

VERY PRELIMINARY AND INCOMPLETE.  
Please do not cite without permission.

## Vertical Integration and Production: Some Plant-Level Evidence<sup>\*</sup>

Ali Hortaçsu  
University of Chicago  
and NBER  
[hortacsu@uchicago.edu](mailto:hortacsu@uchicago.edu)

and

Chad Syverson  
University of Chicago  
and NBER  
[syverson@uchicago.edu](mailto:syverson@uchicago.edu)

December 2006

---

<sup>\*</sup> Very preliminary and incomplete; please do not cite without contacting the authors. We thank Daron Acemoglu, Austan Goolsbee, and Tom Holmes for helpful discussions and seminar participants at the Chicago Micro Lunch for their comments. Margaret Triyana provided excellent research assistance. Syverson appreciates funding from the NSF (award no. SES-0519062), the John M. Olin Foundation, and the Stigler Center. The research in this paper was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the Chicago Census Research Data Center. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau. This paper has been screened to insure that no confidential data are revealed. Support for this research at the Chicago RDC from NSF (awards no. SES-0004335 and ITR-0427889) is also gratefully acknowledged. Both authors can be contacted at the Department of Economics, University of Chicago, 1126 E. 59<sup>th</sup> Street, Chicago, IL 60637.

## **Abstract**

Despite there being a substantial literature exploring the causes and effects of vertical integration, little systematic plant-level evidence exists on the ways production in vertically integrated firms differs from that of unintegrated producers in the same industries. This paper uses a very large dataset on plants' production activities and the organizational structures of their owning firms to show how being part of a vertically integrated structure is reflected in plant and firm attributes. We begin by documenting a novel and ubiquitous fact about vertical integration: large plants (relative to others in their industry) are substantially more likely to be vertically integrated. Despite the ubiquity of this pattern across industries, however, standard models of vertical integration have no explicit role for the scale of operations. We go on to show that integration is more generally related to within-industry differences in plant "type," like productivity measures and capital intensity. We find that these differences mostly embody persistent differences in the plants that are started by (or brought into) firms with vertical structures, but to some extent also reflect changes to formerly unintegrated plants that become integrated. Matching of plant types appears to be important in vertically structure firms: the correlation in these plant-level "type" measures is stronger in firms with vertical structures than those that are purely horizontally structured. We also find evidence that plants with more complex production technologies (at least to the extent that this can be measured using our simple proxies) are more likely to be integrated. We propose a model, based on complementarities of inputs in production, to explain these patterns.

## I. Introduction

There is a considerable literature exploring the causes and effects of vertical integration. This literature has pointed to a large variety of factors being importantly related to integration, including market power, asset specificity, supply uncertainty, incomplete contracting, transaction costs, and regulation.<sup>1</sup> Despite this, there is little systematic plant-level evidence on the ways production in vertically integrated firms differs from that of unintegrated producers in the same (narrowly defined) industries. This applies not only to differences in hard-to-measure attributes like capital specificity, contracting environments, and transaction costs, but even to more basic features like plant scale, factor intensity, and productivity.

In this paper, we hope to begin to fill this gap. We use data on millions of plants—and the organizational structure of firms that own them—to show how being part of a vertically integrated structure is reflected in plant and firm attributes. We show that vertical integration is strongly related to within-industry differences in plant “type,” as reflected in size, capital-labor ratios, and productivity measures. We also demonstrate that these differences mostly embody persistent differences in the plants that are started by (or brought into) firms with vertical structures, but to some extent also reflect changes to formerly unintegrated plants that become integrated. We document that the correlation in these plant-level “type” measures is stronger in firms with vertical structures than those that are purely horizontally structured. Finally, there is evidence that plants with more complex production technologies (at least to the extent that this can be measured using our simple proxies) are more likely to be integrated.

We formulate a theory to explain these patterns. Because they hold across such a broad array of industries—indeed, throughout the entire manufacturing sector—they cannot result from industry-specific strategic settings, legal environments, technologies, or product markets. Instead, we propose that these patterns highlight the importance of a general *attribute* of technology that is typically thought to be commonplace and is empirically well founded:

---

<sup>1</sup> The size of the literature precludes comprehensive citation. Surveys include Perry (1989), Salop (1998), and Joskow (2005). Coase (1937) is often cited as the seminal paper of the literature. Much of the recent industrial organization research on integration has focused on foreclosure (market power) implications. Examples of recent theoretical and empirical work with broader views of the determinants of integration within and across industries include Antras (2003), Acemoglu, Aghion, Griffith, and Zilibotti (2004) and Acemoglu, Johnson, and Mitton (2005).

complementarities among inputs.<sup>2</sup> We describe the theory in more detail below, but its central idea can be put simply. If inputs are complements, the profit gain from vertically integrating to reduce input market distortions (e.g., hold-up due to specific investments, double marginalization, or supply uncertainties) grows in a plant’s “type” within its industry.<sup>3</sup> As a result, we are more likely to see plants with high type (which is empirically connected to observables like plant scale, capital intensity, and productivity levels) in integrated structures.

An illustrative case of the connection between within-industry plant type differences and the propensity for plants to be integrated regards plant scale and vertical integration. We are aware of no theory in the large theoretical literature discussed above with any explicit predictions about the relationship between vertical integration and the scale of production operations. Yet as we will shortly demonstrate, plant scale (measured by the total value of its product shipments relative to other plants in its industry) is clearly related to the likelihood that a plant is vertically integrated, and this relation is quantitatively very important.<sup>4</sup>

After documenting the scale-integration relationship in the next section, we broaden our investigation to look at how other type measures relate to integration and explore the source of these differences. We follow up these empirical explorations with a model that we propose explains the observed patterns.

## II. An Illustrative Case: Vertical Integration and Plant Size

Figure 1 shows the integration-size relationship in our dataset of 1.8 million observations of plants in hundreds of manufacturing industries over a 20 year period. In simplest terms, it plots the probability that a plant is vertically integrated as a function of the plant’s within-industry size percentile. It is clear from the figure that intra-industry differences in integration

---

<sup>2</sup> This is not to say that industry-specific strategic, legal, or technological factors are unimportant to vertical structure, but rather that the applicability of any one of these explanations is limited to the possibly small set of particular industries in which they apply.

<sup>3</sup> See, for example, Tirole (1988), Carlton and Perloff (2005), and Joskow (2005) for discussions of ways that vertical integration can enhance efficiency, particularly with regard to reducing distortions in input markets.

<sup>4</sup> Note that we are referring to the relationship between vertical integration and the scale of a *plant* (i.e., a unique location where production takes place, like a factory), not *firm* scale. One might expect vertically integrated firms (those having plants in vertically connected industries) to be larger on average than unintegrated firms for purely mechanical reasons, since integrated firms by our definition own multiple plants, while many unintegrated firms own only one plant. No such mechanical relationship needs hold for plants. Large firms with vertical plant structures could in principle be large only via the extensive margin; i.e., all plants could be the same size and larger firms simply own more plants.

likelihoods are related to size (no standard error bounds are plotted because they would be virtually imperceptible in the figure given our sample size). The size “effect” also has a substantial magnitude. While industries’ smallest plants are almost never integrated, the median-sized plant is integrated 7 percent of the time, and a whopping 70 percent of plants in the top percentile of their industry size distribution are integrated.

We note that this connection between scale and integration status has, to our knowledge, never been documented empirically or, for that matter, even theoretically addressed. Yet as we show below, it is part of a robust pattern of within-industry heterogeneity in production measures being tied to the likelihood that a plant is in an integrated structure within its firm.

The figure is constructed as follows. The first step is to compute the within industry-year size percentiles for each of our roughly 1.8 million plant-year observations. The data are from U.S. Economic Census microdata, specifically the 1977, 1982, 1987, 1992, and 1997 Census of Manufactures (CM). (This data will be described in more detail below.) There are roughly 360,000 plants in each CM. We measure plant size as the value of its output—its total value of shipments adjusted for inventory changes.

Industries are defined using the 1987 Input-Output Industry Classification System. This is the taxonomy used by the Bureau of Economic Analysis (BEA) for constructing the Benchmark Input-Output tables. Within the manufacturing sector, this system closely mimics the SIC 4-digit system, though there is some aggregation of SIC industries, and more rarely, SIC industries are split among input-output (I-O) industries.<sup>5</sup> While the CM data does not contain plants’ I-O industry classifications, it does contain their SIC codes, so reclassification is straightforward using the BEA’s published concordance.<sup>6</sup> Given every plant’s industry, it is

---

<sup>5</sup> The SIC industries that are aggregated in the input-output taxonomy are typically those that sell different outputs to a “final demand” sector (e.g., personal consumption expenditures or gross private fixed investment) and use similar intermediate materials inputs and production processes. Since the input-output classification system is primarily concerned with goods and services transfers within the production chain, it places less importance on distinguishing products that vary only from the standpoint of final demanders. Since we share the focus on within-production-chain transfers here as well, the input-output classification system is appropriate for our analysis. The largest such aggregation, in terms of the number of industries involved, in the 1987 input-output system is industry 180400, “apparel made from purchased materials.” This one input-output industry consists of the 23 four-digit SIC industries in 231x-238x. These SIC industries use similar inputs and production processes to make various apparel products primarily for personal consumption. Examples include industries like “mens’ and boys’ neckwear,” “women’s, misses’, and juniors’ dresses,” and “robes and dressing gowns.”

