American Economic Review 2014, 104(4): 1120–1148
http://dx.doi.org/10.1257/aer.104.4.1120

Vertical Integration and Input Flows[†]

By Enghin Atalay, Ali Hortaçsu, and Chad Syverson*

We use broad-based yet detailed data from the economy's goods-producing sectors to investigate firms' ownership of production chains. It does not appear that vertical ownership is primarily used to facilitate transfers of goods along the production chain, as is often presumed: roughly one-half of upstream establishments report no shipments to downstream establishments within the same firm. We propose an alternative explanation for vertical ownership, namely that it promotes efficient intrafirm transfers of intangible inputs. We show evidence consistent with this hypothesis, including the fact that, after a change of ownership, an acquired establishment begins to resemble the acquiring firm along multiple dimensions. (*JEL* G32, G34, L14, L22, L60, M11)

Many firms own links of production chains. That is, they operate both upstream and downstream establishments, where the upstream industry produces an input used by the downstream industry. We explore the reasons for such ownership using two detailed and comprehensive datasets on ownership structure, production, and shipment patterns throughout broad swaths of the US economy.

We find that most vertical ownership does *not* appear to be primarily concerned with facilitating physical goods movements along a production chain within the firm, as is commonly presumed. Upstream units ship surprisingly small shares of their output to their firms' downstream establishments. Almost one-half of upstream establishments do not report making shipments inside their firms. The median internal shipments share across upstream establishments in vertical production chains is

*Atalay: Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637 (e-mail: atalay@uchicago.edu); Hortaçsu: Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637 (e-mail: hortacsu@uchicago.edu); Syverson: Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, IL 60637 (e-mail: Chad.Syverson@chicagobooth.edu). Older versions of this paper previously circulated as "Why Do Firms Own Production Chains?" and "Vertical Integration and Production: Some Plant-Level Evidence." We thank Daron Acemoğlu, Phillipe Aghion, Luis Garicano, Bob Gibbons, Austan Goolsbee, Oliver Hart, Tom Holmes, Bengt Holmström, Tom Hubbard, Javier Miranda, Ezra Oberfield, Stephen Redding, Lynn Riggs, Andrei Shleifer, Chris Snyder, Steve Tadelis, Birger Wernerfelt, and seminar participants at the ASSA meetings, UC Berkeley, Chicago, Chicago Fed, Harvard, HEC Montreal, IIOC, LSE, MIT, NYU Stern, Princeton, SITE, and the US DoJ for helpful discussions and comments. Margaret Triyana provided excellent research assistance. Hortaçsu thanks the NSF (SES-1124073 and ICES-1216083) and Syverson thanks the NSF (SES-0519062), the John M. Olin Foundation, and the Stigler Center for funding. The research in this paper was conducted while the authors were Special Sworn Status researchers of the US 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 US 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.

[†]Go to http://dx.doi.org/10.1257/aer.104.4.1120 to visit the article page for additional materials and author disclosure statement(s).

0.4 percent if shipments are counted equally, and is less than 0.1 percent in terms of total dollar value or weight. Even the ninetieth percentile internal shippers are hardly dedicated makers of inputs for their firms' downstream operations, with 62 percent of the value of their shipments sent outside the firm. (However, a small fraction of upstream establishments—slightly more than 1 percent—*are* operated as dedicated producers of inputs for their firms' downstream operations, and these establishments tend to be quite large. We will discuss this further below.) These small shares are robust to a number of choices we made about the sample, how vertical links are defined, and whether we measure internal shares as a percentage of the firm's upstream production or its downstream use of the product.

If firms do not own upstream and downstream units so the former can provide intermediate materials inputs for the latter, why do they? Our results suggest that a primary purpose of ownership may be to mediate efficient transfers of *intangible* inputs within firms, mirroring Grant's (1996) "organizational capabilities" theory of the firm.[1] Managerial oversight and planning strike us as important types of such intangibles, but these need not be involved.[2] Other possibilities include marketing know-how, intellectual property, and R&D capital.[3] This explanation is consistent with small amounts of shipments within vertically structured firms, and even with an absence of internal shipments altogether.

That vertical integration is often about transfers of intangible inputs rather than physical ones may seem unusual at first glance. However, as observed by Arrow (1975) and Teece (1982), it is precisely in the transfer of nonphysical knowledge inputs that the market, with its associated contractual framework, is most likely to fail to be a viable substitute for the firm. Moreover, many theories of the firm, including the four "elemental" theories as identified by Gibbons (2005), do not explicitly invoke physical input transfers in their explanations for vertical integration.[4] That said, many salient "parables" in the theory of the firm literature, such as the GM-Fisher body example, are about physical goods transfers.[5] This, of course, does not preclude integration from also involving physical input transfers in some cases. As noted above, we find a small number of establishments that are clearly dedicated producers for their firms' downstream production units. However, these are the exception rather than the rule. Thus it appears that the "make-or-buy" decision

[1] We discuss this and related papers in Section IVA.

[2] In contexts like hotel or business services franchising, vertical integration often does not involve transfers of physical goods. Our article, however, focuses on vertical integration and shipments in the *goods-producing* sectors of the economy, like manufacturing, where one may think physical goods transfers across plants are important.

[3] These inputs might be just as likely to be transferred from the firm's "downstream" units to its "upstream" ones as vice versa. The names reflect the flow of the physical production process, not necessarily the actual flow of inputs within the firm.

[4] To quote Gibbons, the four elemental theories of the firm are "(i) a 'rent seeking' theory, which can be discerned in informal theoretical arguments by Williamson (1971, 1979, 1985) and Klein, Crawford, and Alchian (1978); (ii) a 'property rights' theory, which can be discerned in formal models by Grossman and Hart (1986); Hart and Moore (1990); and Hart (1995); (iii) an 'incentive system' theory, which can be discerned in formal models by Holmström and Milgrom (1991, 1994); Holmström and Tirole (1991); and Holmström (1999); and (iv) an 'adaptation' theory, which can be discerned in informal theoretical arguments by Simon (1951); Williamson (1971, 1973, 1975, 1991); Klein and Murphy (1988, 1997); and Klein (1996, 2000)." (pp. 200–201)

[5] Some of the most highly cited early empirical work on vertical integration, including Monteverde and Teece (1982), Masten (1984), and Joskow (1985), focused on situations with physical input transfers. Anderson and Schmittlein (1984) is an interesting example where the integration of the sales force is considered; as they note, often no transfer of title to goods was necessary to the sales organization, even if it was outside the firm. Once again, however, our focus in this article is on the goods-producing sectors of the economy.

(at least referring to physical inputs) can explain only a fraction of the vertical ownership structures in the economy.

We find additional patterns in the data that are consistent with the intangible inputs explanation. First, we show that establishments in vertical ownership structures have higher productivity levels, are larger, and are more capital intensive than other establishments in their industries. These disparities, which we interpret as embodying fundamental differences in establishment "type," mostly reflect persistent differences in establishments started by or brought into vertically structured firms. In other words, while there are some modest changes in establishments' type measures upon integration, the cross-sectional differences primarily reflect selection on preexisting heterogeneity. Controlling for firm size explains most of these type differences; establishments of similarly sized firms have similar types, regardless of whether their firm is structured vertically, horizontally, or as a conglomerate.

Second, by studying how establishments' behavior changes with changes of ownership, we provide suggestive evidence of flows of intangible inputs within vertically structured firms. Acquired establishments begin to resemble existing establishments in their acquiring firms along two key dimensions. First, the acquired establishments start shipping their outputs to locations that their acquirers had already been shipping to. Second, they begin producing products that their acquirers had already been manufacturing.

Besides being consistent with the "organizational capabilities" theory of the firm, these patterns evoke the equilibrium assignment view of firm organization advanced by Lucas (1978); Rosen (1982); and more recently by Garicano and Rossi-Hansberg (2006) and Garicano and Hubbard (2007). To the extent that intangibles are complementary to the physical inputs involved in making vertically linked products, equilibrium assignment typically entails the allocation of higher-type intangible inputs to higher-type establishments in each product category. If establishment size is restricted by physical scale constraints, better intangible inputs will also be shared across a larger number of establishments. Simply put, higher-quality intangible inputs (e.g., the best managers) are spread across a greater set of productive assets. Some of these assets can be vertically linked establishments, but their vertical linkage need not necessarily imply the transfer of physical goods among them.

Furthermore, there may not be anything special about vertical structures per se. The evidence below suggests that firm size, not structure, is the primary reflection of input quality. Larger firms just happen to be more likely to contain vertically linked establishments. In this way, vertical expansion by a firm may not be altogether different than horizontal expansion and is a mode of expansion that is much less likely to raise antitrust concerns. A typical horizontal expansion involves the firm starting operations in markets that are new but still near to its current line(s) of business, under the expectation that its current abilities can be carried over into the new markets. Physical goods transfers among the firm's establishments are not automatically expected in such expansions, but inputs like management and marketing are expected to flow to units in the new markets. Vertical expansions may operate similarly. Industries immediately upstream and downstream of a firm's current operations are obviously related lines of business. Firms will occasionally expand into these lines, expecting their current capabilities to prove useful in the new markets. And, just as with horizontal expansions, transfers of managerial or other

nontangible inputs will be made to the new establishments. Yet no physical good transfers from upstream to downstream establishments need occur.

The upshot is that the organizational capabilities and assignment views of the firm are consistent with large firms composed of high-type establishments operating (often) in several lines of business. Common ownership allows the firm to efficiently move intangible inputs across its production units. Many of these units will be vertically related, making these segments "vertical" in that the firm owns each end of a link in a production chain. But the chain need not exist for the purpose of moderating the flow of physical products along it.

This scenario is consistent with the evidence we document here, and in particular with our primary result about the lack of goods shipments within vertically structured firms. The remainder of the article lays out the evidence and tests the hypothesis in more detail. It is organized as follows. The next section describes our data sources. We then explain in Section II how we use these data to measure vertical integration and shipments sent along vertical chains, within firms. Section III reports the empirical results. Section IV discusses flows of intangible inputs across establishments, within firms. We conclude in Section V.

I. Data

We use microdata from two sources, the US Economic Census and the Commodity Flow Survey, and aggregate data from the Annual Wholesale Trade Survey and the Annual Retail Trade Survey. We discuss each dataset in turn.

Economic Census.—The Economic Census (EC) is an establishment-level census that is conducted every five years, in years ending in either a "2" or a "7." Establishments are unique locations where economic activity takes place, like stores in the retail sector, warehouses in wholesale, offices in business services, and factories in manufacturing. Our sample uses establishments from the 1977, 1982, 1987, 1992, and 1997 censuses. We specifically use those establishments in the Longitudinal Business Database, which includes the universe of all US business establishments with paid employees.[6] The data have been reviewed by Census staff to ensure that establishments can be accurately linked across time and that their entry and exit have been measured correctly.

Critically, the Economic Census contains the owning-firm indicators necessary for us to identify which establishments are vertically integrated.[7] (We discuss in Section II how we make this classification.) Additionally, the Census of Manufactures portion of the EC also contains considerable data on establishments' production activities. This includes information on their annual value of shipments, production and nonproduction worker employment, capital stocks, and purchases

[6] Establishment-level data from before 1977 are almost exclusively for the manufacturing sector, precluding proper classification of vertical ownership for manufacturing plants owned by firms that are in fact vertically structured, but only into nonmanufacturing sectors (e.g., firms that own a manufacturing establishment and a retail store selling the product the establishment makes).

[7] The firm identifiers are designed to capture ownership patterns that exist across establishments. See online Technical Appendix C.1 for a discussion of the audits and checks performed by the US Census Bureau to achieve accurate portrayal of ownership patterns.

of intermediate materials and energy. We use these production data to construct establishment-specific output, productivity, and factor intensity measures; details are discussed further below and in online Technical Appendix A. In some cases, we augment the base production data with microdata from the Census of Manufactures materials and production supplements, which contain, by establishment, product-level information on intermediate materials expenditures (at the six-digit level) and revenues (at the seven-digit level).[8]

Commodity Flow Survey.—The Commodity Flow Survey (CFS) contains data on shipments originating from mining, manufacturing, wholesale, and catalog and mail-order retail establishments, spanning approximately 600 four-digit Standard Industrial Classification (SIC) industries.[9] The survey defines a shipment as "an individual movement of commodities from an establishment to a customer or to another location in the originating company." The CFS takes a random sample of an establishment's shipments in each of four periods during the year, one in each quarter. The sample generally includes 20 to 40 shipments per period, though establishments with fewer than 40 shipments during the survey period simply report all of them.[10]

For each shipment, the originating establishment is observed, as well as the shipment's destination zip code (exports report the port of exit along with a separate entry indicating the shipment as an export), the commodity, the mode(s) of transportation, and the dollar value and weight of the shipment.

We use the microdata from the 1993 and 1997 CFS; the former contains roughly 110,000 establishments and 10 million shipments, and the latter 60,000 establishments and 5 million shipments. As with the Economic Census, each establishment has an identification number denoting the firm that owns it. Both the establishment and the firm numbers are comparable to those in the EC, so we can merge data from the two sources. We match the 1993 CFS to the 1992 EC; this will inevitably lead to some mismeasurement of ownership patterns, but we expect this will be small given the modest annual rates at which establishments are bought and sold by firms.

Annual Wholesale Trade Survey and Annual Retail Trade Survey.—These datasets provide information on aggregate sales and purchases of four-digit retail and wholesale industries. We use these datasets to help determine whether two industries are vertically linked.

[8] For very small EC establishments, typically those with fewer than five employees, the US Census Bureau does not elicit detailed production data from the establishments themselves. It instead relies on tax records to obtain information on establishment revenues and employment and then imputes all other production data. We exclude such establishments—called Administrative Records (AR) establishments—from those analyses that use establishment-level measures constructed from the Census of Manufactures (e.g., productivity). While roughly one-third of establishments in the Census of Manufactures are AR establishments, their small size means they make up a much smaller share of industry-level output and employment aggregates.

[9] Hillberry and Hummels (2003, 2008) and Holmes and Stevens (2010, 2012) use the CFS microdata to investigate various effects of distance on trade patterns. They do not make the within- and between-firm distinctions that we make here. See online Technical Appendix C.2 for a description of the sampling methodology used to construct the CFS.

[10] The length of the survey period is two weeks for the 1993 Commodity Flow Survey and one week for the 1997 CFS.

II. Measuring Vertical Ownership and Shipments within Firms' Production Chains

This section explains how we use our data to determine which businesses are vertically integrated and whether the CFS shipments we observe are internal or external to the firm.

A. *Determining Which Industries Are Vertically Linked to One Another*

We define vertically linked industries as *I-J* industry pairs for which a substantial fraction—1 percent in the baseline specification—of industry *I*'s sales are sent to establishments in industry *J*.[11] To compute the fraction of sales of industry *I* output that are sent to industry *J*, we use information from the 1992 Bureau of Economic Analysis Input-Output Tables, the 1992 Economic Census, the 1993 Commodity Flow Survey, the 1993 Annual Wholesale Trade Survey, and the 1992 Annual Retail Trade Survey. We define industries by their four-digit SIC code. We apply the classification of vertically linked industries implied by these data to our entire sample.[12]

To measure the value of shipments sent by industry *I* establishments to industry *J* establishments, we first compute the shipments of commodity *C* sent to industry *J* using the 1993 CFS. Commodities are defined by their Standard Transportation Commodity Code (STCC).[13] Note that the Commodity Flow Survey records neither the receiving establishment nor the receiving industry of each shipment; the algorithm that we use to impute the value of commodity *C* shipments sent to industry *J* plants is described, in detail, in online Technical Appendices B.1, B.2, and B.3. We then sum over all commodities that each industry *I* ships to determine the share of *I*'s sales going to *J*, thereby indicating which *I-J* industry pairs are vertically linked.

For most industries, we rely primarily on the Input-Output Tables, which track quantities of interindustry flows of goods and services, to perform these calculations. However, the I-O Tables treat the entire wholesale and retail sectors as single, monolithic industries, with no distinction as to the types of products their establishments distribute. Additionally, they do not keep track of shipments by manufacturers to (or through) wholesalers or retailers, instead measuring only those inputs directly used by wholesalers and retailers in the production of wholesale and retail services (e.g., in the I-O Tables, cardboard boxes are a major input used by the wholesale sector, even though they are not a primary product shipped by this sector). To achieve better measurement of the flow of goods through the wholesale and retail sectors, we use a different algorithm that incorporates additional data from the Annual Wholesale Trade Survey and the Annual Retail Trade Survey. These calculations are detailed in online Technical Appendix B.[14]

[11] The 1 percent cutoff used to define substantial vertical links is, of course, arbitrary. We have checked our major findings using a 5 percent cutoff and found few differences. (The overall level of integration is of course lower in this more stringent case.)

[12] Applying one vertical structure to the entire sample is made necessary by the lack of CFS microdata before 1993 and changes in the way the CFS records commodities between 1993 and 1997. Given that the input-output structure of the economy is fairly stable over time, we do not expect a large impact on our results.

