firm that both controls the transmission network and is also a competitor in the power market that relies on this transmission network may have the incentive and ability to use the terms and conditions of access to that network to reduce competition in the competitive power market. Continuing price regulation, and in particular cross subsidies that attract inefficient
competitive entry, creates additional complexities regarding the incentives to enter a market and the consequences of foreclosure strategies and I will not pursue those issues here.

Potential opportunities for anticompetitive use of vertical integration or vertical restraints do not require that there be a monopoly either upstream or downstream or an essential facility or input to which downstream suppliers require access. Ordover, Salop and Saloner (1990) analyze a model where there is a duopoly made up of two identical firms that produce a *homogeneous* input upstream and engage in Bertrand competition (effectively perfect competition here) and a duopoly downstream where the firms sell *differentiated* products and also compete in prices (Bertrand competition). Absent vertical integration the upstream firms sell inputs to the downstream firms at a price equal to their marginal production cost. The downstream firms take this input price into account and maximize profits given the demand elasticities they face for the products they each produce, yielding a classic Bertrand equilibrium with differentiated products downstream. The assumption of Bertrand competition downstream means that prices are strategic complements and will be at a level above marginal cost in equilibrium. If one firm selling a differentiated product raises its price then the competing firm will also find it profitable to raise its price. Each firm’s price in turn depends on the price it pays for inputs. If one downstream firm can somehow induce the input price paid by the other downstream firm to rise, while its own input costs do not rise, downstream prices will rise in equilibrium and the profits of the firm that induces it’s rivals costs to rise will rise as well. This is the basic mechanism through which a “raising rivals’ cost” strategy operates.

How could one downstream firm induce the input prices paid by its competing downstream firm to rise without increasing its own input costs? If one downstream firm vertically integrates with
one upstream input supplier and can commit not to sell inputs to the other downstream firm, the remaining independent upstream input supplier now becomes the monopoly supplier to the remaining downstream firm. Accordingly, it will have the ability to raise the price it charges for the inputs it sells to the remaining unintegrated downstream firm. The unintegrated downstream firm will respond to higher input prices by raising its own prices and the other (now vertically integrated) downstream firm will respond by raising its prices in response. Downstream prices rise as do the profits of the vertically integrated firm and the unintegrated upstream firm. This strategy cannot be sustained if the vertically integrated firm cannot commit to withhold input supplies from the market or if it is profitable for the remaining independent upstream and downstream firms to merge as well, effectively recreating the original duopoly situation. Riordan (1998) examines a situation where there is a downstream dominant firm plus a competitive fringe and shows how backward vertical integration can lead to higher prices upstream and downstream as well.

**Other Theories:** 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). “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). However, 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.

George Stigler (1951) proposed a theory of vertical integration based upon Adam Smith’s famous maxim 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. 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.

Acemoglu, Johnson and Mitton (2005) propose a theory in which imperfections in capital markets, regulation, and/or contracting costs create incentives for vertical integration though its ability to overcome these market imperfections by moving transactions from imperfect markets into less costly hierarchical allocation structures. This theory provides a bridge between the traditional neoclassical theories of vertical integration and substitute vertical contractual restraints that arise either as responses to market power or efforts to create or enhance market power and those theories that focus more on the comparative costs of internal organization and an array of market-based contractual alternatives to internal organization.

There is clearly no shortage of neoclassical theories identifying potential incentives for and
consequences of 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 that are alternative “hybrid forms” to either simple linear spot market contracts or vertical integration.

3. Incomplete Contract and Related Transaction Cost Theories

The economic research that looks at vertical integration from a broader organizational cost/benefit perspective falls into two interrelated literatures: Transaction Cost Economics (TCE) theories generally identified with Oliver Williamson (1975, 1983, 1985) and property rights or control rights theories generally identified with Oliver Hart and his co-authors (Grossman and Hart 1986, Hart 1995, Hart and Moore 1990). The foundation of both transaction cost economics theories and control rights theories of vertical integration is the recognition that contracts are incomplete and that contractual incompleteness potentially leads to contractual hazards that adversely affect ex ante investment incentives and the efficiency of ex post performance. 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 vertical integration.