<sup>6</sup> A given plant is assigned to a unique industry. Some plants do produce final products that fall under different four-digit SIC industries, however. The Census Bureau classifies such plants based on their primary product (almost always the product accounting for the largest share of revenue).

straightforward to compute each plant's output percentile within its industry and year.

The next step in constructing the figure is identifying vertically integrated plants. The Economic Census microdata contains owning-firm identification numbers for virtually every plant in the nonfarm private sector. We use these firm identifiers to determine the complete set of industries in which a firm operates, including when applicable industries outside of the manufacturing sector.<sup>7</sup> We identify if the same firm owns plants in industry pairs that comprise a substantial link in a vertical production chain. We define a “substantial link” to exist between one industry and another based on the relative volume of trade flows between those two industries. Specifically, a substantial link exists between Industry A and any industry from which A buys at least five percent of its intermediate materials, or any industry to which A sells at least five percent of its own output. The industry pairs that comprise such links are determined using the BEA's Benchmark Input-Output Tables.<sup>8</sup> When a firm owns one or more plants in both ends of a substantial vertical link, then all the firms' plants in those two industries are considered to be vertically integrated. Note that while we use ownership information across all industries to determine which plants are and are not integrated, our empirical analysis uses only the manufacturing plants in these vertical links, because the detailed production data we use is only available in the manufacturing sector.

As an example of how integration status is determined, consider a plant in I-O industry 490100, a.k.a. pumps and compressors. According to the Benchmark Input-Output Tables, this industry receives at least five percent of its total intermediate inputs from three upstream industries: 370200 (iron and steel foundries), 530400 (motors and generators), and 690100 (wholesale trade). Of its customers outside of final demand sectors, it sells more than five percent of its output to only a single I-O industry: 110000 (construction). A pump-compressor plant is labeled as vertically integrated, then, if its firm also owns a steel foundry, a motor-generator plant, an establishment housing a wholesaling operation, or a construction office. The corresponding plant(s) in the vertically linked upstream or downstream industry (industries) are also considered vertically integrated. Notice that integration is defined at the plant, not firm,

---

<sup>7</sup> Our sample begins with the 1977 CM because comprehensive firm ownership information outside manufacturing is obtained from the Standard Statistical Establishment List, which is only available since 1975.

<sup>8</sup> We use the 1987 tables. Given that the I-O structure of the economy is fairly stable over time, we do not expect those intertemporal differences in vertical commodity flows that we miss by using a single table over our whole sample to have a large impact.

level. If an integrated plant's owning firm also owns other plants that are not in a vertical production chain, these plants are not considered vertically integrated simply because the firm owns some plants that are.<sup>9</sup> Recall again that while we identify integrated plants by observing firms' operations throughout the economy, only the manufacturers are included in our analyses below: here, the pump-compressor plant and, if applicable, the steel foundry or motor-generator plant(s). There is not adequate production information in the microdata for establishments engaged in wholesale or construction activity.

Given the output percentile and vertical integration status of each plant-year observation, the final step in creating Figure 1 is pooling each percentile's plants across industries and counting the fraction of plants that are vertically integrated. The result is plotted against the percentile in the figure. The relationship's lack of noise is notable. We stress that the figure is essentially raw data—a frequency count by category. No functional forms have been imposed or parametric approximations made to produce it. The power-law-like relationship is purely a factor of the underlying process driving vertical integration in the data.

To further assure that the result is not an artifact of some peculiar aggregation phenomenon, we have constructed analogous plots for three separate industries in Figure 2. We chose these industries, pumps and compressors, inorganic and organic chemicals, and miscellaneous plastics, not because of considerations about their particular patterns of integration, but rather because they happen to at (respectively) the 75<sup>th</sup>, 90<sup>th</sup>, and 99<sup>th</sup> percentiles of numbers of industry establishments. (We need industries with many establishments for such plots because they only meaningful if there are hundreds of plant-year observations for the particular industry.) Comparing these industry-specific plots, it is apparent that the convex pattern observed in the pooled sample is separately present in each industry. It is true that the *level* of vertical integration varies across industries. This presumably reflects the influence, discussed before, of industry-specific technological or strategic considerations in chemicals that make integration more advantageous. However, *changes* in integration likelihoods across industry plants of different sizes are similar across industries.

The patterns in Figures 1 and 2 are not artifacts of using the (clearly arbitrary) five-percent input- or output-share thresholds to identify “substantial” vertical links. We have checked our results for robustness to different cutoffs and found little qualitative difference.

---

<sup>9</sup> There is one exception to this case that will be discussed in the empirical work below.

Figure A1 shows an example of this. It shows the integration-scale relationship when ten-percent share cutoffs are instead used to define which vertical links are substantial. Besides the obvious difference that levels of vertical integration are lower (as they must, due to the more restrictive definition), the comparative implications match those from before.

One caveat that should be kept in mind is that our vertical links are based on industry-level flows. We thus in effect assume that flows between plants in a firm match the industry data. While we would prefer to measure directly how much plants buy from and ship to specific upstream and downstream plants, we cannot observe plant-level flows in the data. We observe plants' total input and output quantities, but not their source or destination. This caution aside, we are confident that our vertical integration measures, since they are based on actual ownership information at a very detailed level of disaggregation—data which to this point has been rarely available to researchers, and often only for a particular industry if so—closely reflect the actual patterns of transfers of goods and services within firms.

### **III. Data**

Throughout this paper we use establishment-level microdata from the U.S. Economic Census. The Economic Census is conducted every five years, in years ending in either a “2” or a “7”. Our sample uses establishments from the 1977, 1982, 1987, 1992, and 1997 censuses. We exclude data from before 1977 because plant-level data was available almost exclusively for the manufacturing sector before this time. This precludes proper classification by vertical integration status of manufacturing plants that are owned by firms that are in fact integrated, but only into non-manufacturing sectors (say, for example, a firm that owns a manufacturing plant and a retail store where it sells the product it makes).

Critically, the Economic Census contains the owning-firm indicators necessary for us to identify which plants are vertically integrated. (See the discussion in Section II of how this classification is done.) In addition, the Census of Manufactures portion of the Economic Census also contains considerable data on plants' production activities. This includes information on annual value of shipments, production and nonproduction worker employment, production worker hours, book values of capital equipment and structures, intermediate materials purchases, and energy expenditures. We use this production data to construct plant-specific output, productivity, and capital intensity measures that will be a primary focus of the empirical work.



Details on these measures are discussed further below and in the Data Appendix.

We augment the base production data with microdata from the Census of Manufactures materials supplement. This auxiliary file contains, by plant, six-digit SIC product-level information on intermediate materials expenditures. We construct from this measures of plants' materials usage employed in some of the tests below.

For very small plants, typically those with less than five employees, the Census Bureau does not elicit detailed production or materials expenditure data from the plants themselves. It instead relies on tax records to obtain information on plant revenues and employment and then imputes all other production data. We exclude such plants—called Administrative Records (AR) plants—from those analyses below that use plant-level measures (e.g., productivity) that would otherwise be constructed from imputed data. While roughly one-third of plants in the Census of Manufactures are AR establishments, they typically comprise a much smaller share of industry-level output and employment aggregates because of their small size.

## **IV. Empirical Results**

### *IV.1. Vertically Integrated Plants as High-“Type” Plants*

We begin our broadening of the integration-scale connection explored above by generalizing that relationship to other plant-level measures of “type.” We think of plants' types as being combinations of idiosyncratic demand and supply fundamentals that affect plant profitability.<sup>10</sup> In industry equilibrium, these are tied to plant observables like size, productivity, and (in some cases) capital intensity.

We use five such measures in our empirical work. They are not independent, but do differ enough in construction to allow us to gauge the consistency (or lack thereof) of our results. One, seen above already, is the plant's logged output. This is its reported value of sales deflated by the appropriate price index for the plant's four-digit SIC industry.<sup>11</sup> Three are productivity measures that differ in their measure of inputs: output per worker, output per worker-hour, and total factor productivity. Each measures plant output in deflated revenue terms and is expressed as the log of the output-input ratio. The other measure is the log of the plant's capital-labor ratio

---

<sup>10</sup> Foster, Haltiwanger, and Syverson (2005) solve a model of industry equilibrium where producers differ along both demand and cost dimensions, and show that plant type can be summarized as a single-dimensional index of demand, productivity, and factor price fundamentals.

<sup>11</sup> Industry-specific price deflators are from the NBER productivity database.

(capital stock per worker-hour). Further details on these measures are in the Data Appendix.

