

A PRICE THEORY OF VERTICAL AND LATERAL INTEGRATION*

PATRICK LEGROS[†] AND ANDREW F. NEWMAN[‡]

March 2009; revised March 2013

This article presents a perfectly-competitive model of firm boundary decisions and study their interplay with product demand, technology, and welfare. Integration is privately costly but is effective at coordinating production decisions; nonintegration is less costly, but coordinates relatively poorly. Output price influences the choice of ownership structure: integration increases with the price level. At the same time, ownership affects output, since integration is more productive than nonintegration. For a generic set of demand functions, equilibrium delivers heterogeneity of ownership and performance among ex-ante identical enterprises. The price mechanism transmutes demand shifts into industry-wide re-organizations and generates external effects from technological shocks: productivity changes in some firms may induce ownership changes in others. If the enterprise managers have full title to its revenues, market equilibrium ownership structures are second-best efficient. When managers have less than full revenue claims, equilibrium can be inefficient, with too little integration.

Keywords: Integration, incomplete contracting, market price, organizational welfare loss.

JEL: D21, D23, D41, L11, L14, L22 .

*We had useful comments from Roland Benabou, Patrick Bolton, Phil Bond, Estelle Cautillon, Paola Conconi, Jay Pil Choi, Mathias Dewatripont, Matthew Ellman, Robert Gibbons, Ricard Gil, Oliver Hart, Bengt Holmström, Kai-Uwe Kühn, Giovanni Maggi, Armin Schmutzler, Ilya Segal, George Symeonidis, Michael Whinston, four anonymous referees and the editors of this journal, and workshop participants at Boston University, Washington University, ESSET Gerzensee, the Harvard/MIT Organizational Economics Seminar, the NBER ITO Working Group, and Toulouse. Legros benefited from the financial support of the Communauté Française de Belgique (projects ARC 98/03-221 and ARC00/05-252), and EU TMR Network contract n° FMRX-CT98-0203.

[†]Université libre de Bruxelles (ECARES) and CEPR.

[‡]Boston University and CEPR.

I. INTRODUCTION

There is wide agreement that organizational design is a crucial determinant of the behavior and performance of the business firm. Yet there is little consensus that it matters at the industry level. Indeed, Industrial Economics, at bottom the study of how firms deliver the goods, has largely ignored the internal organization of its key players; instead, the stage on which they perform, the imperfectly competitive market, dominates the show.

There are good reasons for this. Analytical parsimony is one. More fundamental is the presumption that departures from Arrow-Debreu behavior by the individual firm will be weeded out by the discipline of competition: in effect, imperfection *in the market* is the root of all distortion. Any challenge to this view requires a clear demonstration that imperfections *within firms* can, by themselves, affect industry conduct and performance. In other words, if organizational design matters for the concerns of Industrial Economics, it ought to be apparent in the simplest of all industrial models, perfect competition.

This article provides such a model and shows that internal organization — specifically ownership and control in the incomplete-contract tradition of Grossman-Hart (Grossman and Hart 1986; Hart and Moore 1990) — has distinctive positive and normative implications for industry behavior. It presents a simple textbook-style industry model in which “neoclassical black-box” firms are replaced by partnerships of manager-led suppliers who choose ownership structures to govern trade-offs between private costs and coordination benefits. All other aspects of the model are standard for perfect competition: enterprises and consumers are price takers, the formation of the partnerships between suppliers is frictionless, and entry is allowed.

The organizational building block for the analysis is an adaptation of the model in Hart and Holmström (2010). In order to produce a unit of the con-

sumer good, two complementary suppliers, each consisting of a manager and his collection of assets, must enter into a relationship. The managers operate the assets by making noncontractible production decisions. Technology requires mutual compatibility of the decisions made for different parts of the enterprise.

The problem is that decisions that are convenient for one supplier will be inconvenient for the other, and vice versa, which generates private costs for each manager. This incompatibility may reflect a technological need for adaptability (the BTU and sulphur content of coal needs to be optimally tailored to a power plant's boiler and emissions equipment), or occupational backgrounds (engineering favors elegant design and low maintenance; sales prefers redundant features and user-friendliness). Each party will find it costly to accommodate the other's approach, but if they don't agree on something, the enterprise will be poorly served.

The main organizational decision the managers have to make is whether to integrate. If they retain control over their assets and subsequently make their decisions independently, this may lead to low levels of output, since they overvalue the private costs and are apt to be poorly coordinated. Integration addresses this difficulty via a transfer of control rights over the decisions to a third party called an "*HQ*" who, like the managers, enjoys profit, but unlike them, has no direct concern for the decisions. Since *HQ* will have a positive stake in enterprise revenue, she maximizes the enterprise's output by enforcing a common, compromise decision. The cost of this solution is that the compromise will be moderately inconvenient to both managers.

The industry is composed of a large number of suppliers of each type who form enterprises in a matching market, at which time they decide whether to integrate and how to share revenues. The industry supply curve will embody a relationship between the price and ownership structure, as well as the usual

price-quantity relationship. Once this “Organizationally Augmented Supply” curve (OAS) has been derived, it is feasible to perform a textbook-style supply-and-demand analysis of both the comparative statics and the economic performance of equilibrium ownership structures.

One set of findings concerns the positive implications of the structure of the OAS. First is the relationship between market price and ownership structure. At low prices, managers do not value the increase in output brought by integration since they are not compensated sufficiently for the high costs they have to bear. At higher prices, managers value output enough that they are willing to forgo their private interests in order to achieve coordination, and therefore choose to integrate. Thus, demand matters for ownership structure because it affects the market price. And because the price is common to the whole industry, a demand shift provides a natural source of widespread restructuring, as in a “wave” of mergers or divestitures.

Second is a theory of heterogeneity in ownership structure and performance. Endogenous coexistence of different ownership structures, even among firms facing similar technology, is a generic outcome of market equilibrium. The heterogeneity is an immediate consequence of the productivity difference between integration and nonintegration: while managers may be indifferent between the two, integration produces more output. Thus if per-firm demand is in between the output levels associated with each ownership structure, there must be a mixture of the two in equilibrium. In this part of the supply curve, changes in demand are accommodated less by price adjustments than by ownership changes.

Third, product price is also a key source of “external” influence on organizational design, which raises the possibility of supply-induced restructuring. For instance, technological shocks that occur in a few firms will affect market

price and therefore potentially the ownership choice of many other firms. One implication of this external effect is that positive technological progress may have little impact on aggregate performance because of “re-organizational absorption”: the incipient price decrease from increased productivity among some firms may induce the remaining firms to choose nonintegration, thereby lowering their output and keeping industry output unchanged.

These results help with interpreting some recent empirical findings on the determinants of ownership structure and performance in the air travel and ready-mix concrete industries. Evidence from air travel corroborates our basic result concerning the relation of integration to product price: the tendency for major airlines to own local carriers is greatest in markets where routes are most valuable (Forbes and Lederman, 2009; 2010). In their study of the concrete industry, Hortaçsu and Syverson (2007) report that enterprises with identical technology make different integration choices, a finding that is difficult to reconcile with a model of integration that is based only on technological considerations, but is easily understood in terms of the heterogeneity result. Moreover, their other findings, which relate price and the degree of integration in the market, are explained in terms of supply-driven restructuring.

The model is amenable to simple consumer-producer surplus calculations, which deliver two main welfare results. First, the competitive equilibrium is “ownership efficient,” when managers fully internalize the effect of their decisions on the profit, as in small or owner-managed firms. That is, a planner could not increase the sum of consumer and producer surplus by forcing some enterprises to re-organize.

Second, when instead the managers’ financial stakes are lower, there is a generic set of demand functions for which equilibria are ownership inefficient. Specifically, integration favors consumers because it produces more than non-

integration. Thus, inefficiencies assume the form of too little integration: managers without full financial stakes overvalue their private costs. A counterpart to the Harberger triangle measure of deadweight loss from market power can be identified: this “Leibenstein trapezoid” measures the extent of organizational deadweight loss.¹ But in contrast to the case of market power, in which welfare losses are greatest when demand is least elastic, organizational welfare losses are greatest when demand is *most* elastic.

These distortions are unlikely to be mitigated by instruments that reduce frictions. Indeed, if managers have access to positive cash endowments or can borrow the cash with which to make side payments to each other, they are *more* likely to adopt nonintegration, which hurts consumers. This result offers a new perspective on the costs of “free cash flow”: managers use it to pursue their private interests, but here it may take the form of too little rather than too much integration. In a similar vein, free entry into the product market does not significantly affect results: the “long run” OAS typically has a similar shape to the short run OAS, in particular admitting generic heterogeneity, and the long-run welfare results parallel those of the short run.

Related Literature

There is a long line of research that examines various aspects of organizational behavior and how it interacts with the market environment.² Perhaps the earliest relates the degree of competition to managerial incentives, either in a competitive setting (Machlup 1967; Hart 1983; Sharfstein 1988) or in an imperfectly competitive framework (Fershtman and Judd 1987; Schmidt 1997; Raith 2003). The focus there is on the power of compensation schemes, leaving

1. The idea that persistent underperformance or departures from profit maximization may have organizational origins is related to the “X-inefficiency” tradition started by Leibenstein (1966); see also Bertrand and Mullainathan (2003).

2. Legros and Newman (2013) surveys this literature.

organizational design (and firm boundaries in particular) exogenous. There is also a literature that relates market forces to investment in monitoring technologies (Banerjee and Newman 1993; Legros and Newman 1996), but allocations of decision rights, firm boundaries and ownership are not considered.

Earlier work on the external determinants of ownership structure under perfect competition (Legros and Newman 2008) studies how relative scarcities of different types of suppliers determine the allocation of control. It does not examine the effects of the product market, nor does it consider consumer welfare.

Marin and Verdier (2008) and Alonso, Dessein and Matouschek (2008) consider models of delegation in imperfectly competitive settings; firm boundaries are fixed in these models and the issue is whether information acquisition is facilitated by delegated or centralized decision making.

There is also a literature containing models that explain the pattern of outsourcing and integration in imperfectly competitive industries when there is incomplete contracting. McLaren (2000) and Grossman and Helpman (2002) proceed in the Williamsonian tradition, where integration alleviates the hold-up problem at an exogenous fixed cost, and use search frictions to create relationship specificity. McLaren (2000) shows that globalization, interpreted as market thickening, leads to nonintegration and outsourcing. Grossman and Helpman (2002) adds monopolistic competition with free entry to examine the effects of technological and demand elasticity parameters on the propensity to integrate; it also obtains endogenous heterogeneity in organizational choices, but only as a nongeneric phenomenon. Aghion, Griffith and Howitt (2006) also considers the impact of the degree of competition on the degree of integration and finds a nonmonotonic relationship. There is also a literature in international trade (e.g., Antras 2003; Antras and Helpman 2004), which embeds the Grossman and Hart (1986) framework into models of monopolistic competition, and is focused

on how technological or legal parameters (capital intensity, productivity, contract enforceability) affect the the locational as well as firm-boundary decisions of multinational firms.

Generic heterogeneity of ownership is obtained in the recent paper by Gibbons, , and Powell (2012), though via a very different mechanism. It studies a model in which ownership structure, through its impact on incentives, affects the firm's ability to acquire information about an aggregate state, while that information is also incorporated into market prices. Rational expectations equilibrium entails that some firms acquire information while the rest infer it from prices.

None of these papers addresses the impact of organizational design on performance, which is one of the chief motivations for an organizational Industrial Economics. A perfectly competitive framework allows for transparent treatment of this issue. The model in this article highlights the basic relationship between price *levels* and integration embodied in the OAS. Its organizational heterogeneity is coupled with performance differences among firms. And it provides a simple characterization of the impact of equilibrium ownership structures on consumer and social welfare. The model also points to issues that have received little consideration in the Organizational and Industrial Economic literatures, such as the role of corporate governance on consumer welfare, and the connection among industry performance, finance, and ownership structure that stands in contrast to the strategic use of debt that has been studied in the IO literature (e.g., Brander and Lewis 1986).

II. MODEL

This section presents the basic model, where managers are full claimants to the revenue. The basic organizational building block is a single-good, continuous-

action version of Hart and Holmström’s (2010) model. The aim is to derive an industry supply curve that summarizes the relationships among price, quantity and ownership structure. This is best thought of as a “short run” supply curve, for which entry into the industry is limited. Discussion of entry and long-run supply is deferred to Section IV.A.