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 or sellers. In this situation one or both parties to the contractual relationship may have the incentive and ability to behave “opportunistically” to serve their own interests – e.g. seeking to extract a larger share of the quasi rents resulting from the continuing relationship. 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 vertical integration in this case is that internal organizational allocation mechanisms 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….”

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 turn can lead to adaptation problems that adversely affect \textit{ex ante} investment incentives and the \textit{ex post} efficiency of the trading relationship. As this literature has developed, \textit{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 \textit{ex post} bargaining and contractual performance problems. 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 \textit{ex post} quasi rents (the difference in asset values between the intended use in the relationship and the next best use if the relationship is terminated --- 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.

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. \textit{site specificity}: The buyer and the seller are in a "cheek-by-jowl" relationship with one another, reflecting \textit{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 bauxite processing plant and the associated mines (Stuckey, 1983) are examples of site specificity.

\(^2\)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.
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 anticipated *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.

Vertical integration is favored when the benefits of mitigating opportunism problems that may arise as a consequence of specific investments are greater than the costs of other sources of static and dynamic inefficiency that may be associated with resource allocation within bureaucratic organizations that may emerge as a consequence of vertical integration.

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 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 within organizations are also likely to become more significant as monitoring becomes more difficult in large organizations.

Accordingly, the organizational economics literature adopts the perspective that there are benefits and costs of vertical (or horizontal or lateral) integration. Market transactions incur transactions costs associated with writing and enforcing contingent contracts and the inefficiencies \textit{ex ante} and \textit{ex post} resulting from opportunistic behavior that exploits specific investments. Internal bureaucratic allocation mechanisms that may be used as a consequence of vertical integration 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. As Williamson (1971, p. 113) observes:

“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, \textit{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. “

4. Empirical Evidence

4.1 Neoclassical Theories

There is surprisingly little comprehensive empirical analysis that has examined what I have referred to as neoclassical theories of vertical integration and related vertical contractual mechanisms. There are many anecdotes but little systematic empirical analysis. On the other hand, there is an extensive empirical literature that has focused on TCE related theories of vertical integration. In all cases the challenge is to use the theoretical work to formulate testable hypotheses and then to find data that are suitable to test these hypotheses. Developing suitable data sets is often especially challenging.

Most of the empirical literature that examines neoclassical theories of vertical integration focuses on whether vertical integration leads to higher or lower prices, on whether there is evidence of exclusion of competing suppliers from the market, and/or on whether there is evidence of changes in consumer welfare resulting from, for example, changes in product variety. The empirical work effectively reduced these theories into “efficiency” theories of vertical integration that have
consequences that are "good" for consumers and anticompetitive “foreclosure” theories of vertical integration that have consequences that are "bad" for consumers.

Chipty (2001) examines the effects of vertical integration between programming and distribution in the cable television industry. Cable television is an industry in which there are likely to be both vertical externality issues (market power upstream and downstream) that may motivate vertical integration and opportunities for exclusionary behavior by cable distribution networks which often have local distribution monopolies. Chipty finds that vertically integrated cable distributors are more likely to exclude rival cable programming networks and favor their affiliated networks than are unintegrated distributors. Overall, however, consumers are not harmed by this behavior since the resulting changes in prices and product variety appear either not to harm or to benefit consumers overall. Waterman and Wise (1996) examine the same issues for cable television and find extensive evidence of exclusionary behavior by vertically integrated firms. They find little evidence of any downstream price effects but do find that sales (penetration) are higher for affiliated programming services carried by vertically integrated firms than for unintegrated firms. This suggests that vertical integration leads to increased downstream sales effort associated with distributor-owned programming services. Waterman and Wise conclude that the results are consistent with either a foreclosure theory or an efficiency theory of vertical integration.