Each of these five measures has been shown in various empirical studies to be positively related to plant survival. Survival probabilities reflect plant type—usually modeled as productivity—in many models of industry dynamics with heterogeneous producers, like Jovanovic (1982), Hopenhayn (1992), Ericson and Pakes (1995), and Melitz (2003). The productivity-survival link has perhaps been the most extensively studied empirically; see Bartelsman and Doms (2000) for a recent review of this literature. Plant scale and survival was the subject of much of Dunne, Roberts, and Samuelson (1989), and capital intensity’s connection to survival was explored in Doms, Dunne, and Roberts (1995).

We first compare plant type measures across integrated and unintegrated producers. We regress plant types on an indicator for plants’ integration status and a set of four-digit SIC industry by year fixed effects. The coefficient on the indicator (which takes the value of one for vertically integrated plants and zero otherwise) captures the average difference between integrated and unintegrated plants. By including fixed effects we are identifying type differences across integrated and unintegrated plants in the same industry-year, avoiding confounding differences in productivity, scale, or capital intensity levels across industries and time. We estimate this specification separately for each of the five plant type proxies and report the results in Table 1, panel A.<sup>12</sup>

It is clear that vertically integrated plants have higher types. A positive and statistically significant difference between integrated and unintegrated producers exists for all five proxies. Integrated plants have labor productivity levels about 43 percent (output per employee) and 37 percent (output per worker-hour) higher on average than their unintegrated industry cohorts. These are sizeable differences. Earlier work in Syverson (2004) found average within-industry-year interquartile logged labor productivity ranges (for both output per worker and per worker-hour) of roughly 0.65; the gaps seen here are more than a third of this. Total factor productivity differences, while still positive and statistically significant, are much smaller, at just over one percent. Obviously, integrated plants are considerably larger in output terms on average; this result essentially maps the pattern shown in Figure 1 into a single coefficient. Capital intensities are substantially higher in integrated plants as well, with integrated producers having on average

---

<sup>12</sup> Sample sizes differ across the specifications because not all the necessary variables for construction of each is available for each proxy measure for every plant-year observation. In particular, capital information is not available in the 1963 and 1997 CMs.

50 percent more capital per worker hour than unintegrated producers in the same industry. This gap is one explanation why the labor productivity differences among integrated and unintegrated producers are large while the TFP differences are much more modest.

A natural question that follows from these results is the causal nature of vertically integrated plants' advantages. There are three possibilities, and they are not mutually exclusive. The gaps could be due to the fact that becoming part of a vertically integrated structure is associated with an increase in an existing plant's type. Or perhaps newly built integrated plants are more productive than newly built plants in unintegrated ownership structures, and because types are persistent this is reflected in the broader population. Finally, it may be that high-type firms that seek to merge new plants into their internal production chains choose plants that already have high types to add to the firm.

We are able to separately investigate these possibilities in the data. We first see if becoming integrated is associated with unusually high growth in productivity, scale, or capital intensity by comparing changes in these values (computed as five-year differences between CMs) for plants that become vertically integrated to changes over the same period for industry plants remaining unintegrated. Operationally, we regress the growth in plants' type proxies on an indicator dummy for plants that become part of integrated production chains. We again include a full set of industry-year fixed effects to account for industry-specific growth patterns. We must restrict the sample here to continuing plants—i.e., those in both the current and prior CMs—that are unintegrated in the prior CM.

The results are shown in panel B of Table 1. Becoming vertically integrated is associated with labor productivity growth that is about three to four percent higher than growth for continuing plants that remain unintegrated. Despite faster labor productivity growth, however, there are no statistically or economically significant differences in TFP growth. This divergence between labor and total factor productivity growth reflects the fact that integrating plants see substantial investment, raising their capital-labor ratios over seven percent faster than plants that remain unintegrated. This relative capital deepening raises the productivity of labor inputs but not the plants' overall factor-neutral efficiency. Finally, note that plants that become integrated experience slightly faster output growth than that seen by their industry counterparts that remain unintegrated.

To explore the second possible causal mechanism, that new integrated plants are more

productive than newly built plants in unintegrated ownership structures, we compare the productivity levels of newly built plants by regressing all new plants' type measures on a dummy for their integration status and industry-year effects. New plants are defined as those appearing for the first time in a CM. Here we exclude observations from the 1963 CM because of censored entry.<sup>13</sup>

Panel C of Table 1 shows the results. The differences among integrated and unintegrated new plants here are similar to those seen in the broader within-industry comparison discussed above. Labor productivity and capital intensity differences are over 30 percent. The TFP gap is smaller than the labor productivity differences, as before, but in the case of new plants here is more than three times as large as the gap among all plants. Scale differences between new integrated and unintegrated plants are still quite large, though somewhat less pronounced than the gap seen in the overall sample. These results indicate that many of the differences observed between integrated and unintegrated producers' type measures reflect persistent differences present even at the time of the plants' entry.

The third possible source of integrated plants' higher type measures is that firms comprised of high-type, vertically integrated plants seeking to expand through merger or acquisition choose to match with unintegrated plants that are already high-type. We test whether or not this is the case in the data by regressing unintegrated plants' type proxies on a dummy indicating if a plant will *become* vertically integrated by the next CM. Again industry-year fixed effects are included. The estimated coefficient on the dummy captures how to-be-integrated plants compare *before integration* to other plants in their industry that do not become integrated over the same five-year period.

The results, which are in panel D of Table 1, make it clear that soon-to-be integrated plants are different from their industry cohort. They are more productive; their labor productivity advantage is 20 percent and they have TFP levels about three percent higher. They are more capital intensive and considerably larger.

Comparing the type disparities in panels B, C, and D to those seen across all plants in panel A suggest that much of the heterogeneity between integrated and unintegrated plants reflects the effect of differences in the assignment of plant types to integration status. That is,

---

<sup>13</sup> A plant's first appearance in a CM is associated with the start of economic activity at its particular locations; i.e., these plants are greenfield entrants. Existing plants that merely change industries between CMs exist in earlier CMs, and as such are not counted as entrants in our sample.

vertically integrated plants are more productive, larger, and more capital intensive primarily because they were either born into integrated structures that way, or because those are the types of unintegrated plants that firms merge into integrated structures. What gaps not accounted for by these underlying differences are closed due to the faster growth in labor productivity, size, and capital intensity experienced by existing plants when they become integrated.<sup>14</sup>

The last fact pointed to in panel D—that plants that are to become vertically integrated already have higher types than their industry competitors—indicates that firms with vertically integrated structures, when choosing to expand through merger or acquisition, do so by incorporating existing plants that are also high-type. This is true even if the matched plants are not vertically integrated before being incorporated into the new vertical structure. We note that this is precisely the pattern implied by assortative matching equilibria where firms are constructed by matching plants of differing types. For reasons explained below, technological complementarities—a key element of our model explaining the patterns we document here—imply just this sort of assortative matching of plant types within firms. The next section takes this finding as a starting point and pushes the assortative matching prediction further.

#### *IV.2. Assortative Matching within Firms*

As was first shown by Koopmans' (1951) assignment model and Becker's (1973) matching model, complementarities predict assortative matching in equilibrium. If firms' vertical structures are driven by complementarities in plants' inputs, this result has implications for the correlation of plant types within firms. Because the gain to matching with a high-type supplier is growing in the buyer's own type, high-type plants have a higher willingness to pay to match with other high-type plants and will outbid low-type plants to do so. The equilibrium outcome has the highest-type producers matching with highest-type suppliers, slightly lower-type producers match with other similar-type suppliers, and so on down the type distribution. The upshot of the matching process is that plants in vertical production chains should have positively correlated types.

We investigate within-firm correlations, again using our five plant type proxies. Unlike

---

<sup>14</sup> These are of course general patterns across the hundreds of manufacturing industries in our sample. These broad patterns do not imply that the relative importance of these sources of type differences doesn't vary across individual industries. It is quite possible that in certain industries most of the type differences reflect changes that occur when plants become integrated rather than pre-existing type dissimilarities.

the comparisons above that are completely among plants in the same industries, we must now compare type measures across a firm's several plants, where each plant operates in a different industry. Inter-industry technology differences make direct comparison of type measures very difficult; for example, a certain output level may be large for a plant in one particular industry but considered quite small in another. Therefore we use plant type measures that have been normalized by their four-digit SIC industry medians in the year of observation. Since all our measures are in logs, these normalized values correspond to proportional deviations relative to the median, meaning we are correlating the relative positions of a firm's plants *within* their respective industries.

This is operationalized by first computing for each plant the weighted average (the weights are deflated revenues) of these proportional deviations across all *other* plants in its firm. (Obviously, only multiple-plant firms can be included in the sample.) We next regress the plant's own within-industry productivity, output, or capital intensity deviation on their respective weighted averages across the other plants in the firm. A positive relationship means that plants with relatively high type measures relative to their industry median are in firms with other plants that are themselves above their industry's median type, and that low-type plants are in firms with other low-type plants.

Panel A of Table 2 shows the results of this exercise. Columns correspond to different plant type measures. The results are consistent with assortative matching among plants: all five measures are strongly correlated among plants in a firm. Plants with high productivity (or output or capital intensity) levels tend to be in firms with other high-productivity (or high-output or high-capital-intensity) plants, while low-type plants join together in their own firms.