II.A. Environment

1. Technology, Preferences and Ownership Structures

There is one consumer good, the production of which requires the coordinated input of one A and one B . Call their union an “enterprise.” Each supplier can be thought of as collection of assets and workers, overseen by a manager, that cannot be further divided without significant loss of value. Examples of A and B might include “lateral” relationships such as manufacturing and customer support, as well as vertical ones such as microchips and computers. The industry will comprise a continuum of each type of supplier, but for the moment confine attention to a single pair.

For each supplier, a noncontractible decision is rendered indicating the way in which production is to be carried out. For instance, networking software and routing equipment could conform to many different standards; material inputs may be well- or ill-suited to an assembler’s production machinery. Denote the decision in an A supplier by $a \in [0, 1]$, and a B decision by $b \in [0, 1]$. The decision might be made by the supplier’s manager, but could also be made by someone else, depending on the ownership structure, as described below.

As the examples indicate, these decisions are not ordered in any natural way; what is important for expected output maximization is not which particular decision is made in each part of the enterprise, but rather that it is coordinated with the other. Formally, the enterprise will succeed, in which case it generates

1 unit of output, with probability $1 - (a - b)^2$; otherwise it fails, yielding 0.

The manager of each supplier is risk-neutral and bears a private cost of the decision made in his unit (these costs are also noncontractible, else decisions could effectively be contracted upon by contracting on costs). The managers' payoffs are increasing in income, but they disagree about the direction decisions ought to go: what is easy for one is hard for the other, and vice versa. Specifically, the A manager's utility is $y^A - (1 - a)^2$, and the B manager's utility is $y^B - b^2$, where y^A and y^B are the respective realized incomes. A manager must live with the decision once it is made: his function is to implement it and convince his workforce to agree; thus regardless of who makes a decision, the manager bears the cost.³

Managers have limited liability (thus $y^A \geq 0$, $y^B \geq 0$) and do not have any means of making fixed side payments, that is, they enter the scene with zero cash endowments. The significance of this assumption is that the equilibrium division of surplus between the managers, which is determined in the supplier market, will influence the choice of ownership structure. By contrast, if cash endowments were sufficiently large, the "most efficient" ownership structure (from the managers' point of view) would always be chosen, independent of supplier market conditions. Arbitrary finite cash endowments are considered in Section IV.C.

The ownership structure can be contractually assigned. Here there are two options; following the property rights literature, each implies a different allocation of decision making power. First, the production units can remain two separate firms (*nonintegration*), in which case the managers retain control over their respective decisions. Alternatively, the managers can *integrate* into a sin-

3. Thus, the cost function need not be a characteristic of the individual manager, although that is one interpretation, so much as an inherent property of the type of input or employee he oversees: if the managers switched assets, they'd "switch preferences" (as happens when a professor of economics becomes a dean of a business school). Similar results could also be generated by a model in which managers differ in "vision," as in van den Steen (2005).

gle firm, re-assigning control in the process. They do so by selling the assets to a headquarters (HQ), empowering her to decide both a and b and at the same time giving her title to (part of) the revenue stream.⁴ Assume that HQ s always have enough cash to finance the acquisition.

HQ 's payoff is simply her income $y^H \geq 0$; thus she is motivated only by monetary concerns and incurs no direct cost from the a and b decisions, which are always borne by the managers of the two units. As a self-interested agent, HQ can no more commit to a pair of decisions (a, b) than can A and B . Integration simply trades in one incentive problem for another.⁵

2. *Contracts*

The enterprise's revenue is contractible, allowing for the provision of monetary incentives via *sharing rules*: any “budget-balancing” means of splitting the managerial share of realized revenue is permissible. For the benchmark analysis, assume that the entire revenue generated by the enterprise accrues to the managers and HQ .⁶ The assumption will be relaxed later to allow for the possibility that the managers and HQ accrue only a fraction of the revenue, the rest of which goes to “shareholders.”

A *contract* for A, B is a choice of ownership structure and conditional on that a share of revenues accruing to each manager.

4. In fact, as discussed below, giving HQ the power to decide (a, b) implies that she will get a positive share of the revenue.

5. An alternate way to integrate would be to have one of the managers sell his assets to the other. It is straightforward to show (Section II.C.1) that this form of integration is dominated by other ownership structures in this model — the cost imposed on the subordinate manager is simply too great.

6. Since there is a single product produced by the enterprise and the revenue from its sale is contractible, it does not matter in which hands the revenue is assumed to accrue initially. A more general formulation would suppose that the suppliers produce complementary products (e.g., office suite software and operating systems) that generate separate revenue streams, which accrue separately to each supplier; the present model corresponds to the case where these streams are perfectly correlated. A full treatment of that case, which not only admits the possibility of richer sharing rules in which each manager gets a share of the other's revenue, but also would form the basis for a property-rights theory of multi-product firms, is beyond the scope of this paper, so to speak.

- Under nonintegration, a contract specifies a share s accruing to A when output is 1; B then gets a share $1 - s$. By limited liability, each manager gets a zero revenue when output is 0.⁷
- Under integration, HQ buys the assets A, B for prices of π_A and π_B in exchange for a share structure $\mathbf{s} = (s_A, s_B, s_H)$ where $\mathbf{s} \geq \mathbf{0}$ and $s_A + s_B + s_H = 1$.

The revenue shares along with the asset prices π_A and π_B are endogenous, and will be determined in the overall market equilibrium.

3. Markets

The *product market* is perfectly competitive. On the demand side, consumers take P as given and maximize a smooth, quasilinear utility function taking. This optimization yields a differentiable demand $D(P)$. Suppliers also take the (correctly anticipated) price P as given when they sign contracts and make their production decisions.

In the *supplier market*, there is a continuum of A and a continuum of B suppliers with potentially different measures. In the *HQ market*, HQ s are supplied perfectly elastically with an opportunity cost normalized to zero.

II.B. Equilibrium

Equilibrium in this model consists of a *stable match* in the supplier market and market clearing in the product market. There are two types of enterprise, which correspond to the formation of the following coalitions:

7. There is no allowance for third-party budget breakers: as is well known, they can improve performance only if they stand to gain when the enterprise fails. (Note also that because output in case of failure is equal to zero, budget breaking is not effective when managers have zero cash endowments.) For simplicity, borrowing from third parties to make side payments is also not allowed; this assumption is relaxed in Section IV.C.

- Coalitions consisting of one A supplier and one B supplier. The feasible set for these coalitions will depend, among other things, on the product market price and corresponds to the set of payoffs that are achievable through contracts when supplier A has ownership of asset A and supplier B has ownership of asset B .
- Coalitions consisting of one A supplier, one B supplier and one HQ . The feasible set for these coalitions will depend on the product market price and corresponds to the set of payoffs that are achievable through contracts where HQ has ownership of the assets of A, B .

The feasible set of payoffs for each type of enterprise is derived in the following subsection.

Equilibrium may also involve trivial coalitions consisting of singleton agents. These have feasible sets that are independent of the price and coincide with payoffs no larger than an exogenous opportunity cost. For HQ s, this cost is zero, but for the other agents it may be positive.⁸

For a given coalition, the contract that is chosen determines the decisions that will be taken and thereby the probability that the enterprise produces a positive output. Though output is a random variable at the enterprise level, the law of large numbers implies that industry output is deterministic. Hence, once coalitions are formed and contracts are signed, there is a well defined *industry supply* $S(P)$.

DEFINITION 1. An equilibrium consists of a partition of agents into coalitions, a payoff to each agent and a product price P satisfying:

- (1) *Feasibility*: the payoffs to the agents in an equilibrium coalition are feasible given the equilibrium price P ;

8. While there are many other (larger) potential coalitions, the production technology is such that none of them can achieve payoffs different from what can be achieved by unions of the ones already described.

- (2) *Stability*: no coalition can form and find feasible payoffs for its members that are strictly greater than their equilibrium payoffs;
- (3) *Market Clearing*: the total supply in the industry $S(P)$ is equal to the demand $D(P)$.

II.C. Choice of Organization

Consider a matched pair A and B who will accrue the entire enterprise revenue P in case of success (0 in case of failure). For each possible contract, they anticipate behavior and resulting payoffs, and choose one that optimizes B 's payoff while guaranteeing A his equilibrium payoff. To construct the Pareto frontier for A and B , it is convenient to treat each ownership structure in turn.

1. nonintegration

Since each manager retains control of his activity, given a share s , A chooses $a \in [0, 1]$, B chooses $b \in [0, 1]$. A 's payoff is $(1 - (a - b)^2)sP - (1 - a)^2$ and B 's is $(1 - (a - b)^2)(1 - s)P - b^2$.

The (unique) Nash equilibrium of this game is:

$$(1) \quad a = 1 - s \frac{P}{1 + P} \quad , \quad b = (1 - s) \frac{P}{1 + P}.$$

The resulting expected output is:

$$(2) \quad Q^N(P) \equiv 1 - \frac{1}{(1 + P)^2}.$$

For a given value of s , as the revenue P increases, A and B are more willing to concede to each other (a decreases and b increases). Output is therefore increasing in the price P : larger values raise the relative importance of the revenue motive against private costs, and this pushes the managers to better

coordinate. The functional forms generate a convenient property for the model, namely that the output generated under nonintegration does not depend on s , i.e., on how the managers split the firm's revenue. This will also be true of integration.

Of course, the managers' *payoffs* depend on s ; they are:

$$(3) \quad u_A^N(s, P) \equiv Q^N(P)sP - s^2 \left(\frac{P}{1+P} \right)^2$$

$$(4) \quad u_B^N(s, P) \equiv Q^N(P)(1-s)P - (1-s)^2 \left(\frac{P}{1+P} \right)^2.$$

Varying s , one obtains a Pareto frontier given nonintegration. It is straightforward to verify that it is strictly concave in u_A - u_B space, a result of the convex cost functions. The total managerial payoff

$$(5) \quad U^N(s, P) \equiv Q^N(P)P - (s^2 + (1-s)^2) \left(\frac{P}{1+P} \right)^2$$

varies from $\frac{P^2}{1+P}$ at $s = 0$ (or $s = 1$) to $(\frac{3}{2} + P) \left(\frac{P}{1+P} \right)^2$ at $s = \frac{1}{2}$.

Nonintegration has clear incentive problems: A and B managers put too little weight on the organizational goal in favor of their private benefits. The alternative is integration, but this introduces other incentive problems: whoever makes the decision will put too little weight on the private costs. An extreme version of this is when integration gives title to A or B : in this case the decisions made are so costly to the other manager that the ownership structure is dominated by nonintegration. To see this, note that if B were to have control (the argument is similar for A), he would choose a and b to maximize his own payoff $(1 - (a - b)^2)(1 - s)P - b^2$, which entails $a = b = 0$. This maximizes A 's cost, and the total surplus is only $P - 1$, which is less than $\frac{P^2}{1+P}$, the lowest

nonintegration surplus. Thus B -control is Pareto dominated by nonintegration, and the only other organizational form of interest is when control is given to an HQ .

2. Integration

Consider an integration contract in which the shares of the success revenue are $\mathbf{s} = (s_A, s_B, s_H)$, and suppose that HQ has financed the asset acquisition with cash.⁹ As long as $s_H > 0$, HQ will choose to maximize output since her objective function is $(1 - (a - b)^2)s_HP$. Hence the decisions that will be taken by an HQ must satisfy $a = b$; assume that HQ opts for $a = b = 1/2$, which minimizes the total managerial cost $(1 - a)^2 + b^2$ among all such choices. The cost to each manager is then $1/4$. Since HQ s compete and have zero opportunity cost, the purchase prices for the assets must total s_HP . Total managerial welfare under integration is therefore $U^I(P) \equiv P - 1/2$, which is fully transferable between A and B via adjustments in \mathbf{s} or the asset prices. The reason for transferability is simple: the actions taken by HQ and the costs borne by A and B do not depend on their shares. Neither, of course, does integration output. Hence, the Pareto frontier under integration is $u_B = P - 1/2 - u_A$.