[13] A list of STCC codes can be found in pages 117 to 167 of "Reference Guide for the 2008 Surface Transportation Board Carload Waybill Sample," published by Railinc. There are roughly 1,200 commodities represented in the 1993 Commodity Flow Survey.

[14] In a previous draft, we employed a cruder methodology to identify pairs of vertically linked industries, defining industry *I* as upstream of industry *J* provided either (i) *J* buys at least 5 percent of its intermediate materials from

B. *Classifying Shipments as Internal or External to the Firm*

To classify shipments sent by upstream establishments in the Commodity Flow Survey as internal or external to the firm, we first must merge the CFS and EC data. This can be done straightforwardly using the two datasets' common establishment and firm identifiers. Of critical importance is the fact that the Commodity Flow Survey contains the destination zip code of each shipment, while the Economic Census records establishments' zip codes.

Our sample consists of establishments that are at the upstream end of firms' production chains. That is, establishments in our sample are those that are in some four-digit SIC industry, *I*, for which there exists some other establishment in the same firm that is in industry, *J*, where industries *I* and *J* are vertically linked.

We identify a shipment as internal if the shipping establishment's firm also owns an establishment that is both in the destination zip code and in an industry that is in a downstream vertical link (as defined above) of the sending establishment's industry.[15] The CFS contains shipment-specific sample weights that indicate how many actual shipments in the population each sampled shipment represents.[16] We use these weights when computing the shares of internal shipments, be it by count, dollar value, or weight.

III. Shipments within Firms' Vertical Links

We begin by looking at the patterns of shipments within firms' vertical links. We focus on establishments in the Commodity Flow Survey that are at the upstream end of a vertical ownership structure.

A. *Vertically Integrated Establishments' Shipments—Benchmark Sample*

The combined 1993 and 1997 CFS yield a core sample of about 67,500 establishment-year observations of upstream establishments in firms' production chains. These establishments report a total of roughly 6.3 million shipments in the CFS. Panel A of Table 1 shows the prevalence of internal shipments within this sample. It reports quantiles of the distribution of internal shipment shares across our sample establishments, measured as the fraction of the total number, dollar value, and weight of the establishment's shipments.[17]

Overall, only a small share of vertically integrated upstream establishments' shipments are to downstream units in the same firm. Across the 67,500 establishments, the median fraction of internal shipments is 0.4 percent. The median internal shares

I, or (ii) *I* sells at least 5 percent of its own output to industry *J*. We furthermore did not attempt to make any distinction among wholesale or retail industries. While we prefer the current methodology for its increased accuracy, we reproduce our main analysis using the old methodology in the Appendix and find similar results.

[15] Every establishment is assigned to a unique industry. For establishments that produce products that fall under multiple four-digit SIC industries, the US Census Bureau classifies such establishments based on their primary product, which is almost always the product accounting for the largest share of revenue.

[16] Online Technical Appendix C.2 explains how the sample weights are constructed.

[17] For data confidentiality reasons, the reported quantiles are actually averages of the immediately surrounding percentiles; e.g., the median is the average of the forty-ninth and fifty-first percentiles, the seventy-fifth percentile is the average of the seventy-fourth and seventy-sixth percentiles, and so on.

TABLE 1—ESTABLISHMENT-LEVEL SHARES OF INTERNAL SHIPMENTS

Internal share of:	Percentile 50th	75th	90th	95th	Fraction = 0	Fraction = 1	Weighted mean
Panel A. Benchmark							
Establishment shipment counts	0.4%	7.3%	32.2%	62.7%	49.7%	1.2%	14.6%
Establishment dollar value of shipments	<0.1%	7.0%	37.6%	69.5%	49.7%	1.2%	16.0%
Establishment total weight of shipments	<0.1%	7.1%	38.4%	69.9%	49.7%	1.2%	16.0%

Notes: These tables report shares of upstream establishments' shipments that are internal to their firm. The sample consists of 67,500 establishment-years aggregated from about 6.3 million shipments. For data confidentiality reasons, the reported percentiles are averages of immediately surrounding percentiles, e.g., the median = 0.5 × (forty-ninth percentile + fifty-first percentile).

Specification/sample	Percentile 50th	75th	90th	95th	Fraction = 0	Fraction = 1	Weighted mean	Approx. *N*
Panel B. Robustness checks (*percent share of dollar value shown*)								
1. At least median number of shipments	0.2	6.9	31.7	59.5	45.5	0.3	16.1	34,100
2. No exporters	<0.1	8.6	46.5	78.3	49.7	1.6	19.8	47,400
3. Shipments to any estab. in firm are internal	4.9	25.1	67.5	90.6	22.8	2.6	24.1	67,500
4. County, not zip, determines internal	7.2	39.8	87.1	98.8	25.3	4.2	34.9	67,500
5. 25 least differentiated industries	0.0	2.5	20.0	48.6	61.4	0.6	7.9	2,200
6. Manufacturers in the sample, manuf. can only be upstream of manuf.	0.0	2.8	21.1	49.6	59.0	0.9	7.7	26,000
7. Industries with a prior of high internal shares	4.9	33.8	69.4	86.0	35.7	1.8	25.2	3,900
8. 5 percent cutoff definition for VI	0.0	5.1	32.1	63.3	53.9	0.9	12.1	52,700
9. Remove $I \rightarrow I$ as a potential vertical link	0.0	3.9	30.8	60.7	58.7	1.0	9.8	42,800

Notes: Each row shows, for a different subsample, the distributions of the shares (by dollar value) of upstream integrated establishments' shipments that are internal to the firm. The criteria for inclusion in and size of each subsample are discussed in the text. For data confidentiality reasons, the reported percentiles are averages of immediately surrounding percentiles, e.g., the median = 0.5 × (forty-ninth percentile + fifty-first percentile).

by dollar value and weight are even smaller, at less than 0.1 percent. Almost half of the establishments report no internal shipments at all. Even the ninetieth percentile establishment ships over 60 percent of its output outside the firm.

The exception to this general pattern is the small set of establishments that are clearly dedicated to serving the downstream needs of their firm, the 1.2 percent of the sample that reports exclusively internal shipments. The unusualness of this specialization is even more apparent in the histogram of establishments' internal shipment shares shown in Figure 1. The histogram echoes the quantiles reported above: the vast majority of upstream establishments make few internal transfers. The fractions of establishments fall essentially monotonically as internal shipment shares rise—until the cluster of internally dedicated establishments. Another factor in the unusualness of these internal specialist establishments that is not apparent

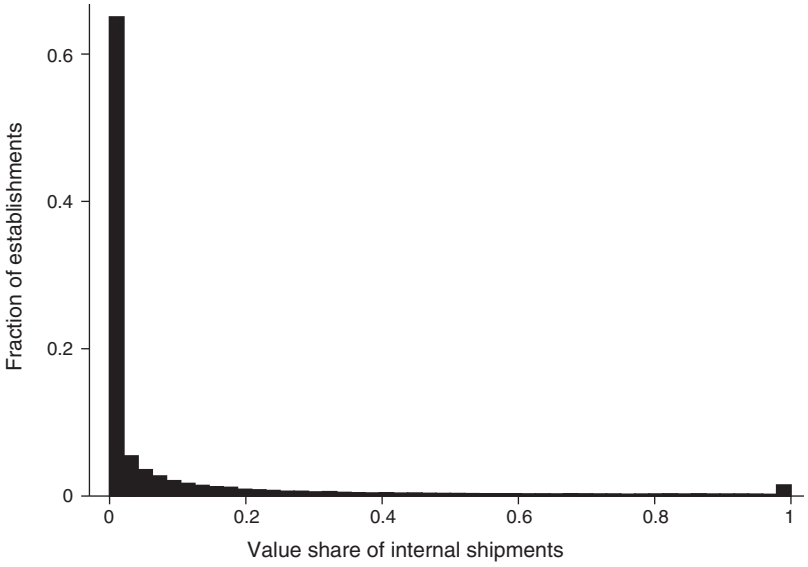


Figure 1. Share of Intrafirm Shipments by Upstream Vertically Integrated Establishments

in the histogram is that they are larger on average. This, along with the internal share distribution being highly skewed, explains why the aggregate internal share of upstream establishments' shipments (the across-establishment sum of internal shipments divided by the across-establishment sum of total shipments) is 16 percent. This is well above the median share across establishments. Thus, internal shipments are more important on a dollar-weighted than an ownership decision–weighted basis but are the exception in either case.[18]

These results imply that the traditional view that firms choose to own establishments in upstream industries to control input supplies may be off target. Clearly, other motivations for ownership *must* apply for those establishments making no internal shipments. Even for those that do serve their own firms, though, their typically small internal shipments suggest that this role may not be primary.[19]

B. *Robustness Checks*

The disconnect between the upstream establishments and their downstream partners, at least in terms of physical goods transfers, is stark and perhaps surprising. We conduct several robustness checks to verify our benchmark results.

First, it is appropriate to review some details of how the Commodity Flow Survey is conducted, specifically with regard to its ability to capture intrafirm shipments.

[18] The distinction between the median internal share and the value-weighted mean internal share mirrors a difference, in the context of international trade, between Bernard et al. (2010) and Ramondo, Rappaport, and Ruhl (2012). Ramondo, Rappaport, and Ruhl show that the bulk of cross-border related-party shipments are due to a small number of very large multinational firms. So, while most multinationals have a small amount of intrafirm flows, the share of international trade occurring through related parties is large.

[19] It is possible in some production chains that an upstream establishment could completely serve its firm's downstream needs with only a small fraction of its output. We show that this possibility is not driving our results in online Appendix D.3, however.

The CFS seeks to measure these shipments, and it makes no distinction between intra- and interfirm transfers in its definition of "shipment." In fact, the survey instructions (US Census Bureau 1997) state explicitly that respondents should report shipments "to another location of your company," save for incidental items like "interoffice memos, payroll checks, business correspondence, etc."

There are several reasons to believe the implied shipments totals are accurate. First, the Census Bureau audits responses by comparing the establishment's implied annual value of shipments from the CFS with that from other sources. If the disparity is well beyond statistical variance, the Bureau contacts the respondent and reviews the responses for accuracy. If integrated establishments were systematically underreporting internal shipments because of confusion or by not following directions, the auditing process would help catch this.

In addition, for establishments in the manufacturing sector, there is an independent measure of internal shipments. The Census of Manufacturers collects data on what it terms establishments' interplant transfers, shipments that are sent to other establishments in the same firm for further assembly. These interplant transfers represent part, but not all, of our internal shipments measure—for example, shipments to wholesalers or retailers are not included in CM interplant transfers.[20] In addition to the difference in definition, these measures are collected using separate survey instruments (often likely to have been filled out by different individuals at the establishment). Despite these differences, we find a strong correlation between the two measures. The correlation coefficient between establishments' log interplant transfers and our CFS–based estimate of internal shipments is 0.52 across our matched sample of about 37,000 establishment-years, and a regression of the latter on the former yields a coefficient of 0.470 (standard error = 0.011).

Robustness: Sample.—In our first series of robustness checks, we consider the impact of modifications to our core sample of upstream vertically integrated establishments. The corresponding distributions of establishments' internal shipments are shown in Table 1, panel B. Each row is a separate check. We show only the distributions of the dollar value shares for the sake of brevity; similar patterns are observed in the shares by shipment counts or total weight.

The robustness check in the first row of panel B uses only establishments reporting at least the median number of shipments (101 shipments) across all establishments in the sample. The point is to exclude those for which sampling error could be higher and for whom extreme values like zero are more likely. This leaves us with a sample of about 34,000 establishment-years making just over 4.2 million shipments. (This is greater than half the establishment-years in the benchmark sample because several establishments report exactly the median number of shipments.) Extreme values are, in fact, rarer in this sample: 45.5 percent report making no internal shipments, down from 49.7 percent in the full sample, and 0.3 percent report exclusively internal shipments, down from 1.2 percent. The remainder of the distribution is not

[20] Restricting shipments to those that are sent for further assembly has a substantial impact on the estimate of establishments' internal shipments. We estimate in online Technical Appendix D.1 that half of our measured internal shipments from manufacturing establishments are sent to establishments outside of the manufacturing sector (and, thus, are not for further assembly).

much different, however. The median fraction of internal shipments is 0.2 percent, and the ninetieth percentile establishment is less likely to ship internally than that in the full sample.

The second check drops any establishment that reports any shipments for export. In the CFS, the destination zip code of shipments for export is for the port of exit, with a separate note indicating the shipment's export status and its destination country. Thus internal shipments to a firm's overseas locations would be misclassified as outside the firm, unless by chance the firm has a downstream establishment in the port's zip code. Focusing on the roughly 47,000 establishments reporting no exports among their roughly 4.3 million shipments avoids this potential mismeasurement. The results are in the second row of panel B of Table 1. The entire distribution is close to the benchmark results above, with the median internal share being less than 0.1 percent and 49.7 percent of establishments reporting zero intrafirm shipments. Missing export destinations are not the source of our results.

The next check counts shipments destined for the zip code of *any* establishment in the same firm as internal, not just those going to locations of downstream links of vertical chains. It is possible that some vertical production may occur outside those chains we identify using the Input-Output Tables. Here, we are taking the broadest possible view toward defining intrafirm transfers of physical goods along a production chain. As seen in the third row of panel B, all quantiles have internal shipment fractions higher than the benchmark, as they must. Still, the median internal share is only 4.9 percent, and the ninetieth percentile is 67.5 percent. About 23 percent of establishments still have no shipments to a zip code of any establishment in their firm, and exclusively internal establishments make up 2.6 percent of the establishments of the benchmark sample.

In the fourth check we make the generous assumption that a shipment is internal if it goes to any *county* in which the firm has a downstream establishment. While unrealistic, this approach accounts for almost any problems with zip code reporting errors or missing zip codes. The results of this exercise are in row 4 of panel B. Not surprisingly, the shares of shipments considered intrafirm are considerably higher, given the easier criterion for being defined as internal. There are more internal specialists or near-specialists: the ninetieth-percentile internal share is 87 percent, and 4.2 percent of establishments report only internal shipments. Even so, a substantial fraction of establishments—25 percent, more than five times the number of internal specialists—report no shipments to counties where downstream establishments in their firms are located. The median internal share across establishments is 7.2 percent.

The fifth check restricts the sample to establishments in the 25 manufacturing industries with the least amount of product differentiation, as measured by the Gollop and Monahan (1991) product differentiation index. The concern is that even our detailed industry classification scheme may be too coarse to capture the true extant vertical links. For instance, it might be that while two industries are substantially linked at an aggregate level, this actually reflects the presence of, say, two separate vertical links within a four-digit SIC industry. In this case, we would not expect many shipments to go from upstream establishments in one link to downstream establishments in another, even though we might infer the two are vertically linked just from comparing the industry-level trade patterns.

Selecting industries with undifferentiated products should reduce product heterogeneity within detailed industries and raise the probability that the industry links we identify as described above hold at a disaggregate level. There are about 2,200 establishment-years in this subset of industries in the CFS. We find that internal shares are actually lower for establishments in the less differentiated industries. The median establishment has no internal shipments, while the ninetieth-percentile establishment's internal share is 20 percent.

A sixth check pertains to wholesale establishments that neither physically receive nor send goods shipments. These establishments—referred to by the Census Bureau as manufacturers' sales offices—instead only prepare the paperwork necessary to market and coordinate their manufacturers' shipments. Manufacturers' sales offices are quantitatively important: in 1997, these establishments' sales were valued at $765 billion (US Census Bureau 2000). Because of the existence of these establishments, our benchmark sample contains some manufacturers that we are classifying to be at the upstream end of a vertical link, but that actually have no same-firm downstream establishments that can actually receive their shipments. For this subset of manufacturers that we would be spuriously including in our benchmark sample, it should be no surprise that the measured share of internal shipments is small. [21]

To assess the significance of this concern, we focus on the manufacturers that are upstream of other manufacturing establishments. We alter the definition of pairs of vertically linked industries to include only manufacturers that are upstream of other manufacturers. This subsample will completely avoid any possible problem with wholesale establishment classification. We report in row 6 that, for the 26,000 manufacturing plant-year observations in this subsample, the median plant has no internal shipments. The value-weighted average internal share is 7.7 percent. The fact that the internal shares are not much different when we focus on manufacturing-to-manufacturing vertical links indicates that the manufacturer's sales offices issue is not skewing our main results.