Notice, however, that we have made no distinction to this point between firms based on the vertical integration status of their plants. Many of the plants in the multi-unit firms sample are in firms that have no vertical structures within them. The positive within-firm correlations in plant types just seen may result from complementarities-driven assortative vertical matching, but they may also reflect horizontal complementarities as well.<sup>15</sup> To separate any possible horizontal

---

<sup>15</sup> A third possibility is that the positive correlations do not reflect complementarities in the production activities of plants at all, but instead result from common firm-level fixed factors (such as management ability) that impact the type measures of all plants in the firm. However, the notion that this mechanism is the sole explanation for the observed patterns is not consistent with the fact seen above that soon-to-be-integrated plants are already of higher-than-average type for their industry.

complementarities from vertical effects, we repeat the analysis but allow the coefficient on the firm's average productivity level to be different if the firm contains vertically integrated plants.<sup>16</sup> The results are shown in panel B of the table. The within-firm correlation of each plant type measure is significantly stronger in vertically integrated firms than that in other multi-unit firms. This is true both in statistical terms and in terms of economic magnitudes. For example, the coefficients relating plants' labor productivities to those of other plants in their firm are roughly 60 percent higher (e.g., 0.361 for exclusively horizontally integrated multi-unit firms and 0.578 for firms with vertically integrated plants). Even for TFP, which exhibits the smallest difference in the correlations, vertically integrated firms have a coefficient 15 percent higher than that for exclusively horizontal firms.

These results indicate that plants in vertically integrated structures are owned by firms with other plants in similar locations in their industries' type distributions. Further, this within-firm correlation in plant types due to vertical complementarities can be distinguished from horizontally-driven spillovers and is in fact quite large. These findings, especially when combined with the results regarding to-be-integrated plants in the previous section, indicate that assortative matching among vertically integrated plants is an equilibrium outcome across manufacturers in many industries.

#### *IV.3. Input Usage and Vertical Integration*

A distinguishing feature of vertically integrated structures is that they allow firms the ability to provide inputs to themselves. In this section, we investigate how plant-level input usage is tied to its integration status.

Specifically, we test, in a rough way, whether having more inputs (or more important inputs in terms of expenditure shares) is correlated with the likelihood of a plant vertically integrating. We are able to measure plants' input usage in the CM materials supplement. This auxiliary file contains, for each plant, detailed (defined at the six-digit SIC level) expenditures on individual intermediate materials. We can therefore count the number of intermediate six-digit materials the plant purchases as well as construct a Herfindahl-Hirschman index of its materials expenditures shares. This index reflects the relative importance of particular inputs among the entire input set. For a given number of inputs, a higher value indicates more concentrated

---

<sup>16</sup> This is the only place in the analysis where we define vertical integration at the firm rather than plant level.

expenditure shares; plants with only a few inputs (or one) that comprise the bulk expenditures will have high index values.

Of course, even with its unusual level of detail, the CM materials supplement offers only noisy measures of the “inputs” in the model. Conceptually, inputs might be thought of as irreducible units of factor goods or services for which any complementarities in the production function would still be present. This could in reality be as detailed as the particular materials or services used by each employee of the firm, or perhaps an even finer definition than that. Obviously, a simple count or disparity index of six-digit SIC materials categories (with a mean count of 5.2 across all plants in the sample) is only going to roughly approximate this concept at best. Still, it strikes us as likely that these measures will still be positively correlated with the conceptual inputs of the model, so the sign of the relationship between vertical integration likelihoods and the materials counts should still be informative.

We estimate a linear probability model of a plant’s vertical integration status (the dependent variable equals one if the plant is integrated and zero otherwise) on both the count and the Herfindahl-Hirschman index of its purchased intermediate materials. Industry-year fixed effects are included in the specification to account for differences in input usage across industries and time periods. The regression outcomes are shown in the first numerical column of Table 3.

Subject to the measurement caveats raised above, we find that plants with more inputs or whose input purchases are more concentrated are more likely to be vertically integrated. An additional input implies vertical integration is roughly 1.3 percentage points more likely.<sup>17</sup> For a fixed number of inputs, an increase in the materials expenditure Herfindahl from 0 to 1 (a large change that reflects the limiting cases of complete diffusion—which actually has a lower bound of  $n^{-1}$ , where  $n$  is the number of inputs—and complete concentration with one input dominating expenditures) is related to a 5.0 percentage point increase in the probability of vertical integration.

#### *IV.4. Integration and the Complexity/Flexibility of the Production Process*

In this section we explore further the notion of technological complexity being related to the likelihood of vertical integration. We first repeat the analysis with the CM materials use data as above, but allow materials use to have an impact on vertical integration likelihoods that varies

---

<sup>17</sup> In future drafts we plan to all allow this influence to vary with the scale of the plant.



with the complexity of industries' production processes. Clearly this exercise is subject measurement difficulties; no straightforward and comprehensive measure of industries' production complexities exists. We use the industry's ratio of value added to revenue as a proxy. This is based on the simple notion that industries that add more value through their production (per dollar of revenue) have more complex technologies, either because there are more steps involved in turning raw and intermediate materials into the industry's product, or because each step is itself more complex, or a combination of these two considerations.<sup>18</sup>

We again estimate a linear probability model of vertical integration on two measures of plants' materials input use (the count and expenditure Herfindahl), as above, but this time interact these measures with the industry's value added to revenue ratio. The results are in the rightmost column of Table 3. The magnitude of the connection between the number of materials a plant uses and vertical integration status is unrelated to our production complexity proxy. The interaction is not statistically significant and any implied effect is economically small and, in fact, is in the opposite direction predicted. However, the probability of integration is significantly more sensitive to the concentration of materials expenditures in industries with higher value added to revenue ratios.

This result offers weak evidence that complexity might play a role in vertical integration likelihoods. However, it explores only industry-level differences in complexity, reducing usable variation and introducing aggregation issues. We therefore turn to plant-level measures of production complexity to further explore the integration-complexity link.

As at the industry level, there is no direct measure of a plant's production complexity in our data. We use two proxies instead. The first is based on ideas similar to those in Acemoglu, Aghion, Griffith, and Zilibotti (2004). They use industry investment intensity (physical capital

---

<sup>18</sup> Much of the earlier empirical work on vertical integration has used the value added to revenue ratio itself as a measure of vertical integration. This may raise in the reader's mind the issue of whether we are explaining vertical integration to explain vertical integration. We do not think this is an issue here for a couple of reasons. First, the ratio as a measure of vertical integration has several well-known drawbacks (see, e.g., Eckard 1979) and as such may not be closely related to the actual extent of vertical integration in an industry. In addition, we have found that, at least for manufacturing, the extent of integration measured using our ownership-based metrics (like the industry's output share of vertically integrated plants, for example) tends to be *negatively* correlated with the simple value added to revenue ratio, further casting doubt on the ratio as an accurate measure of vertical integration. Finally, even if the ratio did actually reflect the extent of vertical integration in an industry, we are looking here at the *within-industry* integration responses to differences in materials use. We are merely letting the size of this relationship vary with the ratio. Thus while the level of vertical integration in the industry could possibly be correlated with the ratio, there is no reason to think a priori that the ratio should be related to integration's sensitivity to input use differences *within* the industry.

investment per value added) as a proxy for “technology intensity.” Since we have access to capital stock data at the plant level, we can use less noisy capital intensity in place of investment intensity. We also normalize by labor rather than value added.

Our second measure, the ratio of wage supplements (which includes benefits paid by the employer) to the sum of those supplements and the wage bill, is meant to capture the inherent *flexibility* of the plant’s technology, or at least its labor inputs. The notion behind this measure is that, within an industry, those plants with more part-time or temporary workers are more likely to adjust their operating scale. Since part-time and temporary workers are typically less likely to receive benefits, flexible plants should tend to have lower wage supplement ratios. To draw on an analogy to electricity generation, these flexible plants serve as a sort of “peak-load” capacity that supplements the “base-load” capabilities of industry plants with more permanent workforces. “Base-load” plants typically have the advantage of scale efficiency, they have the disadvantage of being more susceptible to demand shocks. On the other hand, “peak-load” plants are typically higher marginal cost, but can adapt to demand shocks effectively by scaling their production levels up and down in a relatively costless manner. As such, they typically operate as “spot-market” players, to fill in occasional gaps between demand and supply. One may thus expect “base load” plants to be more likely to be integrated, as integration (especially if it addresses a double marginalization problem) provides a way to dampen demand shocks. If “peak load” plants do not possess much market power, however, there may not be much of a benefit to them or their up/downstream partners from vertical integration.

To see how these technology measures are related to the likelihood of vertical integration, we again estimate a linear probability model. Of course, we already found above that vertical integration was positively correlated with plant size and capital intensity. To see if complexity (at least as is reflected in our capital intensity and flexible labor measures) is related to vertical integration separately from its correlation with plant size, we include a quintic polynomial in plant revenue. Since we also include industry-year fixed effects, this polynomial acts as a flexible control for the impact of within-industry scale differences on integration. The results are presented in Table 4. For the sake of brevity, only the coefficients on the technology complexity measures are shown.