Notice that the cost of integration is fixed, independent of P . This is a *result* of the fact that HQ is an incentive-driven agent who has a stake in the firm's revenue. If she had no stake ($s_H = 0$), HQ would be acting as a "disinterested authority," indifferent among all decisions $(a, b) \in [0, 1]^2$, and hypothetically she could be engaged by the managers to make the first-best choices.¹⁰

The problem with this is that she would be equally happy to choose the

9. The Appendix shows that HQ needs to have some cash in order for integration to emerge, although the level of cash may be arbitrarily small; she will always have a positive success revenue, and that is all that matters here.

10. These maximize $(1 - (a - b)^2)P - (1 - a)^2 - b^2$, yielding $(a^*, b^*) = (\frac{1+P}{1+2P}, \frac{P}{1+2P})$ (thus $a^* \neq b^*$) and costs $\frac{2P^2}{(1+2P)^2}$, which do depend on P . The first-best surplus is $\frac{2P^2}{1+2P}$.

“doomsday option,” setting $a = 0, b = 1$, thereby inflicting maximal costs on the managers and generating zero output. For this reason, disinterested authority is not feasible: HQ would always use the doomsday threat (or a suitably mollified one) to renegotiate a zero-share contract to one with a positive share.¹¹ Anticipating this renegotiated outcome, the managers will give her a positive stake in the first place.¹²

3. *Comparison of Ownership Structure*

Ownership structure presents a tradeoff for the A and B managers: relative to the first best, nonintegration generates too little coordination; integration generates too much. The choice of ownership structure will depend on distributional as well as efficiency considerations: given P , total managerial welfare is constant under integration, independent of how it is distributed, while for nonintegration, it depends on s and therefore on surplus division. In general, neither Pareto dominates the other, so that supplier market equilibrium, which determines the distribution of surplus and s , as well as product market equilibrium, which determines P , will both influence the decision whether to integrate.

The relationship between ownership choice and market price has a simple characterization in the managers’ payoff space. Since the minimum nonintegration welfare is $\frac{P^2}{1+P}$ (corresponding to $s = 0$ or 1), which exceeds $P - 1/2$ if and only if $P < 1$, nonintegration dominates integration at low prices. Because the frontiers are symmetric about $u_A = u_B$, with the integration frontier linear and the nonintegration frontier strictly concave, they intersect twice. Of course, as

11. See Legros and Newman (2012) for a more detailed argument.

12. There are of course other reasons why an agent with significant control rights would have a positive stake of contractible revenue. Moral hazard is one: if enforcing the decisions involves any non verifiable cost, giving HQ a large enough share will ensure she acts. Or, if she has an ex-ante positive opportunity cost of participating, the cash-constrained managers would have to give her a positive revenue share to compensate. Finally, the managers could engage in a form of influence activities, lobbying for their preferred outcomes with shares of revenue.

P increases, both managers may accrue higher payoffs, and the frontiers shift away from the origin. The loci of intersections is given by $|u_A - u_B| = \frac{P}{1+P}$.¹³

To summarize:

PROPOSITION 1.

- (a) Integration is chosen when product price $P > 1$ and $|u_A - u_B| > \frac{P}{1+P}$.
- (b) nonintegration is chosen when $P < 1$ or $|u_A - u_B| < \frac{P}{1+P}$.
- (c) Either ownership structure may be chosen when $P \geq 1$ and $|u_A - u_B| = \frac{P}{1+P}$.

The result is depicted in Figure I, where the $|u_A - u_B| = \frac{P}{1+P}$ locus is represented by the dark heavy curves; the integration regions (I) and nonintegration region (N) are also shown, along with integration (straight) and nonintegration (curved) frontiers for some price exceeding 1.

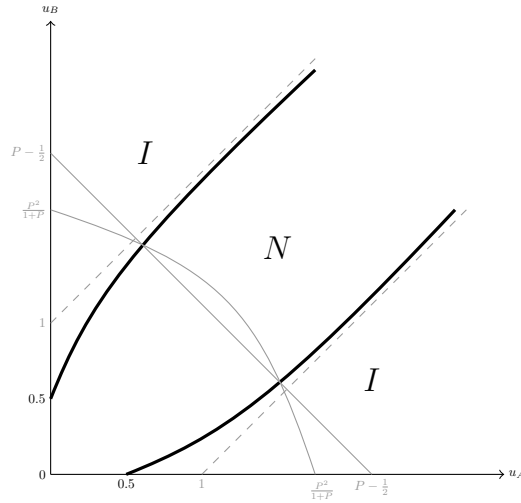


FIGURE I: DIVISION OF THE SURPLUS AND INTEGRATION DECISIONS

13. To see this, use (3) and (4) to get the absolute difference in payoff $|u_A - u_B| = |2s - 1| \frac{P^2}{P+1}$; setting $P - 1/2 = \left(\frac{P}{1+P}\right)^2 (2 + P - s^2 - (1-s)^2)$ to solve for s (solutions exist only for $P \geq 1$) gives $|2s - 1| = 1/P$, from which the result follows.

The last result (c) will play a significant role in generating heterogeneity of organizational form (co-existence of integrated and nonintegrated enterprises).¹⁴

Parts (a) and (b) underscore the role of finite cash endowments, for they show how integration emerges partly as a consequence of inequality in managerial surplus division, and partly in response to the industry price.

If endowments were unbounded, then for any price, the managers would choose the surplus-maximizing organization, namely nonintegration with $s = 1/2$, and settle their distributional claims with cash side payments. But with zero endowments, an given ownership structure, allocations of surplus between the managers must be accommodated via other instruments, namely changes in the shares of revenue or asset prices (for the positive finite case, the situation is similar: see Section IV.C). Under integration, such changes do not alter incentives, because the decisions are made by HQ , who always acts in a revenue-maximizing way. But under nonintegration, changing the shares is nonneutral: inequality of surplus division implies inequality of incentives, and the result is poor overall performance of the nonintegrated enterprise. Integration therefore has a comparative advantage in distributing surplus. If supplier market conditions (relative scarcities, opportunity or entry cost differences) lead to a mismatch between the division of surplus and nonintegration's optimal division of incentives, integration prevails. Thus, surplus division is one determinant of ownership structure.

The other determinant of ownership structure is the industry price level. Some intuition for its role can be obtained by considering different exogenously set payoffs \hat{u}_A for the A s. Suppose first that $\hat{u}_A = 0$. Then under nonintegration $s = 0$, and the A manager has *no* interest in revenue. He sets $a = 1$, thereby bearing no cost. Manager B 's cost will increase with P , since his large interest

14. Though nonconvexities in the overall Pareto frontier create an incentive to engage in them, for simplicity lotteries between ownership structures are not allowed.

in revenue will induce him to increase his concession to A . As revenue becomes large, B 's cost is driven close to the maximum, which because of the convexity of the cost functions, leads to large total cost. Integration has the benefit of forcing A to compromise; he needs to be compensated with a positive share of revenue or the asset sale proceeds, but if the total revenue is large, this is a small sacrifice. Thus, when P is high enough, B will prefer integration. If P is small, though, then the fixed cost of integration is not worth its improved output performance, which has little value in the market.

On the other hand, for a larger value of \hat{u}_A , the revenue share under nonintegration is closer to $1/2$, and A will take account of the revenue as well as his private costs. Letting \hat{u}_A increase with P in such a way as to keep s close to $1/2$ (i.e., moving along the 45° -line in the figure), the concession increases with P . The decisions will remain on either side of $\frac{1}{2}$, so output will fall short of the integration level, but at high revenues this is a small gap. Taking account of the private costs, nonintegration remains preferable.

As the price increases, the contribution of inequality to the integration decision diminishes: if the price is large, nonintegration is chosen only if the surplus division is close to perfect equality.¹⁵

The incomplete contracts literature has tended to emphasize the technological (supply-side) aspects, though distributional aspects have received some attention (Aghion and Tirole 1994, Legros and Newman 2008). The present analysis emphasizes the additional role played by demand, and Proposition 1 illustrates the interplay of demand (P) and distribution (u_A, u_B).

15. A measure of inequality on the indifference loci is $\frac{|u_A - u_B|}{u_A + u_B} = I(P) \equiv \frac{2P}{(1+P)(2P-1)}$, which has a maximum value of 1 at $P = 1$ and declines monotonically to 0 as P gets large. Integration occurs whenever inequality exceeds $I(P)$; thus integration becomes more likely as enterprise value increases.

II.D. Industry Equilibrium and the “Organizationally Augmented Supply”

Industry equilibrium comprises a general equilibrium of the supplier, HQ and product markets. To focus on the role played by market price in determining organizational design, assume that the B suppliers have a zero opportunity cost of participating in this industry and that the measure of A suppliers exceeds that of the B s, which is equal to unity. Thus some of the A s will remain unmatched and receive their opportunity cost of \underline{u}_A . Stability then implies that matched A s receive \underline{u}_A as well.

The simplest case is when $\underline{u}_A = 0$. To derive the industry supply, suppose that a fraction α of firms are integrated and a fraction $1 - \alpha$ are nonintegrated. Total supply at price P is then almost surely (because of the continuum of enterprises and the law of large numbers):

$$\alpha + (1 - \alpha)Q^N(P).$$

When $P < 1$, $\alpha = 0$ and total supply is just the output when all firms choose nonintegration. At $P = 1$, α can assume any value between 0 and 1: by Proposition 1, managers are indifferent between the two forms of organization, but as output is greater with integration, total supply increases with α . When $\alpha = 1$ output is 1 and stays at this level for all $P \geq 1$.

Write $S(P)$ to represent the supply correspondence, where α depends on P as described in the previous paragraph. The supply curve is represented in Figure II.

An equilibrium in the product market is a price and a quantity that equate supply and demand: $D(P) \in S(P)$. There are three distinct types of industry equilibria illustrated in Figure II, depending on where along the supply curve the

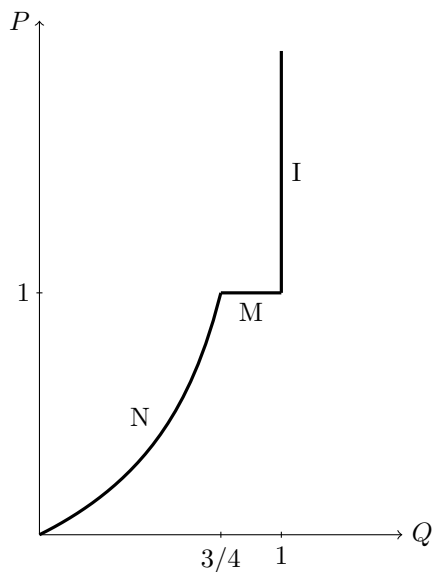


FIGURE II: ORGANIZATIONALLY AUGMENTED SUPPLY CURVE WHEN $\underline{u}_A = 0$

equilibrium price occurs: those in which firms integrate (I), the mixed equilibria in which some firms integrate and others do not (M), and a pure nonintegration equilibrium (N). The model predicts a monotonic relationship between price and integration. As demand increases, the equilibrium price increases, inducing a greater tendency to integrate.

For small positive values of \underline{u}_A (specifically, $\underline{u}_A < 1/2$) the supply curves are the same as when $\underline{u}_A = 0$ except that (i) the shift to integration happens at a price $P^*(\underline{u}_A)$ that is greater than 1 and increasing in \underline{u}_A ; and (ii) supply is positive only if P is large enough to generate a surplus of at least \underline{u}_A .¹⁶

The airline industry displays a pattern of ownership between major and regional carriers that can be interpreted as movement along the OAS (see Forbes

16. Use the equilibrium condition $u_A = \underline{u}_A$, the indifference locus $u_A = u_B - \frac{P}{1+P}$, and A 's value of integration there $u_A = P - \frac{1}{2} - u_B$ to obtain $P^*(\underline{u}_A)$ as the solution to $\underline{u}_A = \frac{P^2}{2(1+P)} - \frac{1}{4}$. The case $\underline{u}_A > \frac{1}{2}$ leads to the interesting theoretical possibility of a nonmonotonic supply curve.

and Lederman, 2009; 2010). The good required by a traveller will often consist of two complementary flight segments, one served by the regional carrier and one served by the major, which corresponds rather closely to the situation modeled here. Majors own some regional flights (integration) and outsource to independent firms for others (nonintegration).