There is substantial heterogeneity across industries in the share of internal shipments. An additional check of our methodology is to compute the average internal shares for industries, such as automobile part manufacturers or petroleum refiners, for which we have a prior belief that internal shares are important. In particular, we compute the internal shares of the industries reviewed in Lafontaine and Slade (2007). To the extent that these industries were initially chosen as subjects of study because of the prevalence of internal shipments, our measured internal shares should be exceptionally high. For the 12 four-digit industries mentioned in Lafontaine and Slade (2007)—surface mining of coal, underground mining of coal, soft drink bottling, crude oil refining, cyclic crudes and intermediates, other industrial organic chemicals, men's footwear, cement, auto parts manufacturers, aerospace parts manufacturers, bulk petroleum wholesalers, nonbulk petroleum wholesalers—25 percent of shipment value was internal to the firm. The fiftieth- and seventy-fifth-percentile

[21] Consider the example of a firm owning two establishments, one in auto assembly (SIC 3711) and the second in auto wholesale (SIC 5010). Our methodology would identify the auto assembler to be at the upstream end of a vertical link. If the auto wholesaler is a manufacturers' sale office, one should *not* expect shipments from the upstream plant to stay within the firm.

internal shares were 4.9 percent and 33.8 percent, respectively (see row 7).[22] Furthermore, not only do establishments in the Lafontaine and Slade (2007) industries have higher internal shares, conditional on vertical integration status, but these establishments are also more likely to be a part of a vertical structure in the first place. Of the establishments surveyed in the Commodity Flow Survey, 42 percent are in our benchmark sample. For the subset of establishments in the Lafontaine and Slade (2007) industries, 67 percent are included in our sample of upstream establishments.[23] In summation, our algorithm yields higher internal shares for the establishments in the industries for which we have a prior, based on previous studies, that vertical integration is motivated by the flow of physical inputs. This gives us confidence that our algorithm is correctly identifying the low internal shares for the other industries in the sample.

The remaining robustness checks in the panel explore the impact of varying the definition of vertically linked industries. Row 8 of the table shows the results using a 5 percent cutoff, while row 9 keeps the 1 percent cutoff, but removes the possibility that an industry can be vertically linked with itself. Both of these robustness checks reduce the number of establishments that are defined to be at the upstream end of a production chain. The 5 percent cutoff sample contains about 53,000 establishment-years and 5.0 million shipments, while the "No $I \rightarrow I$" rule produces a sample with about 43,000 establishment-years and 4.0 million shipments. In both of these subsamples, the median and ninetieth-percentile internal shares are slightly smaller than in the benchmark.

All in all, our benchmark results appear robust to several sample and variable definition changes. Additional robustness checks along these lines are provided in online Technical Appendix D.1.

Robustness: Accounting for Actual Downstream Use.—We measure an upstream establishment's internal shipments above as a share of its total shipments. There are cases where this ratio might be misleading as to the extent of intrafirm product movements. Consider a hypothetical copper products company with two establishments: an upstream mill that produces copper billets and a downstream establishment that processes billets into pipe. Suppose the downstream establishment needs \$10 million of billets to operate at capacity. Now say the upstream mill produced \$100 million of billets in a year. If the mill shipped \$10 million of billets to the pipe-making establishment and the remaining \$90 million elsewhere, we would compute its internal shipment share as 10 percent. Yet the firm would be completely supplying its downstream needs internally. The difference in the scales of operations between upstream and downstream establishments creates this misleading internal share.

While this may raise the question of why the firm wouldn't own enough pipe establishments to use its upstream production, in this section we create an alternative measure of internal shipment shares that can account for inherent differences in operating scales across industries. Instead of using upstream establishments' total

[22] We were unable to report results for many of the industries identified in Lafontaine and Slade (2007) because these industries were in the service sector or there were too few observations in our dataset to pass Census data-confidentiality regulations. See online Technical Appendix E for a discussion of the industries in this subsample.

[23] These figures are 59 percent and 84 percent, respectively, when establishments are weighted according to the value of their shipments.

shipments as the denominator in the internal shipment share measure, we instead calculate firms' downstream use of products they make upstream. We then construct internal shipments shares as intrafirm shipments of upstream establishments divided by the *minimum* of two values, either the firm's total upstream shipments as above or the firm's reported downstream use of the upstream product. Hence, the internal share of the hypothetical copper firm above would be 100 percent, rather than 10 percent as before, because the firm provides all the copper it uses downstream.

While the CFS offers a random sample of establishments' shipments, we unfortunately do not have a random sample of establishments' incoming materials. This precludes us from directly measuring "internal purchase shares" in the same way we measure internal shipment shares. But for a subset of firms we can construct internal shipments as a fraction of downstream use. To do so, we must first restrict our CFS sample to those where we observe *all* the upstream establishments of a firm, at least for a given product. If firms served downstream needs from upstream establishments not in the CFS, we would not observe their non-CFS establishments' shipments and therefore would not know they are internal. Hence, we look here only at CFS establishments where we observe all of the firm's establishments in a particular industry.[24] We use the Economic Census to find this subset of establishments, which ends up being about 11,000 establishment-years. If we calculate these shares as before, this subsample looks similar to the entire sample. For example, 53.8 percent of these establishments report making no internal shipments, and the ninetieth percentile establishment ships 36.5 percent of its output internally.

We then match these upstream establishments' shipments to downstream usage within the firm. We construct three downstream usage measures. The first simply aggregates the materials purchases of all of the firm's downstream manufacturing establishments. These purchases are reported by every establishment in the Census of Manufactures. The firm's downstream use of upstream products is simply the sum of all its intermediate materials purchases. We can compute these downstream use measures for about 4,400 firm-year observations. To compute internal shares, we add up the internal shipments of the firms' upstream establishments to use as the numerator.[25]

The second measure of downstream usage matches upstream shipments to downstream usage by product. We use the detailed materials purchase information from the Census of Manufactures materials supplement, which collects establishments' materials purchases by detailed product. We compute each firm's upstream

[24] Observing all of the establishments in a given industry isn't exactly sufficient for this particular robustness check. Even in cases for which all upstream establishments are sampled in the CFS, we won't observe all of the upstream shipments, since each survey respondent reports only a sample of the shipments that it makes.

[25] There are two measurement problems with this first approach that will tend to bias our internal shares measures in opposite directions. First, because we required only that we observe all of a firm's establishments making a *particular product* in the CFS, we might be missing internal shipments from firms' other upstream establishments (this is much less of a problem in our other two downstream use measures below, since they are matched by firm-product, rather than just by firm). This will cause us to understate the true internal shipment share. The second measurement issue arises because we can only observe materials purchases for downstream establishments in the manufacturing sector. If some upstream products are used in the firms' nonmanufacturing establishments, we will not include these in our downstream usage measures. This will lead us to overstate internal shipment shares. As a practical matter, both of these measurement concerns are probably second order. Our restricted sample has a large fraction of firms with only a few establishments. So, if a firm's upstream establishments are in the CFS and its downstream establishments in manufacturing, it is likely those represent all the establishments the firm owns.

Table 2—Internal Shipments as the Share of the Smaller of Upstream Shipments or Downstream Usage

Downstream usage measure	Value share of shipments: Percentiles			
	50th	75th	90th	95th
Firm's total downstream manufacturing materials purchases	0.3%	13.8%	67.4%	134.3%
Firm's downstream use of two-digit product	0.0%	15.4%	118.8%	403.2%
Firm's downstream use of four-digit product	0.0%	18.5%	125.4%	687.0%

Notes: This table reports shares of upstream establishments' shipments that are internal to their firm, as a fraction of the smaller of (i) the total shipments of a firm's upstream establishments or (ii) the firm's downstream use of a product. Sample construction and sizes are detailed in the text. For data confidentiality reasons, the reported percentiles are averages of immediately surrounding percentiles, e.g., the median $= 0.5 \times$ (forty-ninth percentile + fifty-first percentile).

shipments by product using the shipment commodity codes available in the CFS. Product-specific shipments are computed at the two-digit level. (We use only 1993 CFS data here because a change in the commodity coding scheme made it difficult to match the 1997 CFS commodity codes with the materials codes in the Census of Manufactures.) We sum the same firm's reported downstream use of that two-digit product from the Census of Manufactures. The internal shipment share is the ratio of the firm's internal shipments of the product divided by its reported downstream use of that product. We are able to match approximately 5,500 firm-material combinations.

The third and final measure of downstream materials usage repeats the procedure above, except matches at the more detailed four-digit product level. Because the greater detail makes finding matches less likely, we have a sample of about 2,400 such firm-product combinations.

The results from these exercises are shown in Table 2. Recall that we now compute internal shipments as their share of the *smaller* of (i) the firm's (or firm-product's) total upstream shipments or (ii) the firm's downstream usage. Again, only the dollar-value shares are shown for brevity. The first row shows shares computed using the firm-level match where internal materials usage is aggregated across all materials. The second row shows results from the sample of matched firm-products at the two-digit level; the third shows the firm-product match sample at the four-digit level.

All three measures of downstream usage still imply that most vertical ownership structures are not about serving the downstream material needs of the firm. The median share across establishments of internal shipments as a fraction of the smaller of the firm's upstream shipments and its downstream use is 0.3 percent in the first (firmwide) downstream use measure. The share of this subsample reporting zero internal shipments is 44.4 percent. For the second measure of internal usage (firm-product matching at the two-digit level), 60.2 percent of the firms report no internal shipments. For the third measure (firm-product matching at the four-digit level), 65.3 percent of the sample report no internal shipments.

One thing to note about the results is that some shares are above one. It is possible that this reflects in part the fact that we classified all upstream establishments' shipments as internal if their destination zip code was where the firm owned a downstream establishment; in fact, some of these shipments may have gone to an establishment not owned by the firm, but in the same zip code. But probably some of these

shares reflect measurement error in firms' downstream materials use. For instance, if the firm is outside the manufacturing sector, we may not be able to observe it. A summary measure of the extent of such measurement error is the fraction of observations with implied internal usage ratios above one. For the three downstream use measures above, these shares are 6.7, 11.7, and 12.5 percent, respectively.

Thus, the small internal shares we were finding before do not seem to simply reflect the fact that most integrated structures have considerably larger upstream than downstream establishment scales. In fact, we still find a large number of cases (slightly under one-half of the sample) without any intrafirm shipments. In other words, we know a firm makes a particular product upstream, uses that same product as an input downstream, but does not ship any of its own upstream output to its downstream units.

Shipments of Establishments that Make Firms Become Vertically Structured.—We next look at the internal shipment patterns for a very select subset of establishment-years in our sample. These observations have two properties. First, they correspond to newly vertically integrated establishments on the upstream end of a production chain (they were single-unit firms in the previous Economic Census). Additionally, these establishments have been acquired by firms that, concurrent with the purchase, begin owning establishments in a vertical production chain for the first time. In other words, these are the establishments that *make* these firms vertically structured. These establishments might provide one of the clearest windows into any connection between why firms expand vertically and their internal shipment patterns. Because of the narrow selection criteria, the subsample is small—a total of just over 300 establishment-years in the CFS, reporting about 28,000 shipments—but still offers enough leverage to make a meaningful comparison to the overall patterns discussed above.

This subsample exhibits an even lower prevalence of internal shipments than in the benchmark sample. Sixty-eight percent of these establishments report no internal shipments at all, and the ninetieth percentile of internal shipments is only 10.1 percent. Because the small sample raises questions of whether these differences are statistically significant, we also estimate regressions that project establishments' intrafirm shipment shares on an indicator for these new-VI establishment/firm units and full set of industry-year dummy variables. The estimated coefficient on the subsample indicator in the dollar-value-share regressions is -0.057 (standard error $= 0.009$). (The coefficient is also significantly negative when shares of shipment counts or when weights are used as the dependent variable.) These establishments do in fact have significantly lower internal shipments shares.

Thus even for establishments acquired expressly as part of a firm's move to build a vertically integrated ownership structure, internal sourcing of physical inputs is unusual.

Other Robustness Checks.—We conducted additional, detailed robustness checks on the benchmark results that, for the sake of brevity, we detail in online Technical Appendix D. One explores whether the observed small internal shipment shares reflect the fact that establishments in vertical ownership structures are spaced further apart geographically than is typical. We show this is not the case; in fact, even

vertically structured firms with all of their establishments in a single county have internal shares similar to those in the broader sample. A second robustness check asks whether our definition of vertical ownership, which by necessity requires a firm to operate the upstream and downstream stages of production in separate establishments, misses vertically integrated production practices occurring within a single establishment (and therefore undercounting the within-establishment "shipments" that accompany them). We find no evidence that this is driving our result.

IV. Explanations for Vertical Ownership

The lack of movement of goods along production chains within most vertically structured firms appears to be a robust feature of the data. As mentioned above, we propose that vertical ownership is instead typically used to facilitate movements of intangible inputs, like management oversight across a firm's production units. In this section we document additional facts that are consistent with this theory.

A. *Firms as Outcomes of an Assignment Mechanism*

We first show evidence that establishments' vertical ownership structures are systematically related to persistent differences in establishment "types"—combinations of idiosyncratic demand and supply fundamentals that affect establishment profitability in equilibrium. Further, these type differences primarily reflect "selection" on preexisting differences rather than "treatment" effects of becoming part of a vertical ownership structure. At the same time, we find that these type differences aren't much tied to vertical ownership itself, but rather to being in large firms of any structure. We discuss below how these patterns are all consistent with theories of the firm as the outcome of an assignment mechanism that allocates tangible and intangible assets among heterogeneous firms. Models of such mechanisms—which include Lucas (1978); Rosen (1982); and, more recently, Garicano and Rossi-Hansberg (2006) and Garicano and Hubbard (2007)[26]—offer an explanation for why we might not see many internal shipments within vertical ownership structures while at the same time pointing us toward an alternative explanation for such ownership patterns: namely, facilitating the flow of intangible inputs within the firm.

Establishments in Vertical Ownership Structures Are High "Type" Establishments.—We first focus on the patterns of establishment-level measures of "type" across vertically integrated and unintegrated establishments. We use four measures to proxy for establishment type.[27] They are not independent, but they differ enough in construction to allow us to gauge the consistency of our results. Two are productivity measures that differ in their measure of inputs: output per worker-hour and total factor productivity (TFP). (Both are expressed as the log of the establishment's output-input ratio.) Our third type measure is simply the

[26] These models are in turn built on foundations laid out earlier by Koopmans (1951) and Becker (1973).

[27] Foster, Haltiwanger, and Syverson (2008) present a model of industry equilibrium where producers differ along both demand and cost dimensions, and show that establishment type can be summarized as a single-dimensional index of demand, productivity, and factor price fundamentals.

establishment's log real revenue. The fourth metric is the establishment's log capital-labor ratio (capital stock per worker-hour). Further details on the construction of these measures are given in online Technical Appendix A. Because of data limitations, we can construct these measures only for the roughly 350,000 establishments in each Census of Manufactures.

These empirical type measures have been shown in various empirical studies to be correlated with establishment survival. Survival probabilities reflect establishment type 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 Syverson (2011) for a recent literature review. Establishment 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). Hence, they capture the connection between establishments' supply and demand fundamentals and the establishments' profit and survival prospects.

We first compare establishment type measures across integrated and unintegrated producers by regressing establishment types on an indicator for establishments' integration status and a set of industry-by-year fixed effects. The coefficient on the indicator captures the average difference between establishments in and out of vertical ownership structures. By including fixed effects, we are identifying type differences across establishments in the same industry-year, avoiding confounding productivity, scale, or factor intensity differences across industries and time. We estimate this specification for each of the four establishment type proxies and report the results in Table 3, panel A.[28]

It is clear that establishments in vertical ownership structures have higher types. They are more productive, larger, and more capital intensive. Their labor productivity levels are on average 40 percent higher ($e^{0.337} = 1.401$) than their unintegrated industry cohorts. These are sizable differences. Syverson (2004) found average within-industry-year interquartile log labor productivity ranges of roughly 0.65; the gaps here are half of this. TFP differences, while still positive and statistically significant, are much smaller, at 1.3 percent. Vertical establishments are much larger—4.2 times larger—than other establishments in their industry in terms of real output. Capital intensities are substantially higher in integrated establishments as well, explaining why their labor productivity advantage is so much bigger than the average TFP difference.

A natural question that follows from these results is the nature of vertically linked establishments' type differences. There are three possibilities, and they are not mutually exclusive. The gaps could reflect the fact that newly built establishments under vertical ownership are different from newly built establishments in other ownership structures, and because types are persistent, this is reflected in the broader population. It may also be that high-type firms that seek to merge new establishments into their internal production chains choose establishments that already have high types

[28] Sample sizes differ across the specifications because not all of the necessary variables are available for constructing each proxy measure for every establishment-year observation.