The results indicate that, controlling for size, industry plants that are more capital intensive and those with higher supplemental wage payment ratios are more likely to be

vertically integrated. This is consistent with both the hypotheses above. If capital intensity captures “technology intensity” and its associated potential for creating hold-up problems in inputs, then vertical integration is a likely response to address such problems. Further, if high supplemental wage ratios do indicate “base-load” plants, then such plants are indeed more amenable to vertically integrated structures.

## **V. A (Sketch of a) Model of Plant Type and Vertical Integration (INCOMPLETE)**

We have documented a number of facts relating observables of plants’ production behavior and whether or not they are owned within a vertically integrated structure. Being integrated is correlated with several measures of a plant’s type within its industry: productivity, size, and capital intensity. These differences primarily reflect persistent dissimilarities existing either at plant birth (if it is born into a vertically integrated structure) or before it is brought into the integrated firm, but to a lesser extent also indicates changes that happen when an unintegrated plant becomes integrated. There is evidence that matching on plant types is particularly strong in firms with vertical structures. Further, plants that use a lot of inputs, use a particularly important input, or have more complex/less flexible production processes are more likely to be vertically integrated.

In this section, we sketch out a model that can explain all of these patterns. It has two key elements. First, inputs are assumed to be complements in the production function. This is a common assumption and one which we consider to have strong empirical support. Second, high-type producers in an industry use more inputs in equilibrium. The combination of these two elements imply that any distortions in input provision (which could come from one or more of many possible vertical inefficiencies forwarded in the literature—for example, hold-up due to specific investments, double marginalization, or supply uncertainties—we are purposely not precise about the source because it is not crucial to the story) grow in a plant’s “type” within its industry.<sup>19</sup> Therefore the profit gains that can be obtained by solving such input problems grows in plant type (which is empirically connected to observables like plant scale, capital intensity, and productivity levels). As a result, we are more likely to see these plants in integrated production structures.

---

<sup>19</sup> See, for example, Tirole (1988), Carlton and Perloff (2005), and Joskow (2005) for discussions of ways that vertical integration can enhance efficiency, particularly with regard to solving inefficiencies in input supplies.

A downstream industry is comprised of a continuum of producers of measure  $L$ . Each producer (indexed by  $i$ , where  $I$  is the set of industry producers) makes a distinct variety of the industry product that is sold to a final demand sector. The representative final demand consumer has preferences over these varieties given by

$$U = y + \left[ \int_{i \in I} \tilde{\delta}_i q_i^\eta di \right]^{\frac{1}{\eta}}, \quad (1)$$

where  $y$  is the quantity of a numeraire good and  $0 < \eta < 1$ . The value  $q_i$  is the quantity of good  $i$  consumed, and  $\tilde{\delta}_i$  is a variety-specific taste shifter with distribution  $g(\tilde{\delta})$  which has its domain over positive values and has a mean of one.

These preferences imply that expenditures on any particular variety depend on the relative price of the variety to an appropriate price index over all industry varieties; i.e.,

$$q_i = \left[ \frac{p_i / \tilde{\delta}_i}{P} \right]^{\frac{-1}{1-\eta}} \frac{R}{P}, \quad (2)$$

where  $R$  is aggregate spending on the industry's varieties (which we take as exogenous), and the industry price index is defined as

$$P = \left[ \int_{i \in I} \left( \frac{p_i}{\tilde{\delta}_i} \right)^{\frac{-\eta}{1-\eta}} di \right]^{\frac{1-\eta}{-\eta}}.$$

We use the composite industry good as the numeraire, so  $P = 1$ .

This implies industry producers face the following demand function:

$$q_i = R \left( \frac{p_i}{\tilde{\delta}_i} \right)^{\frac{-1}{1-\eta}} = R \delta_i p_i^{-1/(1-\eta)}, \quad \text{where } \delta_i \equiv \tilde{\delta}_i^{1/(1-\eta)}. \quad (3)$$

Thus producer  $i$  faces constant-elasticity demand that is shifted both by aggregate taste shifts (through  $R$ ) and idiosyncratic changes in consumers' valuation of its product (through  $\delta_i$ ). It is also straightforward to show that producer  $i$ 's revenues as a function of its output are

$$r_i = (R \delta_i)^{1-\eta} q_i^\eta. \quad (4)$$

Downstream producers make output using a choice from one of two technologies. One uses a single input with constant returns (we denote this input  $x_a$  for reasons that will become obvious momentarily):

$$q_i = x_a. \quad (5a)$$

The other technology incorporates a second input,  $x_b$ . This alternative technology still exhibits constant returns to scale, but now each input exhibits decreasing marginal returns and can (possibly) enter asymmetrically into production:

$$q_i = (ax_a^\alpha + bx_b^\alpha)^{\frac{1}{\alpha}}. \quad (5b)$$

The parameters  $a$  and  $b$  are the importance weights of  $x_a$  and  $x_b$  respectively, and  $\alpha$  is the elasticity of substitution between them. We assume that adoption of the two-input technology entails paying an additional cost beyond that required of the one-input technology.<sup>20</sup> We motivate this extra cost as resulting from having to rearrange the production setup of the plant to accommodate an additional input, setting up exchange agreements with an additional supplier, and so on. These costs, which we think are best characterized as invariant to the total amount of  $x_b$  purchased, are embodied in the model using the two-input technology fixed cost  $f_2$ .

A downstream firm can purchase the inputs from upstream suppliers at prices  $p_a$  and  $p_b$ . Distortions cause these prices to be above the upstream suppliers' marginal costs; that is,  $p_a > c_a$  and  $p_b > c_b$ . We purposefully do not specify the source(s) of such distortions. Their cause does not matter to our results, and in reality they arise from multiple potential sources depending on the particulars of the industry or market. What is important is that the distortions could, at least in principle, be wholly or partially eliminated through vertical integration. For example, double marginalization (see, for example, Spengler (1950)) is one possible distortion, but so are vertical specific-capital investment externalities and their associated hold-up or free-riding problems (e.g., Williamson (1979) or Grossman and Hart (1986); here the “markup” over marginal cost is the shadow cost of the distortion created by the hold-up possibility) and supply uncertainties (e.g., Carlton (1979) or Bolton and Whinston (1993)). For simplicity, we assume that the marginal costs and distortion-driven markups are equal for both inputs. Therefore  $p_a = p_b = p$

---

<sup>20</sup> This feature is similar in spirit to the multi-input setup in Acemoglu, Antras, and Helpman (2005), which allows producers to choose the number of inputs in their technology subject to a coordination cost function that is increasing and convex in the number of inputs.

and  $c_a = c_b = c$ . This could easily be generalized, however, at the complication of additional algebra.

Downstream producer profits are given by

$$\pi_i = (R\delta_i)^{1-\eta} x_a^\eta - px_a \quad (6a)$$

under the one-input technology and

$$\pi_i = (R\delta_i)^{1-\eta} \left( ax_a^\alpha + bx_b^\alpha \right)^\frac{\eta}{\alpha} - px_a - px_b - f_2 \quad (6b)$$

under the two-input technology.

Optimal output, sales revenue, and maximized profits in the one-input case are as follows (we drop producer-specific subscripts, but add indicators for organizational form—1NI indicates a one-input technology producer that is not vertically integrated):

$$x_{a,1NI} = \eta^{\frac{1}{1-\eta}} p^{\frac{-1}{1-\eta}} R\delta \quad (7a),$$

$$r_{1NI} = \eta^{\frac{\eta}{1-\eta}} p^{\frac{-\eta}{1-\eta}} R\delta \quad (7b)$$

and

$$\pi_{1NI} = (1-\eta)\eta^{\frac{\eta}{1-\eta}} p^{\frac{-\eta}{1-\eta}} R\delta = (1-\eta)r_{1NI}. \quad (7c)$$

The analogous values in the two-input case are

$$x_{a,2NI} = \eta^{\frac{1}{1-\eta}} p^{\frac{-1}{1-\eta}} a^{\frac{1}{1-\alpha}} \left( a^{\frac{1}{1-\alpha}} + b^{\frac{1}{1-\alpha}} \right)^{\frac{\eta-\alpha}{\alpha(1-\eta)}} R\delta \quad (8a),$$

$$x_{b,2NI} = \eta^{\frac{1}{1-\eta}} p^{\frac{-1}{1-\eta}} b^{\frac{1}{1-\alpha}} \left( a^{\frac{1}{1-\alpha}} + b^{\frac{1}{1-\alpha}} \right)^{\frac{\eta-\alpha}{\alpha(1-\eta)}} R\delta$$

$$r_{2NI} = \eta^{\frac{\eta}{1-\eta}} p^{\frac{-\eta}{1-\eta}} b^{\frac{\eta}{\alpha(1-\eta)}} \left( \left( \frac{a}{b} \right)^{\frac{1}{1-\alpha}} + 1 \right)^{\frac{\eta(1-\alpha)}{\alpha(1-\eta)}} R\delta \quad (8b)$$

and

$$\pi_{2NI} = (1-\eta)r_{2NI} - f_2. \quad (8c)$$

Note that given these profit functions, a producer would only choose the two-input technology if the following condition holds (here we assume for simplicity that the two-input technology has  $a = b = 1$ ; in other words, moving from the single- to the two-input technology

simply adds the possibility of using another input that enters into the production function symmetrically):

$$\delta > \frac{p^{\frac{\eta}{1-\eta}} f_2}{\left(2^{\frac{\eta(1-\alpha)}{\alpha(1-\eta)}} - 1\right) (1-\eta) \eta^{\frac{\eta}{1-\eta}} R}. \quad (9)$$

This is the first implication of the model: only producers with sufficiently high idiosyncratic demand for their product use two inputs. Since sales are positively related to  $\delta$  (see (4)), larger producers use more inputs. The intuition behind this result is straightforward. Because there is a fixed cost to switch to the lower marginal cost two-input technology, a downstream producer needs a sufficient market size for its product to justify paying  $f_2$ .<sup>21</sup> We will see further below the importance of a particular producer's market size (which is differentiated from its competitors by the producer's value of  $\delta$ ) in determining the distribution of organizational types within an industry.