The evidence suggests first that integrated relationships perform better (fewer delays and cancellations), just as in the model, and second, that integration is more prevalent on routes with high demand or where there are higher costs of delay (regional routes with endpoints that have more of the major's departing flights or that involve a hub). Because P is the revenue lost when a good fails to be delivered, it is a cost of delay as well as a reflection of demand. Whether one interprets the involvement of a hub or a high number of connections as high demand or large delay cost, the model's basic prediction that integration is associated with higher enterprise value (P) is consistent with this evidence.

III. CONDUCT AND PERFORMANCE OF AN ORGANIZATIONAL INDUSTRY

This section discusses three properties of the model that pertain to industry conduct and performance. First, there is robust coexistence of different ownership structures. Second, the organization of one enterprise depends not only on its own technology and managers' preferences but also on prices determined outside it, implying that "local" changes can have industry-wide effects. Third, equilibrium has welfare properties readily characterized in terms of consumer and producer surplus. The model is then used to make sense of evidence from the ready-mix concrete industry.

III.A. Heterogeneity of Ownership Structure

In the mixed region of the OAS (M) there is coexistence of organizational forms (ownership structure) within the industry. Notice that this organizational heterogeneity is an endogenous consequence of market clearing given a discrete set of ownership structures, and occurs even though all firms are ex-ante identical.

There is much evidence of productivity variation within industries even among apparently similar enterprises; Syverson (2011) notes that within 4-digit SIC industries in the U.S. manufacturing sector, a “plant at the 90th percentile of the productivity distribution makes almost twice as much output with the same measured inputs as the 10th percentile plant” and that other work on Chinese or Indian firms find even larger differences. Moreover, as pointed out by Gibbons (2006, 2010), there is much evidence of correlation between organizational and productivity variation. The airline case (Forbes and Lederman 2009, 2010) discussed earlier is an example: major airlines own some of their regional flights (integration) and outsource others (nonintegration), even among flights with the same origin airport on the same day, and despite the superior performance of the integrated relationships. But there is little theoretical work attempting to explain how these organizational and performance differences can persist.¹⁷ The model here provides a simple explanation for part of this correlation, since whenever they coexist, a nonintegrated enterprise generates less expected output than an integrated one.

Observe that while there is only a single price at which the heterogeneous outcome occurs, it is generated by a generic set of demand functions:

17. Beside the recent paper by Gibbons, Holden and Powell (2012) discussed in the Introduction, there is an early contribution by Hermalin (1994), which obtains heterogeneity in incentive schemes in a principal-agent model when the product market is imperfectly competitive.

PROPOSITION 2. Consider any demand function $D(P)$ that is positive and has finite elasticity at $P = P^*(\underline{u}_A)$. There exists a nonempty open interval $(\underline{\delta}, \bar{\delta}) \subset \mathcal{R}_+$ such that for any demand function $\delta D(P)$, with $\delta \in (\underline{\delta}, \bar{\delta})$, there is a mix of nonintegrated and integrated firms in equilibrium.

Recall that the pool of HQ s is large enough to integrate every enterprise. If instead the HQ s are in short supply, heterogeneity is even more endemic. Indeed, at all prices exceeding $P^*(\underline{u}_A)$, managers would prefer integration if the HQ s continued to accrue zero net surplus, but since there are not enough HQ s to go around, some enterprises must be nonintegrated; in equilibrium, HQ s extract enough surplus to render managers indifferent between integration and nonintegration. Similarly, admitting lotteries among ownership structures would also make heterogeneity easier to obtain.

In contrast to the robust co-existence of ownership structures found here, other papers investigating endogenous heterogeneity (notably Grossman and Helpman 2002) have found it to be nongeneric, occurring only for a singular set of parameters. Further consideration of the difference in results is deferred to the discussion of entry in Section IV.A.

III.B. Supply Shocks and External Effects

All enterprises face the same price, so anything that affects it – a demand shift, foreign competition, or a tax on profits – can lead to widespread and simultaneous reorganization, as in a merger or divestiture wave. Straightforward demand and supply analysis can be used to study these phenomena. For instance, growth in demand might raise the price from below $P^*(\underline{u}_A)$ to above it, resulting in a merger wave as firms switch from nonintegration to integration.

By the same token, the organization of a particular enterprise depends not only on its own technology and managers' preferences but also on prices (prod-

uct price and surplus division) determined outside it. In particular, technological “shocks” that directly affect some firms may induce reorganizations to *other firms* that are unaffected by the shock, as well as to themselves; in fact, sometimes *only* the unaffected firms reorganize, as in the following example.

A positive technological shock (e.g., a product or process innovation) raises the success output in joint production to $R > 1$ for a fraction z of the B-suppliers. For these affected enterprises, expected output is now equal to $Q^N(RP)R$ under nonintegration and to R under integration. Assuming that $\underline{u}_A = 0$, so that $P^*(\underline{u}_A) = 1$ (positive u_A is similar), managers are indifferent between the two ownership structures when $PR = 1$: integration occurs for the innovating firms if the new equilibrium price is greater than $1/R$. For the unaffected firms, the supply correspondence is unchanged. The industry supply is a convex combination of the supplies for the affected and unaffected firms. In particular, the supply is increasing in z .

Let demand have constant elasticity, $D(P) = P^{-\epsilon}$, with $\epsilon > 1$. In the absence of a shock ($z = 0$), the market clearing condition $S(P) = D(P)$ requires that $P = 1$; in this case $S(1) = D(1) = 1$ and though managers are indifferent between the two ownership structures, market clearing requires that *all firms are integrated*.

Consider two cases.

Homogeneous shocks: $z = 1$. All firms success output is now $R^* > 1$. If all firms are integrated, which requires that in the new equilibrium, $P^* > 1/R^*$, the market clearing condition is $R^* = (P^*)^{-\epsilon}$, or $P^* = 1/R^{*(1/\epsilon)} > 1/R^*$, where the inequality follows from the fact that R^* and ϵ both exceed 1. If a positive measure of firms were nonintegrated, then $P^* \leq 1/R^*$, but then demand $(P^*)^{-\epsilon} \geq (R^*)^\epsilon > R^*$ would exceed supply. Thus, the only equilibrium has no change in organization after the shock – *all firms remain integrated*, and

industry output increases to R^* .

Heterogeneous shocks: $z < 1$. Suppose the affected enterprises are subject to a larger shock $R > R^*$, where the average productivity change is the same, that is:

$$zR + 1 - z = R^*.$$

Since supply is increasing in z , the new equilibrium price cannot exceed 1. On the other hand, since supply is bounded above by R^* , from the calculation done above for homogeneous shocks, equilibrium price will always exceed $1/R^* > 1/R$, so *none of the shocked firms re-organize*. In fact, market clearing will require that *at least some of the unshocked firms reorganize* by becoming nonintegrated: if the price falls below 1, all of them do, and if it remains at 1, they cannot all remain integrated, for supply would be R^* , exceeding demand, which is 1. This is an example of an organizational *external effect*: the impetus for organizational change may come from outside the firm, transmitted by the market.

Because the newly nonintegrated enterprises produce less than they did before, there is a “reorganizational dampening” effect from the heterogeneous shocks: in contrast to the homogeneous case, aggregate output must end up being *less* than R^* . In fact, in case the price remains at 1, which obtains for an open set of parameter values, *all* of the productivity increase is absorbed by reorganization; some managers (the innovating B s) benefit, but consumers do not.¹⁸ It would be difficult to get this kind of effect in a competitive model with neoclassical firms: two distributions of shocks that lead to the same aggregate production set, as in this case, would lead to the same output and price outcome.

18. For the price to remain at 1 requires that demand exceed supply when all unshocked enterprises are nonintegrated, or $zR + (1 - z)Q^N(1) \leq D(1) = 1$; using the defining equation for R and $Q^N(1) = \frac{3}{4}$, this is equivalent to $z + 4R^* < 5$.

This example may be summarized by saying:

- A firm benefiting from a significant change in technology need not re-organize.
- A firm that undergoes a large re-organization need not have experienced any change in technology.
- Re-organizational dampening may substantially absorb the aggregate benefit of heterogenous technological improvements.

Much empirical work in the property rights or transaction cost tradition on the determinants of integration has focused on “supply-side” factors, e.g., asset specificities or complementarities (see, e.g., Whinston 2001 for a summary). The present model points to the importance of demand: once taken into account, the simple intuition of supply-side analysis may be overturned.

III.C. Welfare

Since integration is the more productive ownership structure, consumers benefit when more firms integrate. A natural question to ask is whether there are equilibria that suffer from “too little” integration. If one follows the industrial organization convention and focuses on total welfare of all the industry’s participants, then the benefits that consumers experience from more integration must outweigh any losses in terms of private costs that others (particularly managers) might suffer.

Consider a central planner who imposes an ownership structure on each enterprise, but allows participation, noncontractible decisions, prices and quantities to be determined by the market. Define an equilibrium to be *ownership efficient* if welfare cannot be increased by imposing on some enterprises an ownership structure that differs from the one they choose in equilibrium. Since

such forced re-organization may hurt the B managers (A managers always get their opportunity cost), if the planner can make lump sum transfers from the beneficiaries of the reorganization to the managers, then ownership inefficient equilibria are also Pareto inefficient. Notice this criterion is weak in the sense that it does not empower the planner to set the managerial shares s , let alone the decisions a and b , for any enterprise.¹⁹ In what follows, confine attention to the most straightforward case, where $\underline{u}_A < \frac{1}{2}$.

1. *Full Revenue Claims*

The analysis of ownership efficiency follows the familiar competitive logic. Let π denote the success revenue of the managers. Since every A manager gets his opportunity cost regardless of the ownership decision, a marginal B manager effectively decides to integrate if the revenue from the extra expected output generated by integration, valued at π per unit, exceeds his private cost. If in turn π is also the value to consumers of that extra output, then their willingness to pay equals the cost, and the outcome is efficient. In equilibrium, consumers value the extra output at the price P . Thus, $\pi \equiv P$ implies:

PROPOSITION 3. When managers have full residual claim on revenues, equilibria are ownership efficient.

2. *Managerial Firms*

One thing that distinguishes the welfare analysis of an organizational industry from the standard competitive one is that often the architects of ownership

19. A stronger concept of efficiency would also allow the planner to impose the share s . In this case, it is welfare maximizing to set $s = 1/2$ whenever there is nonintegration, which would never be part of an equilibrium when the opportunity cost \underline{u}_A is low and managers have no initial wealth. However, this higher welfare would be generated at the expense of the B -managers in favor of the A s, who would not be able to make lump-sum compensating transfers. Indeed, as is shown in Section IV.C, if the A s had the cash to make such transfers, they would choose $s = 1/2$ themselves.

structure do not have a full pecuniary stake in the enterprise. If the managers claim only a fraction of the revenue per unit of output ($\pi < P$), with the remainder accruing to “shareholders,” then they will tend to underweight output, relative to its social value, in favor of their private costs. Production and organization decisions are governed by the managers’ unit return π rather than the market price P , and the industry supply curve will be shifted to the left: output under nonintegration will be smaller at each price than in the case of full residual claims (integration output will continue to be equal to 1). More interestingly, the market price at which firms integrate will increase, since managers will be content to integrate only when $\pi \geq P^*(\underline{u}_A)$. The fact that marginal output has a social value of P means that in the managerial firm case, the choice of organization need not maximize social welfare: there is now a possibility of too little integration.

To pursue this point, consider the simplest possible representation of diluted pecuniary stakes, in which shareholders are “passive” and receive a price-independent dividend of $1 - \gamma$ per unit of revenue. The shareholders, like *HQs*, only care about income. However they are unable to choose either the revenue share accruing to the managers or the contractual variables (ownership structure and s), decisions over which remain with the managers.²⁰ Similarly, the planner is not empowered to alter γ .