Table 3—Establishment Attributes by Vertical Ownership Structure

	Output per hour	TFP	Output	Capital-labor ratio
Panel A. Within-industry differences				
Indicator for vertical establishments	0.337*	0.013*	1.443*	0.424*
	(0.002)	(0.001)	(0.004)	(0.003)
Approximate observations	970,000	879,000	991,000	937,000
Approximate observations [vertical establishments]	232,000	219,000	237,000	228,000
Panel B. Differences among new establishments				
Indicator for vertical establishments	0.281*	0.032*	1.228*	0.330*
	(0.004)	(0.003)	(0.009)	(0.006)
Approximate observations	240,000	213,000	248,000	233,000
Approximate observations [vertical establishments]	42,000	38,000	43,000	41,000
Panel C. Comparing unintegrated establishments: to-be-vertical versus remaining nonvertical				
Indicator for to-be-vertical establishments	0.197*	0.002	1.258*	0.246*
	(0.005)	(0.003)	(0.010)	(0.007)
Approximate observations	403,000	367,000	410,000	390,000
Approximate observations [to be vertical]	16,000	15,000	16,000	16,000
Panel D. Changes upon entering vertical ownership				
Newly vertical indicator	0.034*	−0.009*	0.015*	0.033*
	(0.005)	(0.004)	(0.007)	(0.009)
Approximate observations	348,000	300,000	356,000	327,000
Approximate observations [newly vertical]	16,000	15,000	16,000	16,000

Notes: This table shows establishment "type" comparisons between establishments in (or to-be-in) vertical ownership structures and their nonvertical counterparts. Panel A compares across all establishments for which type measures are available. Panel B compares new establishments. Panel C compares prior period types among nonvertical establishments that will become part of vertical ownership structures by next period to those remaining nonvertical. Panel D compares changes in type for establishments that become part of vertical ownership structures to changes for unintegrated establishments that remain so. All regressions include industry-year fixed effects; industries are defined according to the BEA's IOIND classification. Samples are comprised of non-administrative-record manufacturing establishments. See text and online Technical Appendix A on the construction of type measures and additional details.

*Significant at the 5 percent level.

to add to the firm. Finally, becoming part of a vertical ownership structure might be associated with a change in an existing establishment's type.

We can separately investigate these possibilities. To see if new vertically structured establishments are different from newly built establishments in other ownership structures, we reestimate the type specification above on a subsample that includes only new establishments.[29] To test if firms already composed of high-type vertically linked establishments expand by purchasing unintegrated establishments that already have systematically higher types, we regress unintegrated establishments' type proxies on a dummy variable indicating if an establishment will *become*

[29] New establishments are defined as those appearing for the first time in the Economic Census, which is associated with the start of economic activity at its particular location. In other words, these establishments are greenfield entrants. Existing establishments that merely change industries between ECs are not counted as entrants in our sample. New establishments are an important part of the formation of vertically integrated structures in the economy: Entering integrated establishments account for roughly two-fifths of the employment, and three-fifths of the capital stock, of all new establishments in a given EC. This specification excludes observations from the 1977 EC because of censored entry.

vertically integrated by the next Economic Census. (Again, industry-year fixed effects are included.) The estimated coefficient on the dummy variable captures how soon-to-be-vertically-owned establishments compare *before integration* to other establishments in their industry that will not become integrated during the period. Finally, to test if becoming part of a vertical ownership structure is associated with systematic changes in an establishment's type, we regress the intercensus growth in establishments' type measures on an indicator for establishments that become part of integrated production chains during the period. All these specifications include industry-year fixed effects, so we are always comparing establishments within the same industry and time period.

Panels B through D of Table 3 show the results, with panel B comparing new establishments, panel C comparing the types of unintegrated establishments before integration, and panel D comparing establishment type changes. Comparing the type disparities in these panels to those in panel A suggests that much of the heterogeneity between establishments in and out of vertical ownership structures reflects differences in the assignment of establishment types to integration status. As panels B and C show, most of the vertically integrated establishments' higher productivity levels, scale of operations, and capital intensities already existed either when they were born into integrated structures or before they were merged into integrated structures. For example, labor productivity and capital intensity are on average about 30 percent higher for new establishments in vertically integrated structures than for other new establishments. This is about three-fourths of the analogous gap observed among all establishments. Similarly, unintegrated establishments that will soon become part of vertical ownership chains are already considerably more productive, larger, and more capital intensive than unintegrated establishments that will remain so. Thus, most of the differences observed in panel A of the table reflect "selection" effects. At the same time, the results in panel D make clear that, for labor productivity and capital intensity in particular, those gaps not accounted for by preexisting differences in type are closed due to the faster growth in experience by existing establishments when they become integrated. Thus, we cannot ignore the possibility that integration has some direct effects on establishment types.[30]

Firm Size, Not Structure, Explains Most Establishment Type Differences.—The fact that establishments in vertical ownership structures are different naturally leads to the question of whether *firms* with vertical structures are different. And indeed, as we show in online Technical Appendix D.5, firms with vertical ownership structures are larger on average (whether measured by total employment or revenues) than other firms with multiunit organizational structures, be it those that own multiple establishments in a single industry or those that own establishments in multiple industries, none of which comprise substantial vertical links as defined above.

Given that firms with vertical structures tend to be the largest, it's natural to ask whether the differences in establishment types seen above simply reflect underlying

[30] These are, of course, general patterns across the hundreds of manufacturing industries in our sample. They do not imply that the relative importance of these sources of type differences doesn't vary across individual industries. It is possible that in certain industries most of the type differences reflect changes that occur when establishments become integrated rather than preexisting type dissimilarities.

Table 4—Establishment Type Differences, Controlling for Firm Size

	Output per hour	TFP	Output	Capital-labor ratio
VI indicator	0.050*	0.007*	0.237*	0.049*
	(0.004)	(0.003)	(0.009)	(0.006)
Approximate observations	231,000	220,000	235,000	227,000
Approximate observations [VI indicator]	195,000	185,000	199,000	191,000

Notes: This table shows the results from regressing establishment-level type measures on an indicator for vertically integrated establishments, a set of industry-year fixed effects, and control variables for firm size; industries are defined according to the BEA's IOIND classification. The sample consists of non-administrative-record manufacturing establishments in multiindustry firms. The firm size control variables include quintics of several measures of the establishment's owning-firm size: (log) employment, (the logarithm of) the number of establishments, and (the logarithm of) the number of industries. These firm size measures are computed by summing over the *other* plants in the firm of the establishment in question.

*Significant at the 5 percent level.

differences among firms. That is, if large firms tend to own systematically larger (and more productive, etc.) establishments, this might explain the distinctive type patterns of establishments in vertical structures, rather than their vertical ownership linkages per se. In other words, the high types of establishments in vertical ownership structures may be a function of firm size rather than firm structure.

To see if this is the case, we rerun the establishment type regressions above while including control variables for firm size. We regress establishment type measures on an indicator for vertically integrated establishments and industry-year dummy variables as above, while now adding flexible control variables for firm size. These control variables are quintics of log firm employment, log number of establishments, and the log number of industries in which the firm operates. We restrict the sample to establishments owned by multiindustry firms, but few differences are seen if single-industry firms are also included. This specification lets us compare establishments that are in firms of the same size, regardless of the firms' internal structures.

Table 4 shows the results of these regressions. Much of the correlation between an establishment's type and its vertical ownership structure goes away once we control fully for firm size. The point estimate for establishments' TFP differences is now half as large and is one-eighth as large for revenue differences. The labor productivity and capital intensity "premia" for vertically integrated establishments are now roughly 5 percent, much smaller than the initial 40 to 50 percent differences reported in panel A of Table 3.

Hence, much of what makes establishments in vertical ownership structures different isn't really related to vertical ownership itself. Instead, the largest establishments tend to be in the largest firms, and the largest firms tend to own vertically linked establishments. Accounting for this fully explains the TFP and size differences and most of the labor productivity and capital intensity gaps.[31]

[31] This evokes the result in Hortaçsu and Syverson (2007) that vertically integrated ready-mixed concrete establishments' productivity and survival advantages don't reflect their vertical structure per se, but rather that these establishments tend to be owned by firms with clusters of ready-mixed establishments in local markets. (The clusters allow them to harness logistical efficiencies.) Once we compared vertically integrated concrete establishments to nonintegrated establishments that were also in clusters, many of the differences seen between integrated and nonintegrated establishments disappeared.

Discussion.—The results in this section are consistent with theories of the firm as the outcome of an assignment mechanism that spreads higher-quality intangible inputs (e.g., better managers) across better and/or a greater number of production units. Our explanation parallels theories of the firm as a collection of capabilities (or core competencies), which are ubiquitous within the strategic management literature but may be unfamiliar to many economists.

Wernerfelt (1984) and Prahalad and Hamel (1990) are two relatively early examples within this literature. In these papers, firms' primary choices are not over which products to produce, but instead over which intangible inputs ("resources" for Wernerfelt or "core competencies" for Prahalad and Hamel) to cultivate and exploit. In particular, make-or-buy decisions are not the primary reason for vertical integration. Instead, ownership of establishments in vertically linked industries is a byproduct of firms' exploitation of their core competencies (Prahalad and Hamel).

Grant (1996) is a third example of the resource-based view of the firm. For Grant, a firm's most important resource is its workers' knowledge. The role of the firm is to allow its workers to share their knowledge with one another and to coordinate and aggregate these workers' knowledge (see also Aghion and Tirole 1994).

Finally, Montgomery and Hariharan (1991) provide empirical support for these theories. The authors show that, when firms expand, they enter industries for which the resource (e.g., capital, advertising, or R&D) requirements match the requirements of the industries in which the firm had already been producing.

Note that if the intangible inputs mediation explanation for vertical ownership is correct, the distinction between "downstream" and "upstream" becomes one of convenience rather than an accurate depiction of intrafirm transfers. The names reflect the flow of goods through the physical production process, which may be nonexistent or otherwise very small; they do not necessarily indicate the flow of inputs within the firm. Further, verticality itself need not be an important distinction under this alternative explanation. Vertical firm expansions are simply a particular way in which a firm applies its intangible capital to new but related lines of business. No flows of goods between the firms' vertically related establishments are necessary, just as with a typical horizontal expansion. This is consistent with the result, above, that firm size rather than structure explains most of the average type differences seen across establishments.

B. *Some Evidence that Vertical Structures Facilitate Intangible Input Transfers*

It is difficult to directly test our "intangible input" explanation for vertical ownership structures because such inputs are by definition hard to measure. Ideally, we would have information on the application of managerial or other intangible inputs (like managers' time-use patterns across the different business units of the firm) across firm structures. Such data do not exist for the breadth of industries which we are looking at here, however. That said, we compile some suggestive evidence for an intangible input mechanism in this section.

Our first test digs deeper into the changes seen in establishments that become vertically integrated, as with those observed in panel D of Table 3. We decompose the changes in labor productivity and capital intensity into their respective components by repeating the exercises, but this time running the specifications separately for

Table 5—Changes in Establishment Attributes upon Integration

	Change upon VI
Output per hour	0.025*
	(0.005)
Output	0.008
	(0.006)
Hours	−0.017*
	(0.006)
Capital-labor ratio	0.030*
	(0.009)
TFP	−0.011*
	(0.004)
Capital	0.013
	(0.009)
Production workers	−0.010
	(0.006)
Nonproduction workers	−0.047*
	(0.007)
Nonproduction worker share	−0.006*
	(0.001)

Notes: The table repeats panel D of Table 3, but with additional establishment production measures. The sample consists of only the approximately 265,000 non-administrative-record manufacturing establishments that have nonmissing data for all production measures. See text for details. Regressions include industry-year fixed effects; industries are defined according to the BEA's IOIND classification.

*Significant at the 5 percent level.

establishments' capital stocks and labor inputs. To allow for an exact decomposition of these changes, we restrict the sample to establishments for which we observe each of the production measures, ensuring that the changes in the ratios' (log) components add up to the change in ratios. Furthermore, for reasons that will become clear momentarily, we look at the individual changes in two types of labor inputs: production and nonproduction workers.

The results are shown in Table 5. The 2.5 percent average labor productivity change in this sample is driven both by an insignificant 0.8 percent increase in output and by a 1.7 percent decline in hours. The 3.0 percent increase in capital intensity mostly reflects the same decrease in labor inputs, but the (albeit insignificant) point estimate suggests investment may have been higher at these newly integrated establishments than their nonintegrated counterparts, as capital stocks grew 1.3 percent faster in the former.

The most interesting feature of the observed drop in labor inputs is the labor composition shift that accompanies it. The percentage drop in nonproduction workers is more than four times that in production workers. This is also reflected in the drop in nonproduction workers' share of total employment at the establishment.

These changes in capital intensity and labor composition are consistent with an intangible inputs motive for vertical ownership. Capital intensity would rise upon an establishment becoming part of a vertical link if skilled managerial or other intangible inputs have stronger complementarities with capital than labor, for example. These complementarities may originate from the combination of (i) an assignment of better managers to larger firms, and (ii) the fact that some physical capital inputs

come in large lumps and would be more efficiently spread across a large number of workers (see, for example, Oi and Idson 1999). Alternatively, firms with vertical ownership structures might also face lower effective capital costs, which would shift their optimal factor allocation toward a more capital-intensive orientation. Since we know vertical firms are larger on average, and there is evidence that larger firms might be less credit constrained (e.g., Fazzari, Hubbard, and Petersen 1988; and Eisfeldt and Rampini 2009), this is a plausible alternative.

In addition, the relative decline in nonproduction workers upon integration is consistent with some of the establishment's former management, marketing, R&D, or any other staff associated with providing intangible inputs being replaced with the new intangible inputs of the vertically integrated structure. Fewer workers are needed to provide these new inputs in the integrated structure because of centralization and scale returns or greater efficacy. Both of these changes are consistent with the allocation mechanism we discuss above.[32]

Our next tests look for further circumstantial evidence for intangible input movements by examining changes in the behavior of acquired establishments once they are brought into their new firm. We investigate two practices: the products the establishments manufacture and, taking further advantage of our CFS shipments data, the locations to which establishments send their output.

To explore changes in acquired establishments' product mixes, for each acquired establishment we partition the universe of products into four groups, according to the acquiring and acquired firms' production patterns in the previous Census of Manufactures. Group 1 consists of products that were produced neither by any establishment in the acquiring firm nor by any other establishment in the acquired firm.[33] Group 2 are products that were produced by the acquired firm but not the acquiring firm. Group 3 are products made by the acquiring firm but not the acquired firm, and Group 4 includes products made by both the acquired and the acquiring firms. We then compute the sales of the acquired establishments in each of these four groups in the CMs, both preceding and following the change of ownership.[34] A shift in acquired establishments' product mixes away from Groups 2 and 4 and toward Group 3 would indicate that the acquiring firms reorient the establishments toward the firms' existing operations. This reorientation is likely to require some intangible capital of the acquiring firms, be it production knowledge, product design, customer lists, or the like. As such, the reorientation would be circumstantial evidence for the flow of intangibles.

We present our results in panel A of Table 6. There is a marked shift in the acquired establishments' product mix away from what they did before. While the dollar value of production in these groups drops only slightly, because the acquired

[32] As in panel D of Table 3, measured TFP decreases upon integration. This is somewhat puzzling: the sharing of intangibles within newly vertically integrated firms should manifest itself in TFP growth, not decline.

[33] We do not classify products based on those made by the acquired establishment in question, as we are comparing production patterns before and after acquisition. If we grouped products based on the acquired establishment's production, the establishment's sales of any product in Groups 2 or 4—those groups that include products not made by the acquired firm in the prior CM—would be zero by definition. We similarly exclude the establishment's own shipment destinations in the analogous zip code classifications below.

[34] We define products at the seven-digit SIC level. The sample consists of all manufacturing establishments that are part of a merger or acquisition between 1987 and 1997 and for which we have detailed production data from the Census of Manufacturers production supplement.

Table 6—Allocation of Sales/Shipments across Products and Locations for Acquired Establishments

	Group 1	Group 2	Group 3	Group 4
Panel A. Product mix				
Average sales, CM prior to acquisition (millions)	$10.1	$5.5	$7.0	$4.4
Average sales, CM after acquisition (millions)	$14.3	$5.5	$7.8	$4.3
Fraction of establishment sales, CM prior to acquisition (percent)	37.4	20.5	26.0	16.1
Fraction of establishment sales, CM after acquisition (percent)	44.7	17.2	24.6	13.5
Panel B. Shipment locations				
Average sales, CM prior to acquisition (millions)	$61.5	$15.0	$17.9	$8.8
Average sales, CM after acquisition (millions)	$86.3	$13.7	$25.0	$5.2
Fraction of establishment sales, CM prior to acquisition (percent)	59.6	14.5	17.4	8.6
Fraction of establishment sales, CM after acquisition (percent)	69.3	11.0	20.1	4.2

Notes: This table presents, for acquired establishments in the manufacturing sector, the average dollar amounts and shares of sales accounted for by products (shipment locations in panel B) in four different groups, based on the behavior of the acquiring and acquired firms' establishments in the CM prior to acquisition. Shares are weighted according to the revenue of the acquired establishment. Group 1 contains products (locations in panel B) that neither the acquiring firm's establishments nor the establishments in the acquired firm (other than the establishment in question) produced (shipped to in panel B) in the prior CM. Group 2 contains products (locations in panel B) that the acquired firm's other establishments produced (shipped to in panel B) but the acquiring firm's establishments did not. Group 3 contains products (locations) that the acquiring firm's establishments produced (shipped to) but the acquired firm's other establishments did not. Group 4 contains products (locations) that both the acquiring and the acquired firms' establishments produced (shipped to). Dollar figures are stated in terms of real 1987 dollars, using industry-level price indices from the NBER Productivity database. See text for details.

establishments' sales grew on average (by 18 percent), the combined share of the acquired establishments' products in Groups 2 and 4 falls from 36.6 to 30.7 percent. Also consistent with this reorientation is the fact that the establishments' value of sales of Group 3 products increases by 11 percent. (Although here the share drops slightly because most of the acquired establishments' production growth was in Group 1 products—those made by neither the acquiring firm nor the other establishments of the acquired firm—in the previous CM.)[35]

We show in online Technical Appendix D.7 that these basic data patterns remain present in more structured tests. Specifically, we estimate a logit specification for the probability that an acquired establishment will produce a specific seven-digit product after acquisition as a function of the product mix of the acquiring and acquired firms in the previous CM. The probability an acquired establishment produces a given seven-digit product is significantly and economically larger if the product was made by the acquiring firm in the prior CM.