Suppose now that a downstream producer could vertically integrate with its upstream supplier(s), allowing it to obtain the input(s) at marginal cost due to the reduction in associated vertical distortions that come from arms-length buyer-supplier relationships (examples above).<sup>22</sup> However, there is a fixed cost involved in vertically integrating,  $f_{VI}$ , that must be paid for each input the producer integrates into. In the one-input case, this means the revenues and profits of downstream suppliers who integrate are given by:

$$r_{VI} = \eta^{\frac{\eta}{1-\eta}} c^{\frac{-\eta}{1-\eta}} R \delta \quad \text{and} \quad \pi_{VI} = (1-\eta) r_{VI} - f_{VI}. \quad (10a)$$

In the two-input case, we allow for the possibility that a downstream producer could integrate into only one or possibly both inputs. In the former case, assuming without loss of generality that  $a > b$  and the downstream producer vertically integrates production of input A,

---

<sup>21</sup> The two-input technology has lower marginal costs under the current parameterization because it takes advantage of the production function's curvature with respect to individual inputs.

<sup>22</sup> This could easily be generalized to a case where integration allows the downstream firm to obtain the inputs at some cost that is between the upstream sector's marginal cost of production and the distortion-laden prices. This would of course reduce the gains that could be attained through integration, but the basic tradeoffs and influences on them would be just as much at play as they are under the current assumption that integration eliminates distortions altogether. Alternatively, an intermediate reduction in acquisition costs might simply reflect that the integrated producer is less efficient at making the input than specialized upstream producers, but that this efficiency gap is still smaller than the size of the vertical distortion.

profits are obtained by substituting (7b) into (6b) and noting that the per-unit cost of input A is  $c$  while it is still  $p$  for input B. Simplifying yields:

$$r_{2VIa} = \eta^{\frac{\eta}{1-\eta}} p^{\frac{-\eta}{1-\eta}} b^{\frac{\eta}{\alpha(1-\eta)}} \left( \left( \frac{a}{b} \right)^{\frac{1}{1-\alpha}} + 1 \right)^{\frac{\eta(\eta-\alpha)}{\alpha(1-\eta)}} \left( \left( \frac{p}{c} \right)^{\frac{\alpha}{1-\eta}} \left( \frac{a}{b} \right)^{\frac{1}{1-\alpha}} + 1 \right)^{\frac{\eta}{\alpha}} R\delta \quad (10b)$$

and

$$\pi_{2VIa} = r_{2VIa} - \eta^{\frac{1}{1-\eta}} p^{\frac{-\eta}{1-\eta}} b^{\frac{\eta}{\alpha(1-\eta)}} \left( \left( \frac{a}{b} \right)^{\frac{1}{1-\alpha}} + 1 \right)^{\frac{\eta-\alpha}{\alpha(1-\eta)}} \left( \left( \frac{p}{c} \right)^{\frac{\eta}{1-\eta}} \left( \frac{a}{b} \right)^{\frac{1}{1-\alpha}} + 1 \right) R\delta - f_2 - f_{vi}. \quad (10c)$$

If the downstream producer instead integrates into production of both inputs, sales and profits are

$$r_{2VI} = \eta^{\frac{\eta}{1-\eta}} c^{\frac{-\eta}{1-\eta}} b^{\frac{\eta}{\alpha(1-\eta)}} \left( \left( \frac{a}{b} \right)^{\frac{1}{1-\alpha}} + 1 \right)^{\frac{\eta(1-\alpha)}{\alpha(1-\eta)}} R\delta \quad \text{and} \quad \pi_{2VI} = (1-\eta)r_{2VI} - f_2 - 2f_{vi}. \quad (11)$$

Note that in this latter case the fixed cost of integrating  $f_{VI}$  is paid twice, once for each input.

We can now compare producers' profits under these various organizational forms. Since the only source of producer heterogeneity is  $\delta$ , determination of optimal form depends completely on the producer's  $\delta$ .

Downstream producers' variable profits under each organizational form can be rank ordered (though an additional assumption is necessary in one case). The lowest variable profits are those for nonintegrated producers using the single-input technology. They also have the lowest fixed costs (zero).<sup>23</sup>

In the analysis that follows, we make one assumption about the relative size of the fixed costs of changing to the two-input technology and of vertically integrating, namely:

---

<sup>23</sup> An additional fixed cost common to all technologies could easily be added to the model with very little impact on the implied outcomes, since the key determinants of organizational form are relative profit levels that would be unaffected by adding such a cost. The only difference is that producers with very low  $\delta$  draws would not be able to cover their fixed costs and would therefore not operate in the industry. If we were to extend the model to incorporate endogenous entry into the industry, some selection mechanism would be necessary to pin down the equilibrium number of producers, and the presence of a fixed cost that all producers must cover would be a reasonable candidate for such a mechanism.



$$\frac{f_{VI}}{f_2} > \frac{\left(\frac{p}{c}\right)^{\frac{\eta}{1-\eta}} - 1}{\left(a^{\frac{1}{1-\alpha}} + b^{\frac{1}{1-\alpha}}\right)^{\frac{\eta(1-\alpha)}{\alpha(1-\eta)}} - 1}. \quad (12)$$

This condition implies that the costs of vertically integrating are sufficiently large relative to the two-input technology fixed cost. This ensures that the  $\delta$  at which a producer is indifferent between the one-input unintegrated form and the two-input unintegrated form (derived for a special case above) is lower than the  $\delta$  that implies indifference between the one-input unintegrated and integrated forms (this latter  $\delta$  can be easily derived using the expressions above). When the condition holds, there is no value of  $\delta$  at which any one-input producer would be vertically integrated, since it can do strictly better by operating the two-input technology as an unintegrated producer. It seems intuitive that the fixed transaction costs required to incorporate a second input into production would be smaller than those tied to vertically integrating into the production of an input. The former case involves only intrafirm coordination, while the latter requires that integrating firms absorb the possibly manifold transaction costs of dealing with external firms. Hence the left hand side of (12) is likely to be well above one. As long as the vertical distortions present in the arms-length relationship (embodied in  $p/c$ ) are not too large relative to the benefits of adding an additional input, the condition will hold.

To do:

- Show that slope of profit function rises as inputs are integrated. Since integration involves a fixed cost, only high- $\delta$  producers integrate.
- Show how variation in production function parameter  $\alpha$  is related to integration. As  $\alpha$  falls to zero, complementarities between inputs grow. This increase the gain to vertically integrating. This is related to the result in Antras, Acemoglu, and Helpman (2005).

## VI. Discussion and Future Work (INCOMPLETE)

Our analysis has scratched the surface of the connection between vertical integration and plant-level patterns in production activity and “type,” but several open questions remain. We hope to address them in future versions of this paper. Two examples of such questions follow.

*Plant size and the depth and direction of vertical integration.* We define integrated

plants above as those owned by a firm that owns plants in any substantially vertically linked industry. However, there are often several links in a production chain upstream and downstream from a plant. One might expect plants in multi-link integrated structures to be systematically different from those in shorter vertical structures. For example, manufacturing plants whose firms own not only upstream suppliers to those plants but downstream plants as well might be systematically different from plants in the same industry whose firms are only integrated upstream. Since we observe all the industries in which firms operate, we will be able to make such distinctions in our data. This allows comparisons of the scales, productivity levels, and other attributes of integrated plants to the depth (the number of links of the production chain) and the direction (upstream or downstream) of their integration structures.

*Geography and vertical integration.* An interesting question about integration and the boundary of the firm is whether space is a substitute for, or complement to, ownership. It may be that input problems are more easily resolved with an arms-length supplier that is physically near the buyer plant due to lower monitoring costs and so on. If this is true, vertical integration is more likely to be seen in conjunction with greater spatial dispersion. Spatial dispersion may in turn be related to plant size and its interaction with vertical structure. On the other hand, these patterns would be reversed if proximity raises the return to vertical integration, say because coordinated management oversight is easier. We can use the geographic information in the Economic Census, which includes plants' ZIP codes and even street addresses in some cases, to investigate such issues.