For managerial enterprises, ownership inefficiency is quite common:

PROPOSITION 4. Suppose that managers pay a proportion $1 - \gamma$ of the revenue as dividends to shareholders ($\gamma < 1$). There is a generic set of demand

20. Section IV.D provides a summary discussion of how the presence of active shareholders, who choose only their dividend share $1 - \gamma$ optimally, potentially as a function of the price level, has little impact on the results, whereas if they choose ownership structure as well, they are apt to integrate *too often*. As long as the managers cannot “buy the firm” from the shareholders by using cash (borrowing for this purpose will not help since the debt repayment is akin to having $\gamma < 1$ from the managers’ perspective), any second-best contract between shareholders and managers will entail a residual share $\gamma < 1$ for them, leading to ownership inefficient choices, either by them or by shareholders if the latter are “active”.

functions for which equilibrium is ownership inefficient.

To see this, suppose the equilibrium price is P , with the fraction α of integrated enterprises less than one; thus $D(P) < 1$ and $P \leq P^*(\underline{u}_A)/\gamma$. Each nonintegrated enterprise bears managerial cost $\phi(\gamma P)$, while integrated ones bear cost $\frac{1}{2}$.²¹ Suppose the planner forces a single enterprise to increase its output via a small increase in (the probability of) integration $d\alpha$ (there is no effect on the market price). This raises its output by $(1 - Q^N(\gamma P))d\alpha$ and the enterprise's managerial cost by $(\frac{1}{2} - \phi(\gamma P))d\alpha$. Thus the marginal cost of this output increase is $\frac{\frac{1}{2} - \phi(\gamma P)}{1 - Q^N(\gamma P)}$. If this is less than the consumers' marginal willingness to pay, which is equal to the equilibrium price P , then the planner's forced integration has increased welfare, and equilibrium is ownership inefficient.

It is not hard to find prices at which this can happen: simply let $P = P^*(\underline{u}_A)/\gamma$, the market price at which managers are indifferent between the two ownership structures. Then the planner's marginal cost is $\frac{\frac{1}{2} - \phi(P^*(\underline{u}_A))}{1 - Q^N(P^*(\underline{u}_A))}$, which by definition is equal to $P^*(\underline{u}_A)$. Since $\gamma < 1$, $P > P^*(\underline{u}_A)$, i.e., it exceeds the planner's marginal cost, and there is ownership inefficiency. Notice that this equilibrium price also generates coexistence of the two ownership structures. Thus, *with managerial firms, heterogeneity implies inefficiency*. From Proposition 2, the set of demands that generate coexistence, and therefore ownership inefficiency, is generic.

Typically, though, inefficiency is more endemic than heterogeneity.²² Making γ small ensures a wide range of inefficient equilibrium prices, including some

21. Denote by $\phi(\pi)$ the total private cost for a pair of managers under nonintegration when their revenue is π , which is $(s^2 + (1-s)^2)(\pi)^2(1+\pi)^{-2}$, where s satisfies $s\pi Q^N(\pi) - s^2(\pi)^2(1+\pi)^{-2} = \underline{u}_A$. The notation suppresses the dependence of $\phi(\cdot)$ on \underline{u}_A .

22. Since the inverse demand and the planner's marginal cost are continuous in P , the strict inequality between them that holds at $P^*(\underline{u}_A)/\gamma$ also holds on an open interval of prices below that, so there is inefficiency there. Moreover, since $s^2 + (1-s)^2 \geq \frac{1}{2}$, $\phi(\gamma P) \geq \frac{1}{2}(\gamma P)^2(1+\gamma P)^{-2}$, and the planner's marginal cost is bounded by $\frac{1}{2} + \gamma P$. Thus, for ownership inefficiency it suffices that P lie between $\frac{P^*(\underline{u}_A)}{\gamma}$ and $\frac{1}{2(1-\gamma)}$, a range that is nonempty if $\gamma < 2/3$, since $P^*(\underline{u}_A) \geq 1$.

below $P^*(\underline{u}_A)$ (γ need only be less than $1/2$). The reason forced integration is beneficial at such low prices is that nonintegration is especially poor at output production when managers are not full revenue claimants — a kind of “intensive margin” inefficiency, which supplements the “extensive margin” inefficiency wherein they only integrate at excessively high prices.

All of this says little about the size of the welfare loss, or precisely the extent to which the planner should integrate to maximize welfare. For more than marginal amounts of forced integration, there will be a decrease in the market price: this reduces the welfare gains from such policy, both because of the reduction in willingness to pay and (at least when $u_B > u_A$) reduced costs of nonintegration. Thus, demand elasticity plays a role in determining the extent of inefficiency, just as it does in market power settings, but as is shown next, its effect is quite different.

3. Graphical Analysis and the Role of Demand Elasticity

A simple illustration of Propositions 3 and 4 can be obtained when $\underline{u}_A = 0$. In this case, under nonintegration $s \equiv 0$ and A therefore chooses $a = 1$ regardless of the output price. B 's choice of b therefore determines expected output. Since $s \equiv 0$, $\phi(\pi) = \pi^2(1 + \pi)^{-2}$. It is helpful to write this in terms of quantity q using the invertibility of $Q^N(\cdot)$: if $q = Q^N(\pi)$, $\phi(\pi) = c(q) \equiv (1 - \sqrt{1 - q})^2$.

Suppose the enterprise produces expected output $Q \leq 1$ by randomizing between integration with output 1, and nonintegration with expected output q . Then the integration probability α must be $\frac{Q - q}{1 - q}$, and the cost of this production strategy is $\alpha \frac{1}{2} + (1 - \alpha)c(q)$. Let $C(Q) \equiv \min_{q \leq Q} \alpha \frac{1}{2} + (1 - \alpha)c(q)$ (the constraint $q \leq Q$ derives from the requirement $\alpha \geq 0$). The solution is the interior optimum $\frac{3}{4}$ if $Q > \frac{3}{4}$; else the constraint binds and the optimum is $q = Q$. Thus the marginal cost is $C'(Q) = c'(Q)$ for $Q < \frac{3}{4}$ and $C'(Q) = 1$ for $Q > \frac{3}{4}$.

Maximizing B 's payoff $\pi Q - C(Q)$ subject to A receiving his opportunity

cost of 0 implies that the equilibrium level of output solves $\pi = C'(Q)$. If $\gamma = 1$, $\pi = P$ (full revenue claims), yielding the conventional efficient equilibrium characterization that price equals marginal cost, as in Proposition 3. In other words, when $\underline{u}_A = 0$, the full-revenue-claim supply coincides with the marginal cost schedule. Efficient production corresponds to the intersection of demand with this schedule. But for $\gamma < 1$, $\pi = \gamma P < P$: the marginal cost of production is lower than the price (when the price is less than $1/\gamma$), and there is too little production, both because nonintegration produces less than it should (intensive margin) and because there is too little integration (extensive margin).

Figure III illustrates this situation. The organizational deadweight loss ODWL (“Leibenstein trapezoid”) quantifies the welfare losses borne by both shareholders and consumers due to the inefficiently high level of nonintegration.²³ Notice that as the demand function rotates counter-clockwise around the equilibrium price to become flatter, the area of ODWL increases. More generally, a higher demand elasticity *increases* the welfare loss as measured by ODWL:

COROLLARY 1. Suppose that $\underline{u}_A = 0$. Consider two demand functions that generate the same ownership inefficient market equilibrium. If one demand is everywhere more elastic, it generates a larger organizational welfare loss.

The result that higher demand elasticity increases the organizational welfare loss stands in sharp contrast to the theory of monopoly with neoclassical firms: there, higher demand elasticities lead to *lower* welfare losses. This is one example of how organizational distortions may differ systematically from market power distortions.

23. In this case, only the extensive margin inefficiency is operative because the equilibrium price is $1/\gamma$. When the price is below that, the intensive margin effect is also operative, and the ODWL may have other shapes. It should also be apparent that increasing the amount of integration is beneficial to the planner (there is ownership inefficiency) for any elastic demand that is less than one at $P = 1/\gamma$ and greater than $3/4$ at $P = 1$, certainly a generic case.

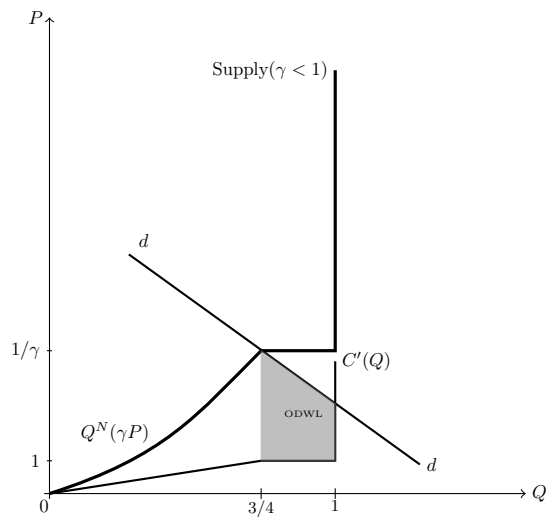


FIGURE III: OWNERSHIP INEFFICIENCY WHEN $\gamma < 1$ AND $\underline{u}_A = 0$

III.D. An Empirical Illustration: Cement and Concrete

Hortaçsu and Syverson (2007) studies the vertical integration levels and productivity of U.S. cement and ready-mixed concrete producers over several decades and many local markets. They find that these industries are fairly competitive, that demand is relatively inelastic, and that there is little evidence that vertical-foreclosure effects of integration are quantitatively important. Three of their major findings are (a) prices are lower and quantities are higher in markets with more integrated firms, (b) high productivity producers are more likely to be integrated, and (c) some of the more productive firms are nonintegrated. They tie the productivity advantages to improved logistics coordination afforded by large local concrete operations, regardless of whether it presides in integrated or nonintegrated enterprises.

Though the negative correlation between integration and prices would go against a single-productivity version of the model, shutting down productivity variation is obviously a simplification imposed to highlight the importance of

demand in determining ownership structure. The importance of supply-side factors (relationship specificity, complementarity of assets and investments, and the like) has received the bulk of the emphasis in the literature.

As shown in Section IV.B, higher productivity implies integration will occur at lower prices, so the Hortaçsu-Syverson findings (a) and (b) are explained by the presence, which is documented in the paper, of multiple productivity levels, interpreted as exogenous — markets with more integrated firms are likely to be those with more productive firms, so equilibrium prices will be lower, assuming equal demands.

However, technological variability by itself (and models which exploit this as their only source of variation in organizational form) cannot fully explain the findings either: if higher productivity causes integration, why aren't all productive firms integrated, in contradiction to finding (c)? Neither supply alone, nor demand alone would seem to be able to explain the facts of the U.S. concrete and cement industries. But together they can.

Suppose, as in Section III.B, that there is exogenous variation in the productivity of the enterprises in the industry, specifically two productivity levels, 1 and R , with $1 < R$. Demand is isoelastic, with $D(P) = \frac{3}{4}P^{-1}$, and $\underline{u}_A = 0$. Letting z be the proportion of high productivity firms, the equilibrium price at $z = 0$ is $P = 1$ and all (low productivity) firms are nonintegrated. As z increases in the interval $(0, 1)$ the equilibrium price will decrease from 1 to $1/R$.

If the price is strictly larger than $1/R$, low productivity firms choose non-integration while high productivity firms choose integration. If price is equal to $1/R$, some of the high productivity firms will be nonintegrated. Either way, high productivity firms are more likely to be integrated, as in finding (b).

As long as $P > 1/R$, the proportion of integrated firms is z , and the equi-

librium price solves:

$$zR + (1 - z)Q^N(P) = \frac{3}{4}P^{-1},$$

where the left hand side is the industry supply given the organizational choices of high and low productivity firms, and the right hand side is the demand. As z increases, the equilibrium price decreases. Thus, consistent with finding (a), price is negatively correlated with the degree of industry integration.

For z above z^* , where $z^*R + (1 - z^*)Q^N(1/R) = \frac{3}{4}R$, the equilibrium price is $1/R$. At this price managers in high productivity enterprises are indifferent between integration and nonintegration. Some high productivity firms will have to be nonintegrated in order to satisfy the market clearing condition, whereas the low-productivity firms continue to be nonintegrated. Hence, there will be heterogeneity of ownership structure among high productivity enterprises, as in finding (c). Of course, this computation assumes the coexistence of integrated and nonintegrated plants is within a local market. For coexistence across markets, it is enough to posit that they have different demands. Either way, the evidence on cement and ready-mix concrete highlights the importance of demand as well as supply for understanding the determinants of ownership and its relation to industry performance.