We conduct a similar exercise looking at changes in the locations to which acquired establishments ship their output before and after acquisition.[36] Again, we partition the acquired establishments' sales into four groups. But here they are based on the locations to which the acquiring and acquired firms shipped prior to the acquisition. Group 1 contains zip codes to which neither the acquiring firm nor any other establishment in the acquired firm shipped before the acquisition. Group 2

[35] Bernard, Redding, and Schott (2010) report substantial turnover in the products that firms produce. Consistent with the results of Bernard, Redding, and Schott (2010), we find that all plants—not only those involved in a merger or acquisition—shift production substantially away from the products other plants in their firm were producing. Similarly, the average establishment stops selling to the locations to which their own-firm establishments had previously been shipping.

[36] Our sample consists of establishments in both the 1993 and 1997 CFS that experienced a change of ownership during that period. The construction of this sample is detailed in online Technical Appendix D.7.

contains zip codes to which other establishments in the acquired firm shipped but no establishments in the acquiring firm did. Group 3 contains zip codes to which the acquiring firm shipped but not the other establishments in the acquired firm, and Group 4 includes zip codes to which both firms shipped output. A shift in acquired establishments' shipping locations away from Groups 2 and 4 and toward Group 3 again suggests a reorientation toward the acquiring firms' existing operations and any intangible capital flows associated with it.

We present these results in panel B of Table 6. The patterns line up with the reorientation story. Both the level and fraction of shipments to zip codes in groups 2 and 4 fall after acquisition. Combined, shipment levels across these two groups fall by 20 percent, and the share going to these two groups drops from 23.1 to 15.2 percent. Concomitant with these drops is an increase in shipments to Group 3 zip codes. Here, shipment levels increase by about 40 percent while their share rises from 17.4 to 20.1 percent. (As with the product mix results, there is an overall increase in reported shipments, mostly coming in Group 1 zip codes.)

We again show using logit regressions in online Technical Appendix D.7 that these basic patterns hold up to more formal testing.[37]

Thus we have seen that acquired establishments have increases in capital intensity driven in large part by reductions in their number of nonproduction workers, a reorientation in their product mix away from their old firm's products and toward their acquiring firms' preexisting product mix, and similar shifts in the destinations of their shipments (and presumably, the identity of their customers as well) away from their old firm's orientation and toward the acquirers'. These patterns are all circumstantial evidence for the flows in intangible inputs that occur within integrated firms. We note, however, that these results are only suggestive—we cannot observe workers' positions within the firm at any finer level than the production/nonproduction worker dichotomy, and we would need much more detailed information on managerial or other intangible inputs to test the theory convincingly. Still, we find the results an intriguing starting point for continued work.

V. Conclusion

We have used data on hundreds of thousands of establishments, the organizational structure of the firms that own them, and their shipments, to explore production behavior in vertical ownership structures. Vertical ownership is not primarily motivated by facilitating the efficient intrafirm transfers of goods along a production chain. Firms' upstream establishments ship only a fraction (and often none) of their output to downstream units inside the firm. This finding is robust to a number of measurement methods. Thus, outside of some exceptional establishments that we find are clearly dedicated to internal production, most vertical ownership appears to have a different reason.

[37] Our results on the reorientation of acquired establishments' operations complement those in Maksimovic, Phillips, and Prabhala (2011). That paper argues that, following a merger or acquisition, the acquiring firm shuts down or sells off establishments outside of the firm's core business segments, while keeping acquired establishments that operate in segments in which the firm already has a large presence or is particularly productive.

Motivated by patterns we document in establishments' "types" within and across firms, we propose an alternative explanation for vertical integration. Namely, that it facilitates efficient transfers of *intangible* inputs (e.g., managerial oversight) within firms. It is plausible that the market would have a more difficult time mediating transfers of knowledge inputs than of physical goods. We provide suggestive evidence in favor of the intangible inputs hypothesis: Acquired establishments begin to resemble—both in terms of their shipment destinations and products produced—establishments from the acquiring firm.

Note that if this explanation is correct, there may not be anything particular about vertical structure within firms; intangible inputs can flow in any direction across a firm's production units. Vertical firm structures and expansions may not be fundamentally different from horizontal structures and expansions. Instead, a more generalized view of firm organization, like the firm as an outcome of an assignment mechanism that matches heterogeneous tangible and intangible inputs, may be warranted, and is consistent with some of the other patterns we document in the data.

REFERENCES

Aghion, Philippe, and Jean Tirole. 1994. "Opening the Black Box of Innovation." *European Economic Review* 38 (3-4): 701–10.

Anderson, Erin, and David C. Schmittlein. 1984. "Integration of the Sales Force: An Empirical Examination." *RAND Journal of Economics* 15 (3): 385–95.

Arrow, Kenneth J. 1975. "Vertical Integration and Communication." *Bell Journal of Economics* 6 (1): 173–83.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. 2014. "Vertical Integration and Input Flows: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.104.4.1120.

Becker, Gary. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–46.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2010. "Intrafirm Trade and Product Contractibility." *American Economic Review* 100 (2): 444–48.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2010. "Multiple-Product Firms and Product Switching." *American Economic Review* 100 (1): 70–97.

Doms, Mark, Timothy Dunne, and Mark J. Roberts. 1995. "The Role of Technology Use in the Survival and Growth of Manufacturing Plants." *International Journal of Industrial Organization* 13 (4): 523–42.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1989. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics* 104 (4): 671–98.

Eisfeldt, Andrea L., and Adriano A. Rampini. 2009. "Leasing, Ability to Repossess, and Debt Capacity." *Review of Financial Studies* 22 (4): 1621–57.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62 (1): 53–82.

Fazzari, Steven M., Robert Glenn Hubbard, and Bruce C. Petersen. 1988. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity* (1): 141–95.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98 (1): 394–425.

Garicano, Luis, and Thomas N. Hubbard. 2007. "The Return to Knowledge Hierarchies." National Bureau of Economic Research Working Paper 12815.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121 (4): 1383–435.

Gibbons, Robert. 2005. "Four Formal(izable) Theories of the Firm?" *Journal of Economic Behavior and Organization* 58 (2): 200–45.

Gollop, Frank M., and James L. Monahan. 1991. "A Generalized Index of Diversification: Trends in U.S. Manufacturing." *Review of Economics and Statistics* 73 (2): 318–30.

Grant, Robert M. 1996. "Toward a Knowledge-Based Theory of the Firm." *Strategic Management Journal* 17: 109–22.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (4): 691–719.

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

Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98 (6): 1119–58.

Hillberry, Russell, and David Hummels. 2003. "Intranational Home Bias: Some Explanations." *Review of Economics and Statistics* 85 (4): 1089–92.

Hillberry, Russell, and David Hummels. 2008. "Trade Responses to Geographic Frictions: A Decomposition Using Micro-data." *European Economic Review* 52 (3): 527–50.

Holmes, Thomas J., and John J. Stevens. 2010. "An Alternative Theory of the Plant Size Distribution with Geography and Intra- and International Trade." National Bureau of Economic Research Working Paper 15957.

Holmes, Thomas J., and John J. Stevens. 2012. "Exports, Borders, Distance, and Plant Size." *Journal of International Economics* 88 (1): 91–103.

Holmström, Bengt. 1999. "The Firm as a Subeconomy." *Journal of Law, Economics, and Organization* 15 (1): 74–102.

Holmström, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7 (1): 24–52.

Holmström, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." *American Economic Review* 84 (4): 972–91.

Holmström, Bengt, and Jean Tirole. 1991. "Transfer Pricing and Organizational Form." *Journal of Law, Economics, and Organization* 7 (2): 201–28.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127–50.

Hortaçsu, Ali, and Chad Syverson. 2007. "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices." *Journal of Political Economy* 115 (2): 250–301.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica* 50 (3): 649–70.

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

Klein, Benjamin. 1996. "Why Hold-Ups Occur: The Self-Enforcing Range of Contractual Relationships." *Economic Inquiry* 34 (3): 444–63.

Klein, Benjamin. 2000. "The Role of Incomplete Contracts in Self-enforcing Relationships." *Revue d'Economie Industrielle* 92 (1): 67–80.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *Journal of Law and Economics* 21 (2): 297–326.

Klein, Benjamin, and Kevin M. Murphy. 1988. "Vertical Restraints as Contract Enforcement Mechanisms." *Journal of Law and Economics* 31 (2): 265–97.

Klein, Benjamin, and Kevin M. Murphy. 1997. "Vertical Integration as a Self-Enforcing Contractual Arrangement." *American Economic Review* 87 (2): 415–20.

Koopmans, Tjalling C. 1951. "Analysis of Production as an Efficient Combination of Activities." *Activity Analysis of Production and Allocation* (13): 33–37.

Lafontaine, Francine, and Margaret Slade. 2007. "Vertical Integration and Firm Boundaries: The Evidence." *Journal of Economic Literature* 45 (3): 629–85.

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9 (2): 508–23.

Maksimovic, Vojislav, Gordon Phillips, and N. R. Prabhala. 2011. "Post-merger Restructuring and the Boundaries of the Firm." *Journal of Financial Economics* 102 (2): 317–43.

Masten, Scott E. 1984. "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics* 27 (2): 403–17.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–725.

Monteverde, Kirk, and David J. Teece. 1982. "Supplier Switching Costs and Vertical Integration in the Automobile Industry." *Bell Journal of Economics* 13 (1): 206–13.

Montgomery, Cynthia A., and S. Hariharan. 1991. "Diversified Expansion by Large Established Firms." *Journal of Economic Behavior and Organization* 15 (1): 71–89.

Oi, Walter Y., and Todd L. Idson. 1999. "Firm Size and Wages." In *Handbook of Labor Economics*. Vol. 3B, edited by Orley C. Ashenfelter and David Card, 2165–214. Amsterdam: Elsevier Science, North-Holland.

Prahalad, Coimbatore K., and Gary Hamel. 1990. "The Core Competence of the Corporation." *Harvard Business Review* 68 (3): 79–91.

Ramondo, Natalia, Veronica Rappoport, and Kim J. Ruhl. 2012. "Horizontal vs. Vertical FDI : Revisiting Evidence from U.S. Multinationals." Unpublished.

Rosen, Sherwin. 1982. "Authority, Control, and the Distribution of Earnings." *Bell Journal of Economics* 13 (2): 311–23.

Simon, H. A. 1951. "A Formal Theory of the Employment Relationship." *Econometrica* 19 (3): 293–305.

Syverson, Chad. 2004. "Product Substitutability and Productivity Dispersion." *Review of Economics and Statistics* 86 (2): 534–50.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49 (2): 326–65.

Teece, David J. 1982. "Towards an Economic Theory of the Multiproduct Firm." *Journal of Economic Behavior and Organization* 3 (1): 39–63.

US Census Bureau. 1997. "Instructions for Completing the Commodity Flow Survey." Washington, DC: US Department of Commerce, Economic and Statistics Administration.

US Census Bureau. 2000. "1997 Economic Census: Wholesale Trade." Washington, DC: US Department of Commerce, Economic and Statistics Administration.

Wernerfelt, Birger. 1984. "A Resource-Based View of the Firm." *Strategic Management Journal* 5 (2): 171–80.

Williamson, Oliver E. 1971. "The Vertical Integration of Production: Market Failure Considerations." *American Economic Review* 61 (2): 112–23.

Williamson, Oliver E. 1973. "Markets and Hierarchies: Some Elementary Considerations." *American Economic Review* 63 (2): 316–25.

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

Williamson, Oliver E. 1979. "Transaction-Cost Economics: The Governance of Contractual Relations." *Journal of Law and Economics* 22 (2): 233–61.

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

Williamson, Oliver E. 1991. "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives." *Administrative Science Quarterly* 36 (2): 269–96.

Vertical Integration and Input Flows
Enghin Atalay, Ali Hortaçsu, and Chad Syverson
Online Appendix

A. Construction of Production Variables

We describe here details on the construction of our production variables. Since the production variable definitions match those previously used in Hortaçsu and Syverson (2007), the descriptions given here will draw heavily on the Data Appendix of that paper.

Output. Establishment output is its inventory-adjusted total value of shipments, deflated to 1987 dollars using industry-specific price indexes from the NBER Productivity Database.

Labor Hours. Production worker hours are reported directly in the CM microdata. To get total establishment hours, we multiply this value by the establishment'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 establishments.

Labor Productivity. We measure labor productivity in terms of establishment output per worker-hour, where output and total hours are measured as described above.

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

$$\mathrm{TFP}_{it} = \mathrm{y}_{it} - \alpha_{lt}\mathrm{l}_{it} - \alpha_{kt}\mathrm{k}_{it} - \alpha_{mt}\mathrm{m}_{it} - \alpha_{et}\mathrm{e}_{it},$$

where the weights α_j are the input elasticities of input $j \in \{l, k, m, e\}$. Output is the establishment's inventory-adjusted total value of shipments deflated to 1987 dollars. While inputs are establishment-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 2-digit industry.

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

Capital-Labor Ratio. Equipment and building capital stocks are establishments' reported book values of each

capital type deflated by the book-to-real value ratio for the corresponding 3-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 establishment are inflated to stocks by dividing by a type-specific rental rate.[1] The total productive capital stock k_{it} is the sum of the equipment and structures stocks. This is divided by the establishments' number of labor hours to obtain the capital-intensity measure used in the empirical tests.

Nonproduction Worker Ratio. Establishments directly report both their number of production and nonproduction employees. Nonproduction workers are defined by the Census Bureau as those engaged in "supervision above line-supervisor level, sales (including a driver salesperson), sales delivery (truck drivers and helpers), advertising, credit, collection, installation, and servicing of own products, clerical and routine office functions, executive, purchasing, finance, legal, personnel (including cafeteria, etc.), professional and technical [employees]. Exclude proprietors and partners." The nonproduction worker ratio is simply such employees' share of total establishment employment.

B. Identifying Pairs of Vertically Linked Industries

The purpose of this section is to detail our methodology of identifying pairs of industries (at the 4-digit SIC level) that are vertically linked to one another. As mentioned in Section II.A of the paper, we classify industry *I* to be upstream of industry *J* if the fraction of shipments from *I* to *J* is greater than some threshold. In the baseline specification, this threshold is set at one percent of the total value sent by establishments in industry *I*. In this section, we describe how we impute the value of shipments sent from SIC industry *I* to SIC industry *J*.

There are two steps to this procedure. First, we must impute how much of each (STCC) commodity *C* was received by any (SIC) industry *J*. Our imputation method is different for *J*'s that are in the wholesale sector, in the retail sector, and in any other sector. The first step is described in Web Technical Appendices B.1, B.2, and B.3. Second, we aggregate over the commodities shipped by each upstream industry, *I*, to generate the estimate of the value of shipments from *I* to *J*. This step is described in Web Technical Appendix B.4.

B.1. Measuring the Flow of Goods through Sectors Other than Wholesale or Retail

For industries outside of the wholesale and retail sector, we start with the 1992 BEA Use Table. This dataset contains information on the amount purchased by different industries of different commodities. Within the BEA Use Table, both industries and commodities are defined according to the BEA's IOIND classification. The main task, for us, is to relate IOIND commodity codes to the Commodity Flow Survey's STCC commodity codes, and to relate IOIND industry codes to the (4-digit) SIC industry codes.

Use $\gamma \in \Gamma$ to refer to IOIND commodities, $\varphi \in \Phi$ to refer to IOIND industries, *C* to refer to STCC commodities, and *I* and *J* to refer to SIC industries. The task at hand is to impute the purchases, P_{CJ}, of commodity *C* by industry *J* using information on the purchases of commodity γ made by industry φ.

[1] 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).

Towards this goal, we use the concordance—between SIC industries and IOIND industries—provided by the BEA.[2] With one minor exception, each 4-digit SIC code can be uniquely matched to a single IOIND industry.[3] We assume that, for the SIC industries J that correspond to the same IOIND industry φ, purchases of γ are proportional to industry J's employment. In other words:

$$P_{\gamma J} = P_{\gamma\phi} \times \frac{Emp_J}{\sum_{J' \in \Lambda(J)} Emp_{J'}}$$

In this equation, Emp_J refers to the total employment in SIC-industry J, and $\Lambda(J)$ denotes the set of SIC industries, J', that correspond to the same IOIND industry as SIC industry, J. The presumption here is that commodity purchases of an industry are proportional to the industry's size and that SIC industries that share a common IOIND industry have roughly similar factor requirements.