## Data Appendix

We describe here details on the construction of our production variables.

*Labor Productivity (worker-based and hour-based).* We measure labor productivity both in terms of plant output per worker and plant output per worker-hour. Output is the plant's inventory-adjusted total value of shipments (deflated to 1987 dollars using price indexes from the NBER Productivity Database). Total employment is reported directly by the plant in the CM, and hours are constructed as described below.

*Total Factor Productivity.* We measure productivity using a standard total factor productivity index. Plant TFP is computed as its logged output minus a weighted sum of its logged labor, capital, materials, and energy inputs. That is,

$$TFP_{it} = y_{it} - \alpha_l l_{it} - \alpha_k k_{it} - \alpha_m m_{it} - \alpha_e e_{it},$$

where the weights  $\alpha_j$  are the input elasticities of input  $j \in \{l, k, m, e\}$ . Output is the plant's inventory-adjusted total value of shipments deflated to 1987 dollars. While inputs are plant-specific, we use industry-level input cost shares to measure the input elasticities. These cost shares are computed using reported industry-level labor, materials, and energy expenditures from the NBER Productivity Database (which is itself constructed from the CM). Capital expenditures are constructed as the reported industry equipment and building stocks multiplied by their respective BLS capital rental rates in the corresponding two-digit industry.

*Labor Hours.* Production worker hours are reported directly in the CM microdata. To get total plant hours, we multiply this value by the plant's ratio of total salaries and wages to production worker wages. This, in essence, imputes the hours of non-production workers by assuming that average non-production worker hours equal average production worker hours within plants.

*Capital-Labor Ratios.* Equipment and building capital stocks are plants' reported book values of each capital type deflated by the book-to-real value ratio for the corresponding three-digit industry. (These industry-level equipment and structures stocks are from published Bureau of Economic Analysis data.) Any reported machinery or building rentals by the plant are inflated to stocks by dividing by a type-specific rental rate.<sup>24</sup> The total productive capital stock  $k_{it}$  is the sum of the equipment and structures stocks. This is divided by the plants' number of labor hours to obtain the capital-intensity measure used in the empirical tests.

*Real Materials and Energy Use.* Materials and energy inputs are plants' reported expenditures on each divided by their respective industry-level deflators from the National Bureau of Economic Research Productivity Database.

---

<sup>24</sup> Capital rental rates are from unpublished data constructed by the Bureau of Labor Statistics for use in computing their Multifactor Productivity series. Formulas, related methodology, and data sources are described in U.S. Bureau of Labor Statistics (1983) and Harper, Berndt, and Wood (1989).

## References

- Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti. "Vertical Integration and Technology: Theory and Evidence." NBER Working Paper 10997, 2004.
- Acemoglu, Daron, Pol Antras, and Elhanan Helpman. "Contracts and the Division of Labor." NBER Working Paper 11356, 2005.
- Acemoglu, Daron, Simon Johnson, and Todd Mitton. "Integration: Finance, Contracts, and Regulation." NBER Working Paper 11424, 2005.
- Antras, Pol. "Firms, Contracts, and Trade Structure." *Quarterly Journal of Economics*, 118(4), 2003, 1375-1418.
- Bartelsman, Eric J. and Mark Doms. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature*, 38(3), 2000, 569-95.
- Bolton, Patrick and Michael D. Whinston. "Incomplete Contracts, Vertical Integration, and Supply Assurance." *Review of Economic Studies*, 60(1), 1993, 121-148.
- Carlton, Dennis W. "Vertical Integration in Competitive Markets Under Uncertainty." *Journal of Industrial Economics*, 27(3), 1979, 189-209.
- Carlton, Dennis W. and Jeffrey M. Perloff. *Modern Industrial Organization, Fourth Edition*. Boston: Pearson Addison Wesley, 2005.
- Coase, R. H. "The Nature of the Firm." *Economica*, 16(4), 1937, 386-405.
- Doms, Mark, Timothy Dunne, and Mark J. Roberts. "The Role of Technology Use in the Survival and Growth of Manufacturing Plants." *International Journal of Industrial Organization*, 13(4), 1995, 523-542.
- Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. "The Growth and Failure of U. S. Manufacturing Plants." *Quarterly Journal of Economics*, 104(4), 1989, 671-698.
- Eckard Jr., E. Woodrow. "A Note on the Empirical Measurement of Vertical Integration." *Journal of Industrial Economics*, 28(1), 1979, 105-107.
- Ericson, Richard and Ariel Pakes. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62(1), 1995, 53-82.
- Foster, Lucia, John Haltiwanger, and Chad Syverson. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" NBER Working Paper 11555, 2005.

- Grossman, Sanford J. and Oliver D. Hart. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4), 1986, 691-719.
- Hastings, Justine and Richard Gilbert. "Vertical Integration in Gasoline Supply: An Empirical Test of Raising Rivals' Costs." University of California Energy Institute Power Working Paper, No. PWP-084.2.
- Hopenhayn, Hugo A. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5), 1992, 1127-50.
- Hortaçsu, Ali and Chad Syverson. "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices." Working Paper, University of Chicago, 2005.
- Joskow, Paul. "Vertical Integration." in Claude Menard and Mary Shirley, eds., *Handbook of New Institutional Economics*. Berlin: Springer/Kluwer, 2005.
- Jovanovic, Boyan. "Selection and the Evolution of Industry." *Econometrica*, 50(3), 1982, 649-70.
- Melitz, Marc J. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6), 2003, 1695-1725.
- Perry, Martin K. "Vertical Integration: Determinants and Effects." in Richard Schmalensee and Robert D. Willig, eds., *Handbook of Industrial Organization*. Amsterdam: North-Holland, 1989.
- Salop, Steven C. "Vertical Mergers and Monopoly Leverage." in Peter Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, London: Macmillan, 1998.
- Spengler, Joseph J. "Vertical Integration and Antitrust Policy." *Journal of Political Economy*, 58(4), 1950, 347-352.
- Syverson, Chad. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics*, 86(2), 2004, 534-550.
- Tirole, Jean. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press, 1988.
- Williamson, Oliver E. "Transaction-Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics*, 22(2), 1979, 233-261.

Table 1. Plant Attributes and Vertically Integrated Plants

	Revenue per employee	Revenue per hour	TFP	Output	Capital-labor ratio
A. Within-Industry Differences					
N	1,063,369	1,046,562	738,889	1,073,978	785,111
R <sup>2</sup>	0.396	0.384	0.550	0.376	0.459
VI indicator	0.359* (0.002)	0.316* (0.002)	0.011* (0.001)	1.522* (0.004)	0.414* (0.003)
B. Changes upon Integration					
N	410,636	402,106	253,945	411,977	276,441
R <sup>2</sup>	0.055	0.052	0.048	0.074	0.084
Newly VI indicator	0.039* (0.004)	0.035* (0.004)	-0.005 (0.004)	0.014* (0.005)	0.071* (0.008)
C. Differences among New Plants					
N	240,922	234,418	151,869	246,464	164,536
R <sup>2</sup>	0.343	0.338	0.501	0.278	0.426
VI indicator	0.343* (0.005)	0.283* (0.004)	0.036* (0.003)	1.266* (0.009)	0.372* (0.007)
D. Comparing Unintegrated Plants: To-Be-Integrated vs. Remaining Unintegrated					
N	653,092	641,092	573,442	662,492	614,048
R <sup>2</sup>	0.289	0.277	0.531	0.234	0.419
To-be-VI indicator	0.238* (0.004)	0.207* (0.004)	0.028* (0.003)	1.369* (0.009)	0.237* (0.006)

Notes: This table shows plant “type” comparisons between vertically integrated (or to-be-integrated) plants and their non-integrated counterparts. All regressions include industry-year fixed effects, where industries are defined by four-digit SIC. Panel A shows comparison using all sample plants. Panel B compares changes in type for unintegrated plants that become integrated to changes for unintegrated plants that remain so. Panel C compares new integrated and unintegrated plants, and panel D compares *prior period* types among unintegrated plants that will become integrated by next period to those remaining unintegrated. Samples are comprised of non-AR manufacturing plants. See text and data appendix on construction of type measures and additional details. An asterisk denotes significance at a five percent level.

Table 2. Within-Firm Correlations of Plant “Type” Measures

A. All Multi-Unit Firms

	Revenue per employee	Revenue per hour	TFP	Output	Capital- labor ratio
N	210,943	207,557	157,939	211,678	163,588
R <sup>2</sup>	0.102	0.097	0.054	0.239	0.122
Firm’s weighted mean	0.555* (0.008)	0.544* (0.008)	0.422* (0.014)	0.721* (0.007)	0.588* (0.009)

B. Breakout of Vertically Integrated Firms

	Revenue per employee	Revenue per hour	TFP	Output	Capital- labor ratio
N	301,097	295,689	219,581	302,418	229,650
R <sup>2</sup>	0.111	0.106	0.067	0.249	0.132
Firm’s weighted mean	0.361* (0.007)	0.353* (0.008)	0.366* (0.013)	0.464* (0.007)	0.407* (0.009)
Firm’s weighted mean X VI firm	0.217* (0.009)	0.215* (0.010)	0.055* (0.018)	0.274* (0.008)	0.199* (0.012)

Notes: This table reports coefficients from a regression of plant “type” measures (productivity, output, or capital intensity) on the output-weighted average of the same type measure among other plants owned by the same firm. In panel B this average is interacted with an indicator for firms that own vertically integrated plants. All type measures are log deviations from industry-year medians (industries defined by four-digit SIC). The sample includes all non-AR manufacturing plants in multi-plant firms. See text and data appendix on construction of type measures and additional details. Standard errors are clustered by firm. An asterisk denotes significance at a five percent level.