Vita (2000) examines the effects of government regulations that restrict vertical integration between gasoline refiners (upstream) and gasoline retailers (downstream). He recognizes that vertical integration could soften competition (foreclosure theory) and lead to higher retail prices or that it could be motivated by efficiency considerations (e.g. double marginalization) in which case

---

3 Though the expansion of direct broadcast satellite service and the entry of wireline telephone companies into video distribution have introduced an important element of competition in recent years.
vertical integration would lead to lower prices, other things equal. He finds that so-called “divorcement” regulations that restrict vertical integration lead to higher retail gasoline prices --- an average increase of 2.6% per gallon. This result is consistent with an efficiency theory of vertical integration. On the other hand, Gilbert and Hastings (2001) examine the effects of vertical integration between gasoline retailers and refiners in the Western U.S. They find that vertical integration leads to higher wholesale market prices, other things equal. They argue that this is consistent with an anticompetitive raising rivals’ cost theory of vertical integration. However, they do not examine effects on retail prices directly. Rosengren and Meehan (1994) use an event study approach that examines the effects on the equity stock prices of unintegrated rivals at the time vertical mergers are announced for a sample of vertical mergers. If the vertical mergers reflect an effort to soften competition (foreclosure) then they expect to find positive abnormal returns for the stocks of rival unintegrated firms. They find no evidence to support the foreclosure theory.

Experimental economics techniques have also been used to test foreclosure theories of vertical integration. Martin, Norman and Snyder (2001) find evidence that vertical integration increases the ability of the upstream firm to withhold output and increase prices consistent with some of the theoretical analysis in Ordover, Salop and Saloner. However, Mason and Philips (2000) find the vertical integration leads to an expansion of output and increased consumer welfare in a similar experimental setting.

Turning finally to the other theories noted in Section 2, there is abundant support in the business history literature for Carlton’s theory that supply security considerations provide a motivation for vertical integration (Chandler, 1964, p. 84). The empirical prediction of Stigler’s theory is that as industries grow the extent of vertical integration should decline and as industries
contract vertical integration should increase. The theory has found some limited empirical support (Levy 1984). Acemoglu, Johnson and Mitton (2005) find evidence that in countries with less developed capital market institutions, industries that are more human capital or technology intensive are more likely to be vertically integrated.

4.2 Organizational and Transaction Cost Economics Theories

The TCE framework has stimulated much more empirical work than either the neoclassical theories of vertical integration outlined above or than the related property rights or control rights literature. 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. Moreover, in the case of TCE related research, the empirical results are much more consistently supportive of the relevant theory than is the case with the neoclassical and other theories of vertical integration that I have just discussed.

There have been at least 500 published papers 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, 1988a; Shelanski and Klein, 1995; Crocker and Masten, 1996; Coeurderoy and Quélin, 1997;
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; Fan 2000); 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 Schmittein 1984, Affuso 2002)). 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 analysis examines whether the incidence of vertical integration or substitute non-standard vertical contractual arrangements 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). The overwhelmingly conclusion of this large number of empirical studies is that the importance of specific investments and other attributes that affect transaction costs are both statistically and economically important causal factors 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.

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. Whinston (2003) argues that the empirical work in this area 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.

5. 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 to anonymous spot market transactions, recognizing that vertical integration is one of many governance alternatives to relying on anonymous spot market contracting. Some of the theoretical work supports an efficiency motivation for vertical integration. Some of the theoretical work supports an anticompetitive foreclosure motivation for vertical integration. Overall, I would argue that there is substantial support in the empirical literature for various efficiency motivations for vertical integration. There is minimal empirical support for anticompetitive foreclosure motivations. This suggests that there is little empirical support for the antitrust law’s traditional suspicion of and hostility toward vertical integration and related non-standard vertical contractual arrangements (Joskow 2002) except under extreme conditions where firms controlling bottleneck monopoly facilities have the incentive and ability to exercise an anticompetitive foreclosure strategy.
References