IV. EXTENSIONS

This section sketches some extensions of the basic model and some directions for future research.

IV.A. Entry

The model considered so far has a fixed population of price-taking suppliers, corresponding to the standard Arrow-Debreu notion of competition. Industrial Economics is also concerned with ease of entry, which in the present model would also include the choice of “side” (A or B). Since managers (particularly B s) earn rents, it is fair to ask what happens if these rents could be competed away.

Consider the following simple model of entry. There is an unlimited population of ex-ante identical potential entrants, each of whom receives zero outside the industry and can choose to become an A or a B . If n_A and n_B are the measures of people choosing A and B respectively, the cost of entry borne by an individual entrant is $e_A(n_A) = e(n_A)$, $e_B(n_B) = \beta e(n_B)$ where $e(\cdot)$ is a non-decreasing function: one side takes a larger investment than the other, unless $\beta = 1$, and resources devoted to training for either side become scarce as the industry expands.²⁴

A (long-run) equilibrium will consist of measures n_A and n_B of entrants on each side; a market clearing product price P and quantity $D(P)$; and feasible payoffs u_A and u_B for each of the entrants, according to which side they take. In equilibrium agents are indifferent among the two occupations and remaining outside the industry, and markets clear; this requires in particular that $n_A = n_B$. The scale (measure of A - B pairs) of the industry will be the common value n . In the product market, supply embodies ownership choice, as before; following Proposition 1 the ownership decisions, and therefore industry output, are determined jointly by $|u_A - u_B|$ and P . The supply of the industry

24. The baseline model is equivalent to having a fixed set of B suppliers, while entry or exit into A or HQ supply is permitted; then $\underline{u}_A = e(1)$, while as before the opportunity cost of HQ s is zero. The long run analysis also allows entry into B supply. This abstracts from issues surrounding the financing of entry, which in light of the discussion in Section IV.C below, alters little.

is $nS(|u_A - u_B|, P)$ where $S(|u_A - u_B|, P)$ is the output generated by a unit measure of enterprises, as described in Proposition 1, and n equates supply and demand.

The indifference condition implies that in equilibrium $u_B = \beta e(n)$ and $u_A = e(n)$. If $e(\cdot)$ is strictly increasing, there is a payoff expansion path $u_B = \beta u_A$. At low enough prices, this path lies in the nonintegration region $|u_A - u_B| < \frac{P}{1+P}$. As price increases, the measure n of A s and B s, also increases, as do the payoffs. Unless $\beta = 1$, eventually the payoff expansion path crosses the $|u_A - u_B| = \frac{P}{1+P}$ loci at a price $\hat{P}^*(\beta)$ and there is a switch to integration. Increases in β starting above 1 (or decreases starting below 1) imply more inequality in the distribution of surplus, which in turn increase the likelihood of integration (see figure IV). This echoes the discussion following Proposition 1.

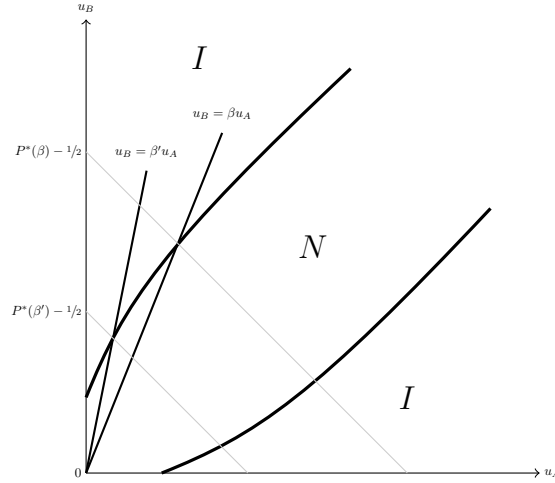


FIGURE IV: PAYOFF EXPANSION PATH FOR $\beta' > \beta$

As in the standard neoclassical case, because of entry, the long run supply will be more price elastic than the short run supply. Indeed, each enterprise's total managerial surplus must be equal to the total cost of entry $(1 + \beta)e(n)$; since surplus is increasing in P , the number of enterprises also increases with P .

Thus the price response of each enterprise is augmented by an increase in their number, leading to a more elastic supply than in the short run (in particular, the integration portion of the long run supply is no longer vertical).

Where managers are indifferent between integration and nonintegration, the quantity supplied will depend on the fraction of integrated enterprises, but their number is fixed. Thus, as in the short run, the supply curve has a horizontal segment at $\hat{P}^*(\beta)$, and there is generic heterogeneity of organizational forms in a free-entry equilibrium.

In the special case where $e(n)$ is a constant, the payoff expansion path is degenerate; the payoffs u_A, u_B are independent of demand. For almost all values of the cost of entry, the vector of equilibrium payoffs will lie either in the non-integration or integration region, making heterogeneity of organizational forms a nongeneric outcome. This is similar to the Grossman and Helpman (2002) finding.

Finally, the short run welfare results extend to the long run; the equilibrium will be ownership efficient if $\gamma = 1$ but not in general otherwise.

IV.B. Scale and Scope

So far, enterprises produce at most one unit of output, but it is not difficult to extend the model to allow for variation in enterprise scale, employee quality or other contractible input level. Suppose that enterprises may increase the volume of successful production by a multiplicative factor $f(l)$, a function of some contractible input l that represents quantity or quality, at cost represented by the schedule $w(l)$. Then expected profit is $(1 - (a - b)^2)f(l)P - w(l)$. For simplicity assume that the private managerial cost is unaffected by l . The input is paid contingently on success: the amount paid, $d(l)$, satisfies $(1 - (a - b)^2)d(l) = w(l)$, with a, b assuming their equilibrium values (equivalently, a loan

can be procured ex-ante to pay the input and the loan is repaid upon success). When enterprises form, managers contract on shares, ownership structure, and scale; then noncontractible decisions a, b are chosen last.

As before, the success probability under integration will be 1; assuming differentiability of $f(\cdot)$ and $w(\cdot)$, the integration scale l^I will satisfy:

$$Pf'(l^I) = w'(l^I).$$

For the nonintegrated enterprise, a straightforward computation shows that its scale l^N satisfies:

$$QPf'(l^N) = w'(l^N),$$

where $Q = Q^N(Pf(l^N) - d(l^N))$. Thus marginal returns to l are greater under integration. From Topkis's theorem, $l^I > l^N$. And as before, there will be a price at which managers are indifferent between integration and nonintegration.

Hence, when they coexist, integrated firms are bigger or “smarter” (in the alternative interpretation of l as quality) than nonintegrated ones.²⁵ Therefore the output gap between the two kinds of enterprises is amplified. Whether these endogenous output differences could explain the large productivity differences that observed in many industries is left for further research. Some indication of its plausibility is offered by Hortaçsu and Syverson (2007), which reports that integrated cement and concrete producers are larger and more capital intensive, as well as more productive, than nonintegrated ones.

Another extension of the model involves the analysis of firm scope, wherein each supplier produces its own distinct good for sale on the market, as in the original Hart-Holmström model. Because it involves strategic interactions

25. This result persists if l represents a vector of complementary inputs, or if managerial costs are a function of the scale, as long as these costs do not grow too quickly with respect to l . Details available upon request.

among different parts of the enterprise, this case is somewhat intricate, particularly if the number of suppliers is greater than two. Legros and Newman (2010) considers such a model. As here, integration yields coordination but at a high cost for the managers of individual product lines. Moreover, an integrated firm provides coordination benefits not only to its members, but also for “outsiders,” who can free ride. The integrated firm is typically tempted to follow the preferences of the outsiders, who therefore enjoy coordination benefits at lower cost than insiders. This mechanism provides a limit to the extent of integration, and the paper shows that the scope of multi-product firms depends on both the mean and the variance of individual product prices.

IV.C. Finance

The responsiveness of ownership choice to industrial conditions in the model emerges in large measure because managers have finite (in fact, zero) wealth. This assumption implies that the ownership structure is dependent on the distribution of surplus between them, and that in turn leads to the relationship between industry price and degree of integration embedded in the OAS. It is worth devoting some attention to the comparative statics of the wealth endowments, as this appears to be connected more generally to the role that financial contracting would have in affecting industry performance.

One important difference between integration and nonintegration is the degree of transferability in managerial surplus: while managerial welfare can be transferred 1 to 1 with integration (that is one more unit of surplus given to B costs one unit of surplus to A) by adjusting the revenue shares, this is not true of nonintegration. If the A manager has cash that can be transferred without loss to the B manager before production takes place, the advantage of integration in terms of transferability is reduced. This has immediate consequences for the

OAS.

PROPOSITION 5. Positive increases in the cash endowments of A suppliers shift the OAS to the left and decrease the level of integration.

In particular, if managers have access to infinite amounts of cash, they will always contract for equal sharing in nonintegrated organizations. However, for any finite cash holdings, there is a level of price high enough for which there will be integration.

These observations apply with equal force when the managers are not full residual claimants on the revenues of the firm. Since integration is output maximizing, giving managers cash increases the inefficiencies from the point of view of consumers and shareholders. Thus, in contrast to previous literature that has suggested that managerial cash holdings may lead to firms that are too large, this model suggests it may generate firms that are too small.²⁶

Borrowing can sometimes be a substitute for cash, but its efficacy depends on the ownership structure. In general, if the income stream one is borrowing against is subject to incentive problems, then borrowing is a poor substitute; if not then it is at best equivalent. Thus, under integration, a cash-poor HQ can finance the acquisition with debt without affecting output, and in fact may borrow essentially up to the value of the enterprise P , provided she puts up at least some cash. The managers could also use debt for side payments, though nothing is gained (or lost) by this: re-dividing the asset purchases or the revenue would accomplish the same thing (recall the integration Pareto frontier is the off diagonal).

By contrast, under nonintegration, borrowing for side payments is more limited. If a debt D has to be repaid when the output is high, it is as if the

²⁶ As in Jensen (1986) managers will tend to use cash for their own interest rather than that of shareholders or consumers, but while in the free cash flow theory they will tend to use cash for “empire building” or over-expansion, here they use cash to keep firms small and buy a “quiet life”.

managers face a success revenue of $P-D$, and expected output will decrease from $Q^N(P)$ to $Q^N(P-D)$; this leads to a decline in ex-post total managerial welfare that has to be weighed against the ex-ante lump-sum payment. In the appendix it is shown that there is no gain to using such contracts when the managerial surplus allocation is either very unequal (along the axes) or approximately equal (in a neighborhood of the 45°-line). Thus the debt-enhanced nonintegration frontier “adheres” to the no-debt frontier in those regions; in particular, debt for side payments does not approximate cash transfers (for instance, cash is Pareto improving near the 45°-line, significantly so near the axes). In between the axes and the 45°-line, small amounts of debt for side-payments may raise managerial welfare, enlarging slightly the set of prices for which nonintegration is chosen. Qualitatively, the responses of the integration decision to prices and surplus division go much as before, but the possibility of borrowing for side payments not only may reduce the output performance of nonintegrated enterprises, but also may increase their number. If the managers are not full revenue claimants, debt may therefore lower total welfare.

PROPOSITION 6. Integrated firms (HQ) may use debt to finance the purchase of assets. nonintegrated enterprises do not use debt for ex-ante side-payments if As receive either zero surplus or share surplus equally with Bs .

Note that debt is used here only to finance ex-ante side payments and not acquisition of assets from outside the relationship. If firms have investment opportunities, e.g., to expand their scale of production, nonintegrated firms are at a disadvantage for borrowing since debt has a depressing effect on expected output.

IV.D. Active Shareholders

The model assumes managers have control over the integration decisions and ownership inefficiencies take the form of too little integration. Since shareholders have similar objectives to consumers — as price takers, they value high output — a reasonable conjecture is that efficiency may be improved when shareholders have control over integration decisions and compensation to managers. Corporate governance is now relevant not only for shareholders but also for consumers.

Providing good incentives is costly to shareholders under nonintegration, since it entails giving up a share of profit to management. Hence, if shareholders have control over the share of revenue accruing to managers, nonintegration will underperform relative to how it does when managers retain all the revenue. By contrast, integration performs well at all managerial shares. Shareholders will therefore tend to integrate “too often.” Thus, while shareholder control may implement a planner’s forced integration as a means of protecting consumers, it can be taken too far.