We employ a similar procedure to impute P_{CJ} from $P_{\gamma J}$. First, we construct a correspondence between IOIND commodities, γ, and STCC commodities, C. The concordance of CFS STCC commodity codes and the BEA IOIND commodity codes is produced in a two-step process. To match STCC commodities to the corresponding SIC industries, we use a table provided to us by John Fowler at the U.S. Census Bureau. Then, to match SIC codes to IOIND commodities, we use the concordance provided by the BEA, described in the previous paragraph. The result of this two-step process is a many-to-many correspondence between IOIND commodities and STCC commodities.

Consider a single IOIND commodity, γ, which we have matched to multiple STCC commodities, C. We assign the purchases (by J) of these multiple C's in proportion to their prevalence in the Commodity Flow Survey. Since a given commodity C can correspond to multiple γ's, we need to sum over the γ's to estimate the flows of STCC-commodity C to SIC-industry J. In other words, our assumptions lead to the following expression for P_{CJ}:

$$P_{CJ} = \sum_{\gamma \in \Theta(C)} P_{\gamma J} \times \frac{V_C}{\sum_{C' \in \Psi(\gamma)} V_{C'}}$$

In this equation, V_C refers to the total amount of commodity C that we observe being shipped in the 1993 Commodity Flow Survey, $\Psi(\gamma)$ refers to the set of STCC commodities that correspond to the IOIND commodity γ, and $\Theta(C)$ refers to the set of IOIND commodities that correspond to the STCC commodity C.

---

[2] The concordance can be found at http://www.bea.gov/scb/pdf/national/inputout/meth/io1992.pdf; see pages M33-M36.

[3] The exception is in agriculture (SIC industries 0100-0299). For these industries, there are certain 4-digit SIC industries that can be matched to multiple BEA IOIND industries. For these industries, we tried several methodologies: dropping all agriculture establishments from our baseline sample, allowing for many-to-many merges, and using our best judgment over the IOIND industry which most closely matched any given 4-digit SIC. In the end we chose the latter methodology. Because establishments within agriculture are not part of the CFS sample, the choice of the methodology has essentially no effect on any of the paper's results. (The only way in which the results of the paper could at all be affected by this choice is if agriculture was an important downstream industry for many of the establishments in our sample. This turns out not to be the case.)

In the construction (SIC industries 1500-1799, 6552), there are also several 4-digit SIC industries that correspond to multiple BEA industries. This is not an issue, however, since the BEA Input-Output tables consider all of construction as a single industry (despite the fact that there are multiple IOIND industries within construction).

B.2. Measuring the Flow of Goods through the Wholesale Sector

The Input-Output Tables treat both the entire wholesale and retail sectors as single industries. Further, they do not keep track of shipments by manufacturers to or through wholesalers or retailers, instead measuring only those inputs directly used by wholesalers and retailers in the production of wholesale and retail services. To better measure the flow of goods through these sectors, we employ algorithms that rely less on the BEA's Use Table.

If industry *J* is in the wholesale sector, we impute the industry's purchases of each commodity *C* using CFS data on establishments' shipments of commodity *C* along with data from the Annual Wholesale Trade Survey (AWTS). Aggregating across establishments in the CFS gives a measure of aggregate sales, T_{CJ}, of each commodity by each wholesale industry. Second, the AWTS contains information on wholesale industries' aggregate commodity purchases and aggregate commodity sales. Using data from the AWTS, we compute the ratio R_J of commodity purchases to sales. Given these two pieces of information, we impute industry *J*'s purchases of commodity *C* as

$$P_{CJ}=T_{CJ}\times R_J.$$

To give an example, establishments in the motor-vehicle-related wholesale industries (SICs 5010-5019) had sales of $159 billion and purchases of $131 billion in 1993. We therefore set R_J = 0.82 (131/159) for all vehicle-related wholesale industries. For each commodity and industry within SICs 5010-5019, we impute aggregate purchases as 82 percent of the shipments of the respective commodity that we observe CFS establishments making.

B.3. Measuring the Flow of Goods through the Retail Sector

When *J* is a retail industry, we utilize the CFS data along with the Annual Retail Trade Survey (ARTS). The U.S. Census Bureau uses the ARTS to collect information on purchases of groups of retail industries. For example, in 1993, establishments in the household appliance industries (SICs 5720–5734) purchased $35.8 billion in intermediate materials. Unfortunately, we do not know how much total merchandise was purchased by each SIC industry within these groupings, nor do we know the amount of any specific commodity purchased by these groups. To impute these values, we rely on data from the Commodity Flow Survey and then hand match commodity-specific shipments to the most appropriate retail industry within the ARTS groupings.

To demonstrate, we continue with our household appliance retailers example. Our hand match specifies IOIND commodities 510102 (calculating and accounting machines), 510103 (electronic computers), 510104 (computer peripheral equipment), and 510400 (other office machines) as those that are sold to SIC 5734 (computer and software stores). Repeating this process for all commodities and industries yields a table of commodity-retail-industry pairs such that the 4-digit retail industry could potentially purchase the given commodity. The amount of the commodity purchased by the industry is assumed proportional to a) the amount of the good shipped, according to the Commodity Flow Survey, b) the 4-digit retail industry's share of employment among its larger grouping of industries, and c) the total amount purchased by the industry group.

For example, suppose we want to impute the purchases of computers (STCC 37531) by computer and software stores. We know that total goods purchases by retailers in SICs 5720–5734 is $35.8 billion. Since

employment in computer and software stores is 14.6 percent (30,000/205,000) of employment in this retailer group, we impute goods purchases of $5.2 billion by computer and software stores.[4] (As in Web Technical Appendix B.1, we are presuming that a) purchases of a given 4-digit SIC are proportional to employment, and b) purchase intensities are constant, across SICs, within the larger group of industries.) To impute the amount of this total that is computers specifically, we multiply the $5.2 billion figure by the value of shipments of computers as a fraction of all commodities that can be purchased by computer and software stores, where both of these commodity values are computed from the CFS (again, mirroring an assumption that we make in Web Technical Appendix B.1).

B.4. Aggregating across the Commodities Shipped by a Given Industry

In the previous three subsections, we have described how to compute the total value, P_{CJ}, of STCC-commodity *C* purchases made by each SIC industry, *J*.

In addition to this information, from the Commodity Flow Survey we are able to compute the amount of each STCC commodity that is shipped by each SIC–industry, *I*. From this, we can compute the fraction, F_{CI}, the fraction of all shipments of (STCC) commodity, *C*, that originate from (SIC) industry *I*.

Thus, to estimate the total values of shipments from *I* to industry *J*, we sum over all of the commodities that industry *I* ships:

$$S_{IJ} = \sum_{C} P_{CJ} \times F_{CI}$$

Note that, in terms of defining pairs of vertically linked industries, the defining traits of an industry are the commodities that it ships and receives. In this way, a wholesaler and manufacturer who ship the same commodity may each be defined to be upstream of a retail industry. To give an example, both Auto Assemblers (SIC 3711) and Auto Wholesalers (SIC 5010) ship large quantities of assembled automobiles (STCC 37111). Our methodology will thus define both of these industries to be upstream of the New and Used Car Dealers (SIC 5511) industry.

B.5. Some Aspects of the Resulting Definitions of Pairs of Vertically Linked Industries

We conclude this section by describing the resulting definition of vertically linked industries. The number of industries, *J*, that are classified as downstream of industry *I* depends on the cut-off that is used to determine whether industries are vertically linked to one another. For the average upstream industry, *I*, approximately 18 (out of the 900 possible downstream SIC industries) have sales for which S_{IJ} is greater than one percent. In other words, under our benchmark definition, there are 18 industries that are downstream of the average industry. Using a five percent cutoff—as we do in the robustness check given in row 8 of panel B of Table 1—the average industry has 3 industries, *J*, that are downstream of it.

Table A1 depicts these patterns for a single upstream industry, Cane Sugar, Except Refining (SIC 2061). This table presents the estimated flows from *I* to *J*. Under the benchmark definition, 15 industries are defined to be downstream of Cane Sugar. Prepared Flour Mixes and Doughs (SIC 2045) is defined to be downstream of the Cane

[4]For these employment figures, see the "Establishment and Firm Size" document at http://www.census.gov/epcd/www/92result.html.

Sugar industry, but Prepared Feeds and Feed Ingredients (SIC 2048) is not. With the five percent cutoff, Grocery Stores (SIC 5411), Farm Product Raw Materials (SIC 5159), Cane Sugar, Except Refining, (SIC 2061), and Beet Sugar (SIC 2063) are the four industries defined to be downstream of the Cane Sugar industry.

C. Data Issues

C.1. Do the Census Firm Identifiers Accurately Reflect Ownership?[5]

The Census Bureau takes particular care to ensure that the firm identifiers used in the Economic Census reflect true ownership patterns that exist across establishments. The primary source of the firm identifiers is the Establishment Identification Number (EIN), originally retrieved by the IRS. Additional surveys and audits, performed by the Census, are aimed at determining whether establishments with different EIN numbers are actually part of a single firm.

The Report of Organization Survey (also known as the Company Organization Survey), conducted by the Census, is designed specifically to correctly measure firms' ownership of different establishments. Firms with more than 500 employees receive and are required to fill out this survey annually. The survey gives firms a list of all establishments currently considered by the Census to be under ownership control by the firm and asks the firms to make any corrections. The survey also asks firms to add any establishments they own that are not currently listed. Note that, among other things, every establishment has an EIN field, so it is easy to handle cases where establishments under the firm's control happen to have different EINs. Also note that the firm itself must report whether it is under more than 50 percent ownership control by some other entity, in which case the Census would consider this other entity to be the owner of all the establishments. The survey forms, as well as additional information about the Report of Organization Survey, can be found at http://bhs.econ.census.gov/bhs/cos/form.html.

In addition, "important" companies (based on their overall prominence or salience within a sector) have a Census Bureau analyst assigned to them. This analyst knows the firm very well, is supposed to check that all company reports conform with his/her knowledge, and is tasked with resolving any discrepancies. This is part of the survey auditing process.

In summation, it seems that, certainly for firms with more than 500 employees, the Economic Census firm identification numbers should very closely reflect the true ownership patterns that exist across establishments.

C.2. The Commodity Flow Survey's Sample Design

In this subsection, we summarize the sampling methodology used to construct the Commodity Flow Survey. See U.S. Census Bureau (1996, 1999) for additional details. The design, over which shipments to sample, comprises a multi-stage process: the first stage over which establishments to survey, the second stage over which weeks of the year to request data, and the third stage over the shipments for the given respondent-week. The primary objective of the sample design is to "estimate shipping volumes (value, tons, and ton-miles) by commodity

[5] This subsection has benefited greatly from conversations with Javier Miranda, a Senior Economist at the U.S. Census Bureau. We are grateful to him for helping us understand how the Census generates its firm identifiers.

and mode of transportation at varying levels of geographic detail." (U.S. Census 1999, p. Appendix C1)

In the first stage, establishments within each industry-region are categorized as either "certainty" or "non-certainty" establishments.[6] Within each industry-region, a cut-off value is selected. Establishments that are larger[7] than the cut-off value are sampled with certainty (these are the "certainty" establishments), while other establishments—the "non-certainty" establishments—are sampled with probability less than one.
In the second stage, for each quarter, sampled establishments are assigned to different reporting periods. For the 1993 Commodity Flow Survey, the reporting period is a two-week interval. For the 1997 CFS, the length of a reporting period is one week. "To avoid potential quarterly cycles, reporting periods in subsequent quarters were assigned so that an establishment did not report at the same time each quarter." (U.S. Census 1996, p. Appendix C1)

Third, for each reporting period, each sampled establishment reports a set number of shipments. The number of shipments that an establishment reports depends on the number of shipments that the establishment actually makes during the reporting period. For respondents that make fewer than 40 shipments in the reporting period, all shipments are reported. Establishments that send between 40(n-1)+1 and 40n shipments are asked to list every n^{th} shipment that they make. For example, an establishment that sends between 41 and 80 shipments is asked to report every other shipment that it made during the reporting period.

Sample weights are constructed from the inverse of the probability that the observed shipment would—ex ante—be included in the sample.[8]

D. Additional Robustness Checks

D.1. Establishment-Level Shares of Internal Shipments

This section contains six additional robustness checks, related to those presented in Section III.B.1. First, we compute the distribution of internal shares using successively more restrictive definitions of vertically linked industries. Then, we compare our measures of establishments' internal shares to the measures constructed directly from the Census of Manufacturers. Third, we discuss how our definitions of a) vertically linked industries and b) establishments' internal shipments differ from the definitions we gave in an earlier draft. Fourth, we examine how robust the measured internal shares are to a definition of vertical links in which retail/wholesale industries are always allowed to be at the downstream end of a vertical link. Fifth, we consider how the measured internal share distribution would look if each surveyed establishments reported a larger fraction of their shipments. Finally, related to the discussion of Appendix C.1, we discuss whether there is a jump in our measured internal shares for establishments in firms with fewer than—or greater than—500 employees, as might be the case if firm identifiers

[6] Here, industries are grouped by their 3-digit SIC code. Regions are defined according to the National Transportation Analysis Region (NTAR) classification. See http://www.census.gov/geo/www/mapGallery/images/ntar000.pdf for a map of the 89 NTARs.

[7] For the 1993 Commodity Flow Survey, an establishment's payroll defines its size. For the 1997 Commodity Flow Survey, a combination of employment, payroll, and sales is used to characterize size.

[8] When computing the sample weights, the Census conducts adjustments for sampling error and survey non-response. See page C2 from U.S. Census (1996) or pages C2–C3 from U.S. Census (1999), for details.

better represented ownership patters for firms with more than 500 employees.

In the benchmark calculations, we define industry *J* to be downstream of industry *I* provided at least one percent of industry *I*'s sales were purchased by establishments in industry *J*. In the body of the paper, we also compute the internal share distribution, using a five percent cutoff rule. In rows 1-3 of Table A2, we consider the effect of changing the one percent cutoff to 10 percent, 15 percent, or 20 percent. As the cutoff increases two things occur: First, our sample of upstream establishments shrinks. Second, for any particular establishment in our sample, fewer shipments are classified as being sent along an internal, vertical link. Increasing the cutoff from 1 to 20 percent reduces the size of our sample by three-fifths. At the same time, however, the distribution of establishments' internal shares is not substantially altered using the more restrictive definition of vertical integration. Under the 20-percent cutoff, the 75th- and 90th-percentile internal shares are 4.5 percent and 28.2 percent, respectively, only somewhat smaller than the values given in Table 1.

Next, we compare our measure of internal shares to a measure derived from the Census of Manufacturers. The purpose is to show that the two internal share measures match up once we have comparable samples and comparable definitions of internal shipments. As mentioned in Section III.B of the paper, the Census records manufacturers' interplant transfers. These are shipments made to other manufacturing establishments, within the same firm, for further assembly. Since the Census of Manufacturers also contains information on each establishment's total value of shipments, it is straightforward to compute an alternate measure of internal shares by taking the ratio of interplant transfers to total value of shipments. We begin our comparison, in rows 4 and 5 of Table A2. Row 4 characterizes the distribution of interplant transfers for the 766 thousand establishments surveyed in the 1992 and 1997 version of the Census of Manufacturers. In row 5, we restrict the sample to the 37,000 establishments which are also included in our benchmark sample of establishments at the upstream end of a vertical link. Within this subsample, 76.6 percent of the establishment report no interplant transfers; the 75th-percentile internal share is 13.2 percent.

We next describe the internal share distribution, using our benchmark methodology, (i.e., using data from the Commodity Flow Survey and the algorithm specified in Section II of the paper). Row 6 restricts the benchmark sample to establishments in the manufacturing sector. For this subsample, the 75th- (90th-) percentile internal share is 6.2 percent (33.7 percent), slightly lower than the values given in Table 1 (7.0 percent and 37.6 percent, respectively).

The difference between rows 6 and 7 originates from differences in what is defined as an internal shipment. Interplant transfers, which are shipments to other establishments for further assembly, only comprise shipments sent to establishments in the manufacturing sector. Our definition, using data from the Commodity Flow Survey, includes shipments sent to same-firm establishments in any sector. In row 7, we only count a shipment as internal if there is a downstream establishment, from the same firm, that is also in the manufacturing sector.[9] The 75th- and

[9] Consider the following example of an establishment, sending a shipment of auto parts to zip code, *z*. Suppose there is a same-firm auto parts wholesaler (SIC 5013), but no manufacturing establishments in an industry that consumes auto parts, in zip code *z*. This shipment would be classified as internal according to the calculations of row 6, but not in the calculations corresponding to row 7.

90th-percentile internal shares are 0.3 percent and 11.7 percent, respectively, reasonably close to the values given in row 5.

To summarize, the interplant transfers variable yields smaller values for internal shipments, compared to the variable constructed from the Commodity Flow Survey, because it omits shipments sent to non-manufacturing establishments. If it were not for this difference, the two variables would be similar to one another.