Table 3. Materials Use and the Likelihood of Vertical Integration

	Indicator for VI	Indicator for VI
N	484,736	484,736
R <sup>2</sup>	0.179	0.171
Number of materials	0.0128* (0.0002)	0.0156* (0.0025)
No. of materials x value-added revenue share		-0.0055 (0.0051)
Materials HHI	0.0504* (0.0033)	-0.0999* (0.0446)
Materials HHI x value-added revenue share		0.2853* (0.0845)

Notes: This table shows estimates from linear probability models where the dependent variable equals one if the plant is vertically integrated and zero otherwise. The explanatory variables are measures of intermediate materials usage taken from the Census of Manufactures materials supplement, which reports plants' expenditures on intermediate materials at the six-digit product level. Both the number of six-digit materials and the Herfindahl-Hirschman index of their expenditure shares are included. One specification interacts these measures with the revenue share of value added in the plant's industry. See text for additional details. Standard errors in the specification including the value added share interaction are clustered at the industry level. An asterisk denotes significance at five percent.

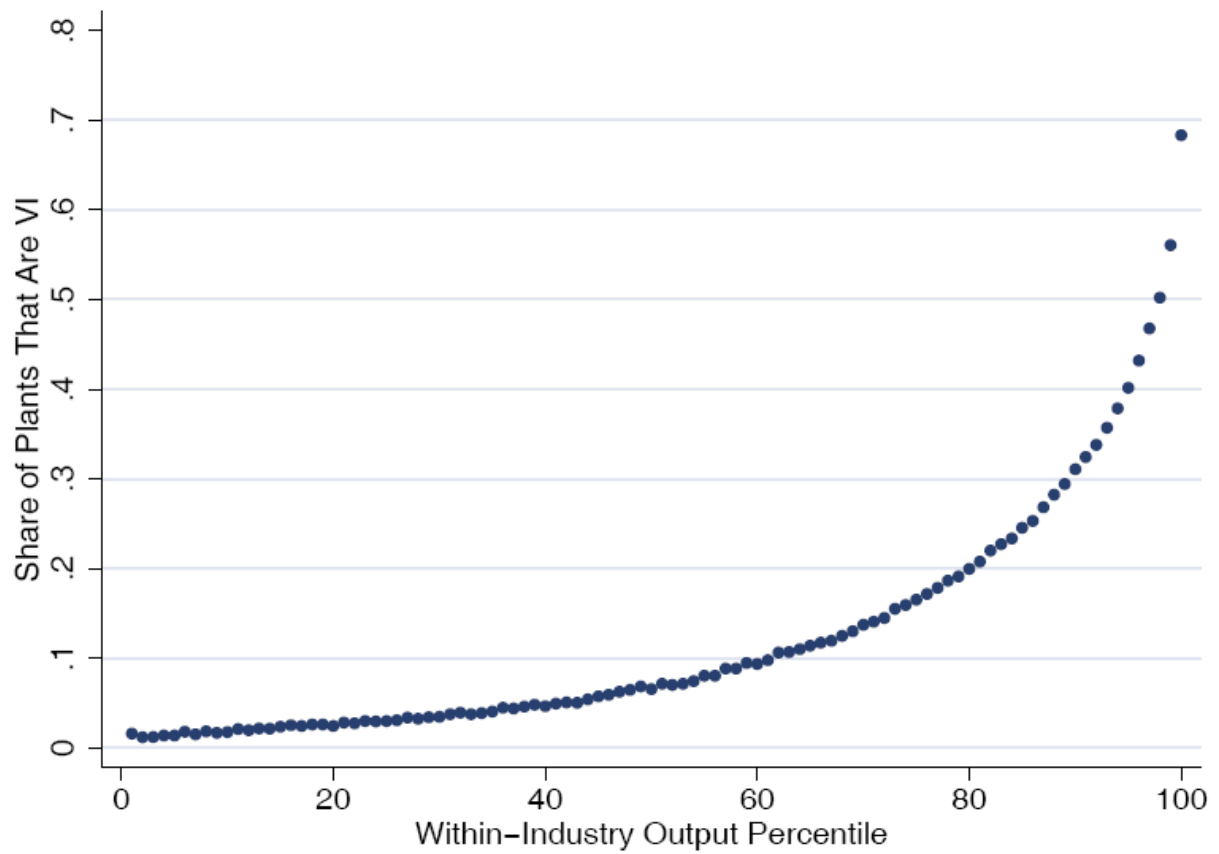


Table 4. Other Technology Attributes and VI Likelihoods

	Technology Measure	
	K/L	Wage ratio
N	786,808	1,073,315
R <sup>2</sup>	0.351	0.344
Technology	0.035* (0.0004)	0.711* (0.008)

Notes: This table shows estimates from a linear probability model where the dependent variable equals one if the plant is vertically integrated and zero otherwise. The explanatory variables measure attributes of plants' technologies. These include capital intensity and the ratio of wage supplements to the sum of paid wages and supplements. See text for additional details. All regressions include a fifth-order polynomial of plant size (measured using deflated revenue) and a complete set of industry-year fixed effects, where industries are defined by four-digit SIC. An asterisk denotes significance at five percent.

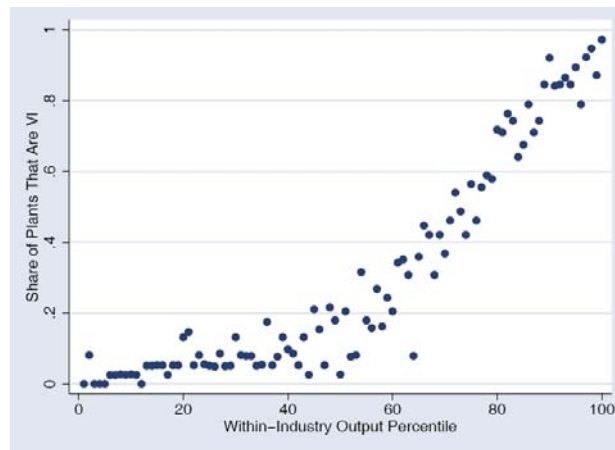
Figure 1. Vertical Integration and Scale



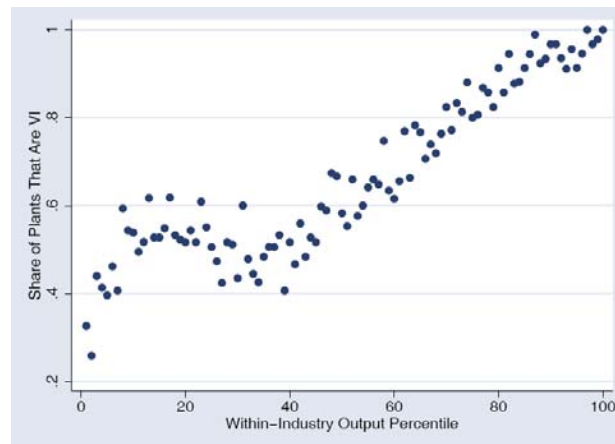
Notes: This figure plots the frequency of vertically integrated manufacturing plants by plant scale (measured by real revenue) percentile within industry-year. N = 1.8 million. See text for details.

Figure 2. Vertical Integration and Scale, Specific Industries

A. Pumps and Compressors (75<sup>th</sup> percentile number of establishments)



B. Inorganic and Organic Chemicals (90<sup>th</sup> percentile number of establishments)



C. Miscellaneous Plastics (99<sup>th</sup> percentile number of establishments)

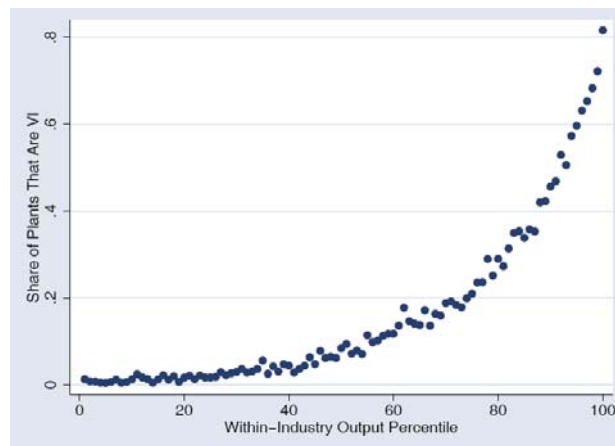


Figure A1. Vertical Integration and Scale: Alternative VI Definition (10 percent “Significant Link” Threshold)