PROPOSITION 7. Suppose that the shareholders control the decision to integrate and the shares of output going to the two managers. The minimum price at which shareholders choose integration is strictly lower than the ownership efficient level.

The policy question of how to regulate governance when consumers are affected by firm’s organizational choices is still open and an important subject for future research.

V. CONCLUSION

By embedding a simple model of the determination of firm boundaries into a perfectly competitive setting, this paper suggests that internal organizational effects can have significant consequences at the industry level, quite apart from the strategic and collusive interactions among firms that have dominated the Industrial Economics literature. The model's organizational industry equilibrium offers an explanation for heterogeneity in firm performance and shows that the response of the industry to demand or technology shocks is dependent on integration decisions. At the same time, it illustrates how market forces can be important determinants of organizational design.

This model of industry equilibrium is based on one important tradeoff between nonintegration and integration: nonintegration is strong on internalizing private costs and weak on coordinating decisions; integration has the opposite strength and weakness. And though there is evidence of the model's relevance for the cement and airlines industries, an organizational Industrial Economics is hardly restricted to this tradeoff. Other underlying models of integration or other aspects of organizational design may generate other relationships among prices, quantities and organization: the OAS may have different shapes for different theories. One benefit of unifying Organizational and Industrial Economics is that market data become a proving ground for different organizational tradeoffs, which can help to identify the relative importance of each in different contexts.

Although the model goes some way toward the objective of jointly determining market structure, ownership structure, and industry conduct and performance, much remains to be done. In particular, rather than assuming perfect competition, environments where market power would play a role could also be considered, so that the degree of competition would be endogenous.

Finally, on the welfare side, the analysis also shows that consumers have an interest in the organization of the firms that make the products they buy. An influential strand of thought asserts that competition in the product market must assure efficient outcomes: firms that do not deliver the goods at the lowest feasible cost, whatever the reason, including inefficient organization, will be supplanted by ones that do.²⁷ The present framework confronts such claims directly. While efficient ownership may prevail when managers have full title to enterprise revenue, in the general case, organizational choices may not be efficient. Much as Leibenstein argued, and as quantified by the eponymous trapezoid, the welfare losses from this kind of organizational inefficiency may be significant.

APPENDIX: PROOFS

A. Proof of Proposition 3

Suppose that demand is perfectly elastic. With integration, B gets $P - \frac{1}{2} - \underline{u}_A$, while with nonintegration he gets $u_B^N(s(P; \underline{u}_A), P) = U^N(s(P; \underline{u}_A), P) - \underline{u}_A$, where $s(P; \underline{u}_A)$ is the sharing rule solving $u_A^N(s, P) = \underline{u}_A$, with $u_A^N(s, P)$ defined in (3).

It follows that the B manager chooses the organization that maximizes the total managerial welfare. Since with perfectly elastic demand, the total welfare coincides with the total managerial welfare, the result follows.

With a less elastic demand, suppose that at the market equilibrium some firms are not integrated. If any firms are forced to integrate, industry supply increases and therefore the new equilibrium will be at $P' < P$ since demand is decreasing. However, the difference in total welfare between forced

27. The “form of organization that survives... is the one that delivers the product demanded by customers at the lowest price while covering costs.” (Fama and Jensen, 1983).

integration at price P' and nonintegration at price P is bounded above by $U^I(P) - U^N(s(P; \underline{u}_A), P)$, which is negative by the previous argument. Hence the initial equilibrium is ownership efficient.

Finally, if in equilibrium all firms are integrated, forcing some firms not to integrate will decrease consumer surplus (since the quantity supplied decreases) and decrease managerial surplus (since managers prefer integration), so equilibrium is efficient in this case as well.

B. Proof of Corollary 1

Consider demand functions D_L and D_H yielding the same ownership inefficient equilibrium with price P^e and proportion $\alpha^e \geq 0$ of integrated firms (thus $D_H(P^e) = D_L(P^e)$), and suppose D_H has larger elasticity than D_L at all prices. Let $\Omega_i(P)$ be the consumer welfare corresponding to the demand $D_i, i = H, L$.

Fix $\gamma < 1$ and write the sums of shareholder and managerial payoffs under integration and nonintegration respectively as

$$V^I(P) = P - \frac{1}{2}$$

$$V^N(P) = PQ^N(\gamma P) - \phi(\gamma P).$$

A planner can choose to force a proportion α firms to integrate in order to maximize total welfare when demand is D_i :

$$\max_{\alpha} \Omega_i(P) + \alpha V^I(P) + (1 - \alpha) V^N(P)$$

$$\text{s.t. } D_i(P) = \alpha + (1 - \alpha) Q^N(\gamma P).$$

The constraint defines a one-to-one relationship between α and P and $\rho_i(\alpha)$ is the price solving the constraint for a given value of α . The problem is then

equivalent to choosing α to solve:

$$(6) \quad \max_{\alpha} \Omega_i(\rho_i(\alpha)) + \alpha V^I(\rho_i(\alpha)) + (1 - \alpha)V^N(\rho_i(\alpha)).$$

Clearly, for any α , $\rho_L(\alpha) < \rho_H(\alpha) < P^e$.

Fact (i) If D_H has larger elasticity than D_L at any price, the two demand curves intersect only once at P^e and $D_H(P) > D_L(P)$ for any $P < P^e$.

(ii) For all $P < P^e$, $\Omega_H(P) - \Omega_H(P^e) > \Omega_L(P) - \Omega_L(P^e)$.

(iii) If $P \leq P' < P^e$ and $D_H(P') = 1$, $\Omega_H(P') - \Omega_H(P^e) + P' - P > \Omega_L(P) - \Omega_L(P^e)$.

Proof. (i) At P^e , since $D_H(P^e) = D_L(P^e)$, D_H is flatter than D_L . If single crossing fails, there is a price less than P^e where the demand curve D_H crosses D_L from above, hence D_H is steeper than D_L at P' , contradicting the assumption that D_H has larger elasticity than D_L at any price. (ii) Now, for $P < P^e$, D_H is above D_L , implying that the incremental consumer welfare is larger for D_H than D_L , as claimed. (iii) By (ii), $\Omega_H(P') - \Omega_H(P^e) > \Omega_L(P') - \Omega_L(P^e)$. It is clear that $\Omega_L(P) - \Omega_L(P') < P' - P$; substituting $\Omega_L(P') > \Omega_L(P) - (P' - P)$ in the previous inequality yields the result. ■

Case 1. Suppose first that the solution of (6) is obtained at $\alpha_L = 1$, leading to an equilibrium price of $P_L = \rho_L(1)$. Then total welfare gain is $G_L(1) = \Omega_L(P_L) - \Omega_L(P^e) + V^I(P_L) - V^N(P^e)$. If the planner imposes $\alpha = 1$ with demand D_H (this may not be optimal), the price will be $P_H = \rho_H(1)$; by the Fact (i), $P_H > P_L$ and the gain in welfare is $G_H(1) = \Omega_H(P_H) - \Omega_H(P^e) + V^I(P_H) - \alpha^e V^I(P^e) - (1 - \alpha^e)V^N(P^e)$. Therefore:

$$G_H(1) - G_L(1) = [\Omega_H(P_H) - \Omega_H(P^e)] - [\Omega_L(P_L) - \Omega_L(P^e)] + P_H - P_L$$

which is positive by Fact (iii). Because the maximum welfare loss with D_H is at least $G_H(1)$, the result follows. Note that this case does not use the fact that $\underline{u}_A = 0$.

Case 2. Suppose now that the solution to (6) is obtained at an interior point $\alpha_L \in (\alpha_e, 1)$, and let $P_L = \rho_L(\alpha_L)$. Then:

$$(7) \quad V^I(P_L) - V^N(P_L) > 0.$$

To see this, use $\Omega'_L(P_L) = -D_L(P_L)$ to obtain the variation of the planner's objective function with respect to α at the optimum as:

$$\frac{d\rho_L(\alpha_L)}{d\alpha} \frac{dQ^N(\gamma P_L)}{dP} (1 - \alpha)\gamma [P_L - c'(Q^N(\gamma P_L))] + V^I(P_L) - V^N(P_L).$$

Because $\frac{d\rho_L(\alpha_L)}{d\alpha} < 0$ and $P - c'(Q^N(\gamma P_L)) = (1 - \gamma)P_L > 0$ (recall from the text that the nonintegrated enterprise sets $\pi = c'(q)$), α_L solves the first order condition only if $V^I(P_L) - V^N(P_L) > 0$, as claimed.

Case 2a. Suppose there is $\alpha \in (0, 1)$ such that $D_H(P_L) = \alpha + (1 - \alpha)Q^N(\gamma P_L)$. Then $\alpha > \alpha_L$, since $D_H(P_L) > D_L(P_L)$. The total gains in surplus when α firms integrate under demand function D_H and α_L firms integrate under demand D_L are therefore:

$$\begin{aligned} G_H(\alpha) &= \Omega_H(P_L) - \Omega_H(P^e) + V^N(P_L) + \alpha[V^I(P_L) - V^N(P_L)] \\ &\quad - \alpha^e V^I(P^e) - (1 - \alpha^e)V^N(P^e), \\ G_L(\alpha_L) &= \Omega_L(P_L) - \Omega_L(P^e) + V^N(P_L) + \alpha_L[V^I(P_L) - V^N(P_L)] \\ &\quad - \alpha^e V^I(P^e) - (1 - \alpha^e)V^N(P^e), \end{aligned}$$

implying:

$$\begin{aligned} G_H(P_L) - G_L(P_L) &= \Omega_H(P_L) - \Omega_H(P^e) - [\Omega_L(P_L) - \Omega_L(P^e)] \\ &\quad + (\alpha - \alpha_L)[V^I(P_L) - V^N(P_L)] \end{aligned}$$

which is positive by Fact (ii), (7), and $\alpha > \alpha_L$. Because the maximum gain under D_H is greater than $G_H(P_L)$ while $G_L(P_L)$ is the maximum gain under D_L , the result follows.

Case 2b. The last possibility is that there does not exist $\alpha \in (0, 1)$ such that $D_H(P_L) = \alpha + (1 - \alpha)Q^N(\gamma P_L)$. Then $D_H(P_L) > 1$. Let $P_H = \rho(1)$. If under D_H all firms are forced to integrate the total surplus is $\Omega_H(P_H) + V^I(P_H)$. Because $V^I(P_L) > V^N(P_L)$, $V^I(P_H) - [\alpha_L V^I(P_L) + (1 - \alpha_L)V^N(P_L)] > V^I(P_H) - V^I(P_L) = P_H - P_L$. Hence:

$$\begin{aligned} G_H(P_H) - G_L(P_L) &= [\Omega_H(P_H) - \Omega_H(P^e)] - [\Omega_L(P_L) - \Omega_L(P^e)] \\ &\quad + V^I(P_H) - \alpha_L V^I(P_L) - (1 - \alpha_L)V^N(P_L) \\ &> [\Omega_H(P_H) - \Omega_H(P^e)] - [\Omega_L(P_L) - \Omega_L(P^e)] + P_H - P_L \\ &> 0. \end{aligned}$$

where the last inequality follows Fact (iii).

C. Proof of Proposition 5

Under nonintegration, cash is a more efficient instrument for surplus allocation than the sharing rule s between managers A and B since a change of s affects total costs. By contrast, when firms are integrated, a change in s has no effect on output or costs and therefore shares are as efficient at allocating surplus as cash. Hence, introducing managerial cash endowments favors nonintegration,

and there will be fewer integrated firms.

Return to the baseline environment where the population of B s is fixed and let $\underline{u}_A = 0$. Assume for simplicity that all A managers have the same cash holding ℓ . A contract now specifies a lump sum transfer t from the A manager to the B manager together with a share s of output and an organization. Because the managerial surplus is transferable under integration, cash transfers are neutral. For nonintegration however, since the total surplus is increasing on $s \in [0, 1/2]$, the ability for the A managers to make cash transfers in exchange for a greater share of output will increase managerial surplus. Final payoffs are then $u_A^N(s, P) - t$ and $u_B^N(s, P) + t$, where u_A^N, u_B^N are given by (3).