Our definitions over which industries are vertically linked and when shipments are counted as internal were slightly different in an earlier draft of the paper (Hortaçsu and Syverson 2009). It turns out that results are qualitatively similar whether one uses the old or new definitions of vertically linked industries or internal shipments. For completeness, we review these old definitions, as well as the internal shares that resulted from these definitions.

In the previous draft, we had a less stringent definition for internal shipments. We did not require that the shipment be destined to an establishment that is in an industry directly downstream to the shipping establishment, only that the destination be an establishment that is on the downstream end of *any* vertical link in a firm.[10] In row 8 of Table A2, we recomputed internal shares for the benchmark sample, with this less strict definition of internal shipments. The median establishment has an internal share of 3.0 percent, and only 29.1 percent of establishments have no internal shipments. Compared to the benchmark calculations, the mean internal share is almost 6 percent larger (16.1 percent, compared to 10.2 percent). Thus, under our old definitions, we were being very liberal when computing internal shipments.[11]

A second difference, compared to the previous draft, originates from the way in which vertically linked industries are defined. In the previous draft, we defined two industries to be vertically linked only using information from the BEA Input-Output Tables. Specifically, a substantial link exists between Industry *A* (using the BEA definition of input-output industries) 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. As we discuss in Section II and Web Technical Appendix B, the old definition is potentially problematic if the downstream industry is retail or wholesale. For wholesalers and retailers, the BEA doesn't keep track of the gross shipments by sent to wholesalers/retailers. Instead, the BEA measures the industries which are used by wholesalers/retailers in the production of wholesale/retail services. Because of this issue, there are several pairs of industries that are, in reality, linked with one another, but are not classified as such under the old definition.

In row 9 of Table A2, we compute internal shares using the old definition of vertically linked industries (but retain the new definition of when shipments are internal to the firm). With the old definition of vertically linked industries, the sample of vertically integrated establishments is less than half as large, 29,900 compared to 67,500 establishment-years. The 75th-percentile (90th-percentile) internal share is 1.0 percent (17.4 percent). These are considerably less than corresponding values of the benchmark calculations for the 75th and 90th percentiles, 7.0

[10] For instance, suppose a firm has two upstream establishments U_1 and U_2, and two downstream establishments D_1 and D_2, and U_1-$D_1$and U_2-D_2 are separate vertical links. According to the old definitions, a shipment from U_1 would be classified as internal if it is destined to *either* D_1 or D_2's zip codes, not just D_1's.

[11] Since a main objective of the paper is to point out that internal shipments are surprisingly small, being liberal in defining internal shipments is innocuous.

percent and 37.6 percent.

In row 10 of Table A2, we compute internal shares using both the old definition of when shipments are classified as internal, and the old definition of when industries are classified as vertically linked. Under these definitions, the median establishment has an internal share of 2.5 percent, the 90th-percentile establishment has an internal share of 57.7 percent, and 2.1 percent of establishments have a 100 percent internal share.

In row 11 of Table A2, we explore the sensitivity of our internal share measures to different assumptions over the extent to which retailers and wholesalers are on the downstream end of vertical links. Given the benchmark sample of 67,500 establishment-year observations, we apply a more liberal definition of when shipments are classified to be flowing within the firm: A shipment is internal to the firm either if a) it was classified as internal, according to the baseline methodology, or b) there is a same-firm retail or wholesale establishment in the destination zip code. Under this extreme assumption, internal shares are only moderately higher. We take this finding to suggest that our original baseline methodology is not causing us to miss too many intra-firm upstream-downstream shipments.

In rows 12 through 15, we check the effect of changing the number of sampled shipments per survey participant on the estimated distribution of internal shares. As a reminder, surveyed establishments are asked to list only 20 to 40 shipments per quarter. As a result, we are almost certainly overstating the fraction of establishments that have 0 percent or 100 percent of their shipments staying within the firm. In this robustness check, we explore the magnitude of this bias induced by the limited sample size. We will try to estimate what would happen if, counterfactually, the CFS had requested many more shipments per respondent.

In this exercise, we assume that each sampled establishment has an establishment-specific probability, *p*, with which any of its shipments are sent internal to the firm. Again, because of sampling variability, *p* won't be equal to the fraction of shipments that are observed to be internal to the firm. We assume that the *p*'s are independently drawn from a Beta(α, β) distribution. (Here, we use the Beta distribution mainly because it is flexible and has the unit interval as its support.) For an establishment with $p_i = p$, the likelihood of observing m_i out of N_i shipments being internal to the firm equals $\binom{N_i}{m_i} \cdot (p_i)^{m_i} \cdot (1-p_i)^{N_i - m_i}$. Given our data on m_i and N_i we can estimate α and β via maximum likelihood. We do so: our estimates of α and β are 0.0280 and 0.955.

With the estimated distribution of *p*'s in hand, we are able to assess how the observed distributions would change with greater or fewer reported shipments per establishment. We report the percentiles of distributions, where we assume that the number sampled shipments per establishment is 1 time, 2 times, 5 times, and 25 times what is actually observed. The results are presented in rows 12-15 of Table A2. Consistent with the robustness check presented in the first row of Table 1B, assuming that the sample sizes are larger decreases the fraction of establishment-year observations with 0 percent and 100 percent internal shares. However, a larger sample size probably would not affect the measured 50th-percentile, 75th-percentile, or 90th-percentile internal shares. For example, if each establishment submitted data on 25 times as many shipments (compared to the number of shipments that they actually recorded), then the 50th- and 75th-percentile internal shares would be 0.6 percent and 8.9

percent, respectively.[12]

Finally, we examine whether the distribution of internal shares differs according to the size of the establishment's firm. As we explain in Appendix C.1, firms with over 500 employees receive the Report of Organization Survey. This survey is aimed at accurately depicting the patterns of establishment ownership, across firms. If there are organizations that own establishments with multiple firm identifiers, and if establishments within these organizations sent shipments to one another, then we would (incorrectly) classify these shipments as being inter-firm ("external") shipments. This would cause us to underreport the extent of internal shipments within vertically integrated firms. Thus, if there are problems, in our dataset, with firm identifiers, we should notice a jump—around the 500 employee cutoff—in our measured internal shares.

As we report in rows 16-19, the internal share measure is larger for establishments in firms with greater than 500 employees. However, the increase around the 500-employee cutoff is small (the value-weighted mean internal shares for the "0–500 employee" group and the "500–1000 employee" group are 6.9 percent and 7.0 percent, respectively). At the same time, internal shares are higher for the "1,000–10,000 employee" group, and even higher for the "greater than 10,000 employee" group. In combination, these results suggest that inaccurate firm identifiers are not causing us to underreport the share of internal shipments.

D.2. Is Geographic Proximity Important?

It's quite likely that some of the low internal shares we see above arise because a firm's establishments are too spatially separated to make internal shipments practical. Of course, if this is the case, this may be a *result* as much a *cause* of the lack of within-firm goods transfers along a production chain. If moving physical products down a production chain was the only reason for vertical ownership, after all, no firm would own vertically related establishments that were located too far from one another to make intra-firm shipments impractical. The fact that firms do own vertically linked producers that are far apart suggests other motives for ownership.[13]

Nevertheless, it is interesting to quantify how much distance matters. We take two approaches. The first is to compute the distribution of internal shipment shares for firms whose establishments are all located close to one another. The second is to compare establishments' shipment distances to the distances they are from other establishments in their firms.

To see shipment patterns of closely-spaced firms, we use the subset of upstream establishments from our CFS sample where *all* of the establishments that their firm owns are in the same county. (This is determined from

[12] Because of the parametric assumption that we make on the distribution of the p's, the internal share distribution resulting from our MLE estimates will not match the observed distribution, even when the number of shipments is set equal to what is actually observed. The Beta distribution has trouble fitting, for example, the small share of establishments with a 100 percent internal share.

[13] Hortaçsu and Syverson (2007) document examples of vertically integrated cement and concrete firms that own clusters of ready-mixed concrete establishments that are remotely located from their upstream cement establishments. These firms, in fact, do not internally supply these clusters with cement. The downstream concrete establishments instead report buying cement in the local market from the firm's upstream competitors. We find evidence that the firms' motives for owning these concrete establishment clusters is to harness logistical efficiencies in a business that shares a common final demand sector (construction) with cement.

the Economic Census, which includes state and county codes for virtually all establishments. This location information is not subject to the limitations of the EC zip code data, where codes for 10 percent of establishments are missing.) This subset is small—2,300 establishment-years and 200,000 shipments—and contains a large number of two-establishment firms with one upstream and downstream establishment each. Nevertheless, it offers a rough gauge the role of distance.

The results are consistent with the patterns described in Table 1. Just under half (46.7 percent) of the upstream establishments report no shipments to downstream units in their firm. The 90th-percentile establishment ships 49.0 percent of the value of its shipments internally. The fraction of establishments with all shipments staying in the firm is above that in the benchmark sample, however, at 2.4 percent. Thus it appears that vertically structured firms with closely located establishments are less likely to make internal shipments on average, but somewhat more likely to contain internally dedicated upstream establishments.

We next compare the shipment distances of our entire sample of upstream establishments in the CFS to their distances from other establishments in their firms (both measured in great circle terms). It's clear from pooling shipments across establishments that internal shipments go shorter distances. In fact, the average external shipment is sent roughly 55 percent further (349 miles versus 225 miles) than the average internal shipment. This may reflect upstream establishments "bypassing" their downstream units with some of their shipments, but it may also reflect composition effects if internally dedicated, high-volume upstream establishments are located close to downstream units in their firm.

We can decompose these contributions to the pooled numbers by looking within establishments. We find that for 8.2 percent of upstream shipment establishments, their farthest-traveling shipment does not go as far as the distance to the nearest downstream establishment in their firm. These establishments account for just over one-eighth of the one-half of our upstream establishments that report no internal shipments, showing the importance of distance. But this also means the other two-thirds of establishments reporting no internal shipments do send output at least as far as their nearest establishment. This pattern isn't unusual across the broader sample. Looking across establishments, the average of the within-establishment medians of reported shipment distances is 267 miles, while the average distance to the closest downstream establishment within the firm is 193 miles.

D.3. Is There Vertical Integration Within Establishments?

Our definition of vertically integrated ownership links requires multiple establishments by definition. A firm must own at least one establishment each in vertically related upstream and downstream industries. This definition could be problematic if firms commonly vertically integrate production within a single establishment. In such cases, the firm would be operating a vertically integrated production process and obviously supplying its own input needs. We would miss this type of integration, however, because we would not classify the establishment as integrated. There would be no shipments from the upstream to downstream parts of the production process in the CFS, since those goods transfers never leave the establishment.

To give a concrete example, consider the two following hypothetical firms. One has two establishments. The upstream establishment refines copper ore into billets which are then shipped to the downstream establishment

to be extruded into pipe. The second firm operates a similar production process in a single establishment: one side refines ore into billets, and the other side turns billets into pipe. We would define the former establishments as vertically integrated, but not the latter, even though each firm operates the same production processes.

How can we tell if this sort of misclassification is a big problem? We compare the materials purchase patterns of establishments that we classify as being in vertical structures to those in the same industry not classified as such. In the context of the above example, we compare the two copper pipe establishments. (Since establishments are classified into industries in the Economic Census based upon their outputs, both the downstream establishment in the first firm and the second firm's establishment would be classified in the same industry, SIC 3351: Rolling, Drawing, and Extruding of Copper.) The pipe establishment in the first firm—the one that we would have classified as in a vertical ownership structure—will list copper billets as an intermediate materials purchase in the Census of Manufactures materials supplement. The second establishment, where billet production is inside the establishment, will list copper ore as a materials purchase. Hence if we see substantial differences in materials use patterns across establishments (in the same industry) that we classify respectively as vertically linked or not, we should be concerned that we are missing a lot of vertical production that occurs "under one roof." On the other hand, a lack of significant differences suggests this sort of misclassification is less of a concern.

We make three such comparisons between the materials use patterns of what we classify as vertically linked establishments and others in their industry. (Again, our analysis is restricted to establishments in the manufacturing sector because of the detailed intermediate materials data requirements.) We first compute the share of each establishment's intermediate materials purchases that is for "raw materials," which we define as the products of the agricultural, fisheries, forestry, or mining sectors—i.e., SIC product codes beginning with "14" or below. We then regress this share on a set of industry-year fixed effects and an indicator equal to one if we classify the establishment as in a vertical ownership link. In essence, we test whether there are significant differences in the intensity of raw materials use across establishments that we classify as vertical and non-vertical in the same industry. We would expect that if the "under one roof" misclassification problem were substantial, we would find that establishments we designate as *non*-vertical would have a larger raw materials share, since a greater portion of the production chain would be operated within the establishment. Again, to return to our example, the pipe establishment in the second firm reports copper ore (a raw material) as a materials purchase, while the establishment in the first firm purchases copper billets.

We run this regression on a sample of over 453,000 establishment-years from the Census of Manufactures. (We don't need the CFS for this.) The coefficient on the vertical ownership link indicator is 0.47 percent, with a standard error of 0.05 percent. Thus establishments we classify as vertical use raw materials more, not less, intensively compared to other establishments in their industry. (Recall that we would expect establishments we classify as vertically linked to use raw materials less intensively). Further, the point estimate of the share difference is small, less than one-fifteenth the average raw materials share of 8.2 percent. Even if we restrict our comparisons only to those roughly 85,000 establishments that report using positive raw materials shares, the vertically linked coefficient is -1.87 percent with a 0.19 percent standard error. The estimated difference is small, relative to the 44 percent average materials share, for establishments that report positive materials purchases.

Our second check aggregates this raw materials use data to the industry level. We add up raw and total materials use of establishments classified as vertical within an industry year, and compare the ratio of the two to the same share computed for non-vertical (again, under our classification) establishments in the same industry. We then conduct a t-test for equality of means across our sample of 1867 industry-years. The mean difference is 0.08 percent, with a standard error of 0.22 percent. Here, there are no significant within-industry differences in raw materials usage intensity across the two types of establishments.

Our final check is also done at the industry-year level. We separately aggregate materials purchases of our designated vertical and non-vertical establishments for each industry year. We then order materials by decreasing intensity of use for each type of establishment (as measured by their aggregate share of purchases). This yields 86,659 industry-year-materials ranks for both vertical and non-vertical establishments. We then compare these ranks within industry-year to see if there are systematic differences. The two ranks move together; the correlation coefficient is 0.74. Table A3 shows the frequency of relative rank orderings for the five most intensively used materials by industries' non-vertical establishments. (Material 1 is the most intensively used.) Only ranks 1 through 7 of vertical establishments are shown for parsimony. If materials usage patterns were exactly the same, we would only see entries on the diagonal of the table. The most intensively used material of an industry's vertical establishments would be the most intensively used among its non-vertical establishments; the second-most used would be so for both types of establishments, and so on. Clearly, this is not the case. However, the general pattern holds. The diagonal is the largest element of a row or column, and the frequency of other pairings falls as they move further away from the diagonal. Hence, these results suggest, as do the raw materials use tests above, that there are not systematic differences in the mix of materials used by establishments we classify as in vertical ownership links and those we do not classify as such.

D.4. Cross-Industry Differences in Internal Shares

Table A4 presents, for different 2 and 3-digit industries, establishments' average internal shares. The first five columns use data from the Commodity Flow Survey, while the final three columns use data from the Census of Manufacturers. [14]

The first column gives, for all establishments surveyed in the Commodity Flow Survey, the fraction of establishments which we classify as being at the upstream end of a vertical link (i.e., these are the establishments in our main benchmark sample). The second column gives, again for all establishments, the fraction that have a positive internal share. Columns 3 through 5 give, respectively, the mean internal shares for establishments that are in our benchmark sample, the mean share for establishments that have at least some internal shipments, and the mean share for all establishments. All industry averages are establishment-sales weighted.

There is substantial variation, across different goods-producing industries. Establishments in petroleum and transportation equipment manufacturing have the largest fraction of establishments within positive internal shares; furniture manufacturers and furniture and lumber wholesalers have the smallest fraction of establishments

[14] Results for select 4-digit SIC industries can be found at the AER webpage corresponding to this article.

with positive internal shares. There is also substantial variation, among industries, in average internal shares, with the highest average internal shares being the fabric and petroleum manufacturing industries, and the lowest average internal shares being the furniture manufacturing industry and the lumber, metals, drugs, chemicals, and beer and wine wholesaling industries.

Columns 6 through 8 display the corresponding calculations, using the interplant transfers variable from the Census of Manufacturers. While both the sample and the definition of internal shares differ between columns 1–5 and columns 6–8, the cross-industry patterns of internal shares are similar across the two sets of calculations. Paper, transportation equipment, and primary metals manufacturing are some of the more vertically integrated industries; furniture manufacturing and printing are two of the least vertically integrated.

The petroleum industry is a bit of an outlier, and deserves extra attention. Petroleum is an industry that has one of the highest internal shares in columns 1–5, but one of the lowest internal shares in columns 6–8. This difference results from the different definitions of internal shipments across the two datasets. The interplant transfers variable, collected in the Census of Manufacturers, asks establishments to give the value of shipments sent to other manufacturing establishments for further assembly. Since shipments by petroleum manufacturers are mainly sent to wholesalers, and not to other manufacturers, the internal shares computed from the Census of Manufacturers tend to be significantly smaller than the internal shares computed using data from the Commodity Flow Survey.