By competition, the marginal manager A must have zero surplus. For a given market price P , if $\ell > U^N(1/2, P)$, A managers are able to transfer $t = U^N(1/2, P)/2$ to the B manager at the signing of the contract in exchange for a share of $1/2$ of the output. In this case, the payoff to the B managers is $U^N(1/2, P)$. If $\ell < U^N(1/2, P)$, the A manager uses a transfer $t = \ell$ in exchange for a share $s_\ell(P)$ of the revenue satisfying:

$$u_A^N(s_\ell(P), P) = \ell.$$

In this case, managers B have, after the transfer of $t = \ell$, a payoff of:

$$u_B^N(s_\ell(P), P) + \ell = U^N(s_\ell(P), P).$$

Recall that $U^N(s_\ell(P), P) = PQ^N(P) - \psi(s_\ell(P))c(Q^N(P))$ with $\psi(s) \equiv s^2 + (1 - s)^2$. Since $\psi(s)$ is a decreasing function of $s \in [0, 1/2]$, and since as ℓ increases, $s_\ell(P)$ increases, total managerial welfare increases in ℓ . By contrast, managerial welfare under integration is $U^I(P) = P - 1/2$ which is independent of ℓ . Therefore the shift to integration happens at higher prices.

D. Proof of Proposition 6

Debt and nonintegration. Suppose that A makes a side payment T to B that is financed with a debt. Specifically, an amount D is repaid in case of success, and by limited liability, there is no repayment in case of failure. Then A 's revenue share s must satisfy $sP \geq D$, and the payoffs to the two managers are respectively $u_A = (1-(a-b)^2)(sP-D)-(1-a)^2$ and $u_B = (1-(a-b)^2)(1-s)P-b^2+T$. By making the change of variable $\hat{s} = \frac{sP-D}{P-D}$, the payoffs can be written as $u_A = (1-(a-b)^2)\hat{s}(P-D)-(1-a)^2$ and $u_B = (1-(a-b)^2)(1-\hat{s})(P-D)-b^2+T$. The game in (a, b) played by the managers is then equivalent to the baseline model of nonintegration with a price equal to $P-D$. The probability of success is $Q^N(P-D)$ and therefore in equilibrium $T = Q^N(P-D)D$. Since $Q^N(P-D)$ is declining in D , consumers are harmed if nonintegrated enterprises borrow for side payments. Moreover, the total managerial surplus is $Q^N(P-D)P - \psi(\hat{s}) \left(\frac{P-D}{1+P-D} \right)^2$, which is decreasing in D for any sized value of \hat{s} .

Now, if $u_A = 0$, then $\hat{s} = 0$, and the optimal choice of D maximizes B 's payoff, which can be written $Q^N(P-D)P - \left(\frac{P-D}{1+P-D} \right)^2$. Maximizing this over D yields as the optimum $D = 0$. Hence, manager A will not use debt to make lump-sum payments to manager B , as claimed.

In general, starting with a contract with share s_0 and no debt, suppose there is a contract with debt $D > 0$ that yields the same payoff for A ; then there is $\hat{s}(D)$ satisfying $\hat{s}(D)(P-D)Q^N(P-D) - \hat{s}(D)^2 \left(\frac{P-D}{1+P-D} \right)^2 = s_0 PQ^N(P) - s_0^2 \left(\frac{P}{1+P} \right)^2$. The total surplus under the debt contract is then $PQ^N(P-D) - \psi(\hat{s}(D)) \left(\frac{P-D}{1+P-D} \right)^2$. Whether B benefits from debt for side payments can now be assessed by comparing the total surplus without debt $PQ^N(P) - \psi(s_0) \left(\frac{P}{1+P} \right)^2$ to the total surplus with debt $PQ^N(P-D) - \psi(\hat{s}(D)) \left(\frac{P-D}{1+P-D} \right)^2$. If $s_0 = 1/2$, $\psi(\hat{s}(D))$ is larger than $\psi(s_0)$ (since $\psi(\cdot)$ is minimized at $1/2$), and $PQ^N(P) - \psi(\frac{1}{2}) \left(\frac{P}{1+P} \right)^2 > PQ^N(P) - \psi(\hat{s}(D)) \left(\frac{P}{1+P} \right)^2 > PQ^N(P-D) - \psi(\hat{s}(D)) \left(\frac{P-D}{1+P-D} \right)^2$.

The surplus without debt strictly exceeds that with debt. Thus debt will not be used under nonintegration in a neighborhood of the 45°-line, as claimed in the text.

Debt and Integration. Unlike the nonintegration case, when there is integration, debt may have substantial value. From Section II.C, HQ must have a strictly positive share s_H of the firm's revenue. The maximum ex-ante transfer she is then willing to make to the managers is $s_H P$. If her cash endowment is positive (however small), she can always finance the rest of the asset purchase with debt, since debt has no distortionary effects on her incentives. Indeed, if she owes D , her interim payoff is $s_H(1 - (a - b)^2)(P - D)$, which still results in $a = b = 1/2$, provided $D < P$.

If however she is fully leveraged, with $D = P$, then neither she nor the managers have a positive stake in the revenue, and ex-post they would all agree to implement $(a, b) = (1, 0)$, resulting in zero expected revenue, in order to minimize costs. Anticipating this, a lender would refuse a loan. It follows that HQ s must finance their acquisitions partly with cash.

E. Proof of Proposition 7

Under integration, shareholders obtain $P - 1/2$ when managers have a zero opportunity cost. Under nonintegration, shareholders have an expected revenue $Q^N(\pi)(P - \pi)$ if they give a total compensation π to the managers. Let $\pi(P)$ be the optimal solution to this problem and $V(P)$ be the value at the optimum. Because the objective function has positive cross partials in (π, P) , $\pi(P)$ is an increasing function of P . By the envelope theorem, $V'(P) = Q^N(\pi(P))$ is positive and since $\pi(P)$ is increasing in P , $V(P)$ is a convex and increasing function of P . Note that $\lim_{P \rightarrow \infty} V'(P) = 1$; hence there is a unique solution to $V(P) = P - 1/2$ at a unique value. Moreover, this solution is less than 1 because

at $P = 1$, $V(1) < 1/2$. Because $\pi(1)$ solves the equation $(1+\pi)^3 = 3-\pi$, we have $\pi(1) \approx 0.38$, and therefore $V(1) \approx 0.29$ which is less than $1/2$ as claimed. Since the minimum price at which it is efficient to switch to integration is $P = 1$ (this follows from Proposition 3 and the fact that managers never choose integration below this price), there is “too much integration” when shareholders have full control.

REFERENCES

- Aghion, Philippe, Rachel Griffith, and Peter Howitt, “Vertical Integration and Competition,” *American Economic Review Papers and Proceedings*, 96 (2006), 97-102.
- Aghion, Philippe and Jean Tirole, “The Management of Innovation,” *The Quarterly Journal of Economics*, 109 (1994), 1185-1209.
- Alonso, Roberto, Dessein, Wouter and Matouschek, Niko: “Centralization Versus Decentralization: An Application to Price Setting By a Multi-Market Firm,” *Journal of the European Economic Association*, 6 (2008), 457-67.
- Antras, Paul and Helpman, Elhanan, “Global Sourcing,” *Journal of Political Economy*, 112 (2004), 552-80.
- Antras, Paul, “Firms, Contacts, and Trade Structure,” *Quarterly Journal of Economics*, 118 (2003), 1357-1418.
- Banerjee, Abhijit V., and Andrew F. Newman, “Occupational Choice and the Process of Development,” *Journal of Political Economy*, 101 (1993), 274-298.
- Bertrand, Marianne and Sendhil Mullainathan, “Enjoying the Quiet Life? Corporate Governance and Managerial Preferences,” *Journal of Political Economy*, 111 (2003), 1043-1075.
- Brander, James A., and Tracy R. Lewis, “Oligopoly and Financial Structure: The Limited Liability Effect.” *The American Economic Review*, 76 (1986), 956-970.
- Fama, Eugene F., and Michael C. Jensen, “Separation of Ownership and Control,” *Journal of Law and Economics*, 26 (1983), 301-325.
- Fershtman, Chaim, and Kenneth L Judd, “Equilibrium Incentives in Oligopoly,” *American Economic Review* 77 (1987),: 927-940.
- Forbes, Silke J., and Mara Lederman, “Adaptation and Vertical Integration in the Airline Industry,” *American Economic Review* 99 (2009), 1831-1849.
- , “Does vertical integration affect firm performance? evidence from the airline industry,” *The RAND Journal of Economics* 41 (2010), 765-790.
- Gibbons, Robert, “What the Folk Theorem doesn’t tell us,” *Industrial and Corporate Change*, 15 (2006), 381-386.

- , “Inside Organizations: Pricing, Politics, and Path Dependence,” *Annual Review of Economics*, 2 (2010), 337-365.
- Gibbons, Robert, Richard Holden and Michael Powell, “Organization and Information: Firm’s Governance Choices in Rational Expectations Equilibrium,” *Quarterly Journal of Economics*, 127 (2012), 1813-1841.
- Grossman, Sanford and Oliver Hart, “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral integration,” *Journal of Political Economy*, 94 (1986), 691-719.
- Grossman, Gene M. and Helpman, Elhanan, “Integration Versus Outsourcing in Industry Equilibrium,” *The Quarterly Journal of Economics*, 117 (2002), 85-120.
- Hart, Oliver, “The Market Mechanism as an Incentive Scheme” *Bell Journal of Economics*, 14 (1983), 366-382.
- Hart, Oliver and Bengt Holmström, “A Theory of Firm Scope,” *The Quarterly Journal of Economics*, 125 (2010), 483-513.
- Hart, Oliver and John Moore, “Property Rights and the Nature of the Firm,” *Journal of Political Economy*, 98 (1990), 1119-1158.
- Hermalin, Benjamin E., “Heterogeneity in Organizational Form: Why Otherwise Identical Firms Choose Different Incentives for Their Managers?” *The RAND Journal of Economics*, 25 (1992), 518-34.
- Hortaçsu, Ali, and Chad Syverson, “Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices,” *Journal of Political Economy*, 113 (2007), 250-301.
- Jensen, Michael C., “Agency Cost of Free Cash Flow, Corporate Finance, and Takeovers,” *American Economic Review*, 76 (1986), 323-329.
- Legros, Patrick and Andrew F. Newman, “Wealth Effects, Distribution and the Theory of Organization” *Journal of Economic Theory*, 70 (1996), 312-41.
- , “Competing for Ownership,” *Journal of the European Economic Association*, 6 (2008), 1279-1308.
- , “A Price Theory of Firm Scope,” mimeo, ECARES, 2010.
- , “Contracts, Ownership, and Industrial Organization: Past and Future,” mimeo ECARES and BU, 2013.
- , “On the Impossibility of Disinterested Authority,” mimeo ECARES and BU, 2012.
- Leibenstein, Harvey, “Allocative Efficiency Vs. ‘X-efficiency’,” *American Economic Review*, 56 (1966), 392-415.
- Machlup, Fritz, “Theories of the Firm: Marginalist, Managerial, Behavioral,” *American Economic Review*, 57 (1967), 1-33.
- Marin, Dalia and Verdier, Thierry, “Power Inside the Firm and the Market: A General Equilibrium Approach,” *Journal of the European Economic Association*, 6 (2008), 752-88.
- Mclaren, John, “Globalization and Vertical Structure,” *The American Economic Review*, 90(2000), 1239-54.
- Raith, Michael, “Competition, Risk, and Managerial Incentives,” *The American Economic Review* 93 (2003), 1425-1436.

- Schmidt, Klaus, "Managerial Incentives and Product Market Competition", *Review of Economic Studies* 64 (1997), 191-214.
- Scharfstein, David, "Product-Market Competition and Managerial *The RAND Journal of Economics*" 19 (1988), 147-155.
- Syverson, Chad, "What determines productivity?" *Journal of Economic Literature*, 49 (2011), 326-365.
- van den Steen, Eric, " Organizational beliefs and managerial vision," *Journal of Law, Economics, and Organization*, 21 (2005), 256-283.
- Whinston, Michael D., "Assessing the Property Rights and Transaction-Cost Theories of Firm Scope," *The American Economic Review*, Papers and Proceedings of the Hundred Thirteenth Annual Meeting of the American Economic Association, 91 (2001), 184-188.