D.5. Firm Size Differences by Firm Structure

Figure A1 plots the densities of firm size (log total employment, since revenue is unavailable outside of manufacturing) for three mutually exclusive and exhaustive sets of multi-establishment firms. One set includes firms with vertical ownership structures.[15] The other two multi-unit organizational structures are single-industry and multi-industry-unintegrated firms.[16]

The figure reveals that each of the (log) employment size distributions is unimodal, though they clearly have different central tendencies.[17] Single-industry multi-unit firms are the smallest and have the most symmetric size distribution. Vertically integrated firms are clearly the largest on average, and their distribution is more skewed than the other firm types. (While not plotted, single-establishment firms are smaller than the multi-unit single-

[15] Recall that we define vertical ownership at the establishment, not firm, level. For our purposes here, however, we define a firm as vertically structured if it owns any vertically linked establishments. As a practical matter, most establishments in what we call vertically structured firms here are also in vertical chains according to our establishment-specific definition.

[16] The distribution of establishments across these firm sets is as follows. Over the entire manufacturing sample, multi-unit establishments of all types accounted for 19.7 percent of establishments, 71.8 percent of employment, and 86.5 percent of the capital stock. Vertically integrated establishments' shares were, respectively, 14.5, 60.4, and 79.2 percent. Multi-unit single-industry establishments accounted for 2.8 percent of establishments and 5.2 and 3.2 percent of employment and capital, while multi-industry unintegrated establishments comprised 3.7, 8.0, and 5.3 percent of establishments, employment, and capital, respectively.

[17] We only plot the 1997 distributions rather than those pooled across years in order to remove any secular shifts in firm sizes. Checks of other years show similarly shaped distributions.

industry firms, as one might expect.) Thus, not only are vertically integrated establishments larger, their firms are as well.

D.6. Establishment Attributes by Vertical Ownership Structure

When using a one percent cutoff rule, 74 percent (=14.5/19.7) of establishments in multi-unit firms are classified, by our methodology, to be part of a vertical production chain. Thus, the comparisons described in Tables 3-5 of the paper are, to a large extent, between establishments in multi-unit firms and establishments in single-unit firms.

For this reason, we re-examine Tables 3-4 using a five percent cutoff rule. With the more stringent definition, a smaller fraction of manufacturing establishments are classified to be part of a production chain. However, as Tables A5 and A6 demonstrate, the differences in establishment attributes by vertical structure are robust to the cutoff rule that we have chosen.

D.7. Flows of Intangible Inputs

In our product mix and shipment destination tests, we use the following algorithm to identify establishments that experienced ownership changes. From the Longitudinal Business Database, we begin with all establishments for which the firm identifier changes between *t* and *t*+1. Since firm identifiers may change across years for a number of reasons, we need to discard the observations which are unrelated to mergers or acquisitions.[18] For the establishments that change firm identifiers, we say that a change of ownership has occurred if they share the same firm identifier with some other set of establishments in period *t*+1, but not in period *t*. We define the other set of establishments to be an acquiring firm, if their firm identifier is the same in both years *t* and *t*+1.[19]

Here, we complement our analysis in the main text of summary data with more formal analyses. In Table A7, we estimate the probability that establishment *i* will produce a given 7-digit product in year *t* as a function of the year *t*-5 production patterns of the acquiring and acquired firms. We find that an establishment is more likely to produce a product in year *t* if either the acquiring or the acquired firm was producing the product. The probability that an acquired establishment produces a given 7-digit product in year *t* is 6 percent higher for products that were produced by the acquiring firm in year *t*-5. Compared to other products in their 4-digit industry, acquired establishments are also more likely to produce the products that its original firm was producing: The probability that

[18] For example, legal reorganizations may cause a change in firm identifiers without an actual change in ownership. For an additional example, multi-unit and single-unit firms are coded differently by the Census: A single-unit establishment that opens a new establishment elsewhere will have its firm identifier change, again without any change in ownership.

[19] An example will help explain how the algorithm works. Consider a two-establishment firm with establishment identifiers I_1 and I_2, firm identifier *F* in year *t*, and firm identifier *G* in year *t*+1. If there are no other establishments in year *t*+1 that have firm identifier *G*, then the algorithm would not identify a change of ownership. On the other hand, if there exists some establishment, I_3 that had firm identifier *G* in year *t*, our algorithm would identify I_3 as the sole establishment in the acquiring firm; I_1 and I_2 would then be classified as members of the acquired firm. Using a different method, Nguyen (1998) constructs a sample of acquired establishments, called the Ownership Change Database. As a robustness check, we re-create Tables A7 and A8 using the Ownership Change Database. The results are presented in the final columns of Tables A9 and A10.

establishment *i* produces a given 7-digit product is 7 percent higher for products that were produced in year *t*-5 by some other establishments of the acquired firm.

To further explore the evolution of acquired establishments' shipping patterns, we run a series of logit regressions to estimate the probability that an acquired establishment *i* will ship to any particular zip code *z* in 1997. In these regressions, the variables of interest measure the shipping patterns of the acquiring and acquired firms in 1993. In addition, we include the following control variables: establishment-by-destination-county fixed effects, control variables for total sales to zip code *z*, the great-circle distance between *i* and *z*, an indicator variable equal to one if there exists an establishment from the same firm in 1997, and an indicator variable equal to one if establishment *i* shipped to *z* in 1993.

Table A8 contains the results from these regressions. An establishment is significantly more likely to ship to a zip code if either the acquiring or acquired firm sold to that zip code in previous years. The probability that establishment *i* sells to zip code *z* in 1997 is 1.2 percent higher when an establishment from the acquiring firm sold to that zip code in 1993. The estimated marginal effect is significantly larger, 4.6 percent, if the establishment from the acquiring firm shares the same 4-digit industry as the acquired establishment. Finally, these marginal effects are economically meaningful. The average probability that an acquired establishment in our sample sells to a particular zip code is 4.0 percent. Furthermore, the acquired establishment *i* is more likely to ship to the zip codes that it used to sell to, and to the zip codes that other establishments in the acquired firm were selling to.

We also estimate these logit regressions with different subsets of the sample of acquired establishments. In Table A9, we re-estimate the probability than an establishment manufacturers a given 7-digit product. Again, we cut the data according to the year of the acquisition (column 1 versus column 2). We also run the logit regression separately for establishments that were in multi-unit firms and single-unit firms (column 3 versus column 4). Finally, we use a dataset—the Ownership Change Database constructed by Sang V. Nguyen of the Census Bureau—as an alternate source of acquired establishments. Coefficient estimates are similar across the different subsamples. In particular, in each specification, the probability that establishment i manufactures a particular 7-digit product is at least 5.5 percent larger when the acquiring firm had an establishment that, in year *t*-5, produced that same product.

Table A10 presents robustness checks related to the estimation of the probability that an acquired establishment ships to a particular zip code. In the first two columns, we show that the estimated effects are similar for establishments that merge earlier or later on. In the third and fourth columns, we show that the estimated effects are similar for establishments that were, in 1992, part of a multi-unit or a single-unit firm.[20] In the fifth column, we estimate the probability of shipping to a particular zip code for establishments in the wholesale, instead of the manufacturing, sector. Finally, in the sixth column, there is no substantial difference in the estimated effects from using the Ownership Change Database to define the set of acquired establishments.

E. Industries Mentioned in Lafontaine and Slade (2007)

[20] Since, in the fourth column, the sample includes only establishments that are in single-unit firms in 1992, the I(in 1993, another establishment, from the acquired firm, shipped to *z*) indicator is 0 for all establishments. Thus, this variable is excluded from the list of independent variables.

In this section, we describe the twelve 4-digit industries used in the robustness check of Table 1B, row 6. These industries are analyzed in the studies reviewed in Lafontaine and Slade (2007). The industries listed below are at the upstream end of their vertical links. (For example, *Auto Parts Manufacturers* refers to the link from automotive parts markers to automotive assemblers.)

Aerospace Parts Manufacturers. Masten (1984) studies the make-or-buy decision for airplane assemblers. We include the parts suppliers (SIC 3724, Aircraft Engines and Engine Parts).

Auto Parts Manufacturers. Several articles, including Masten, Meehan, and Snyder (1989), discuss the relationships between auto parts manufacturers (SIC 3714, Motor Vehicle Parts and Accessories) and auto assemblers (SIC 3711).

Cement. Vertical relationships between cement and ready-mix concrete manufacturers are the focus of Hortaçsu and Syverson (2007). Cement is produced primarily by establishments in 4-digit industry number 3241.

Coal. Establishments that engage coal mining reside in two SIC industries:1221: Bituminous Coal and Lignite, Surface Mining; and 1222: Bituminous Coal, Underground Mining. Joskow (1985) studies integration and contractual relationships between these coal-mining establishments and electricity-generating establishments. He notes that only a small fraction of coal shipments—approximately 15 percent, by value—are within-firm shipments. Indeed, among the Lafontaine and Slade (2007) subsample, the two coal-producing industries have two of the lowest fraction of within-firm shipments: 9 percent for Surface Mining and 12 percent for Underground Mining.

Industrial Gases. Both Lieberman (1991) and Mullainathan and Scharfstein (2001) study vertical relationships between organic chemical manufacturers and their customers. Mullanathian and Scharfstein (2001) focus on producers of vinyl chloride monomer (part of SIC 2869, Industrial Chemicals NEC), while the sample in Lieberman (1991) contains numerous products. The five largest products in their sample are propylene (part of SIC 2869), benzene (part of SIC 2865, Cyclic Organic Crudes and Intermediates), chlorine (part of SIC 2812, Alkalies and Chlorine), toluene (part of SIC 2865), and ethylene (part of SIC 2869). Due to data confidentiality regulations, we cannot include SIC 2812 in our calculations. However, we can include both SIC 2865 and SIC 2869.

Petroleum Refiners and Petroleum Wholesalers. Like the auto industry, petroleum refining and distribution has received substantial interest from industrial organization economists (e.g., Gilbert and Hastings (2005)). The three industries that we include are SIC 2911 (Petroleum Refining), 5171 (Petroleum Bulk stations and Terminals), and 5172 (Petroleum and Petroleum Products Wholesalers, Except Bulk Stations and Terminals).

Shoe Manufacturing. Woodruff (2002) studies the integration decisions of Mexican footwear manufacturers and retailers. We include men's footwear (SIC 3143) in our subsample of industries with a prior of high internal

shipments. We cannot separately report women's footwear (SIC 3144) without violating Census data confidentiality regulations.

Soft Drink Bottlers. Muris, Scheffman, and Spiller (1992) chronicle the evolution of the soft drink industry, in particular the transition towards integration between soft drink bottlers and the two concentrate manufacturers (Coca-Cola and Pepsi). The soft drink bottling industry is SIC 2086.

Other Industries. The Lafontaine and Slade (2007) article reviews several additional industries, which we could not include in our calculations. The majority of these industries are those that are not included in the Commodity Flow Survey's sample frame. These include retail and service industries, but also some of the mining industries. For example, we could not include the iron ore mining industry, which is analyzed in Mullin and Mullin (1997), for this reason. There are also industries included in the CFS sample frame that we could not include. In these industries, there are too few establishments to pass the Census confidentiality requirement. Pulp mills (SIC 2611), analyzed in Ohanian (1994), is an example of one such industry. The other industries that we could not include, for this reason, are women's footwear (SIC 3144) and ship building (SIC 3731).

Appendix References

Harper, Michael, Ernst R. Berndt, and David O. Wood. 1989. "Rates of Return and Capital Aggregation Using Alternative Rental Prices." In *Technology and Capital Formation*, edited by Dale W. Jorgenson and Ralph Landau. Cambridge, MA: MIT Press.

Hastings, Justine S. and Richard J. Gilbert. 2005. "Market Power, Vertical Integration, and the Wholesale Price of Gasoline." *Journal of Industrial Economics*, 53(4): 469-92.

Hortaçsu, Ali and Chad Syverson. 2007. "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices." *Journal of Political Economy*, 115(2): 250-301.

Hortaçsu, Ali and Chad Syverson. 2009. "Why Do Firms Own Production Chains?" U.S. Census Bureau Center for Economic Studies Paper No. CES 09-31.

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

Lafontaine, Francine and Margaret Slade. 2007. "Vertical Integration and Firm Boundaries: The Evidence." *Journal of Economic Literature*, 45(3): 629-85.

Lieberman, Marvin B. 1991. "Determinants of Vertical Integration: An Empirical Test." *Journal of Industrial Economics*, 39(5): 451-66.

Masten, Scott E. 1984. "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics*, 27(2): 403-17.

Masten, Scott E., James W. Meehan Jr., and Edward A. Snyder. 1989. "Vertical Integration in the U.S. Auto Industry: A Note on the Influence of Transaction Specific Assets." *Journal of Economic Behavior and Organization*, 12(2): 265-73.

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

Mullainathan, Sendhil and David Scharfstein. 2001. "Do Firm Boundaries Matter?" *American Economic Review*, 91(2): 195-99.

Mullin, Joseph C., and Wallace P. Mullin. 1997. "United States Steel's Acquisition of the Great Northern Ore Properties: Vertical Foreclosure or Efficient Contractual Governance?" *Journal of Law, Economics, and Organization*, 13(1): 74-100.

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

Nguyen, Sang V. 1998. "The Manufacturing Plant Ownership Change Database: Its Construction and Usefulness." *Journal of Economic and Social Measurement*, 24(3-4): 209-32.

Ohanian, Nancy K. 1994. "Vertical Integration in the U.S. Pulp and Paper Industry, 1900–1940." *Review of Economics and Statistics*, 76(1): 202-07.

U.S. Bureau of Labor Statistics. 1983. "Trends in Multifactor Productivity: 1948-81." Bulletin 2178. Washington D.C.: Government Printing Office.

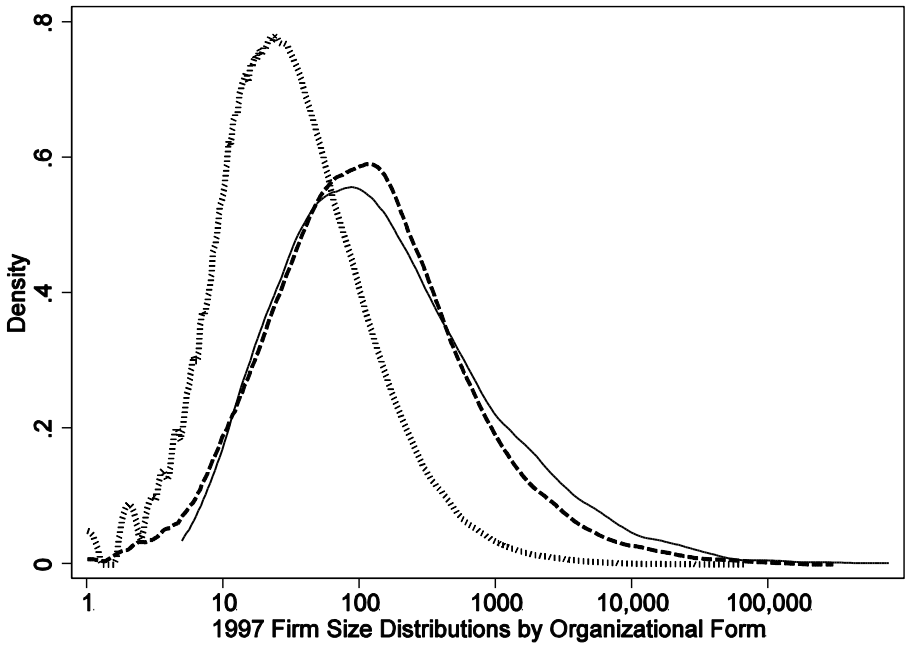
U.S. Census Bureau. 1996. "1992 Census of Transportation, Communications, and Utilities: 1993 Commodity Flow Survey, United States." Washington D.C.: Department of Commerce, Economic and Statistics

Administration.

U.S. Census Bureau. 1999. "1997 Census of Transportation: 1997 Commodity Flow Survey." Washington D.C.: Department of Commerce, Economic and Statistics Administration.

Woodruff, Christopher. 2002. "Non-Contractible Investments and Vertical Integration in the Mexican Footwear Industry." *International Journal of Industrial Organization*, 20(8): 1197–1224.

Figure A1. Firm Size Distributions by Organizational Structure



Notes: This figure shows density plots of the firm size distributions (measured by log total employees) for the three types of multi-establishment firms: single-industry, multi-establishment firms (thick, dashed line); multi-industry, non-VI firms (thin, dashed line); and VI firms (thin, solid line). See text for details.

Table A1: The Flows of Goods from the Cane Sugar, Except Refining, Industry to Other Industries

Downstream Industry SIC	Downstream Industry Name	S_{IJ}
5411	Grocery Stores	28.1%
5159	Farm Product Raw Materials	27.0%
2061	Cane Sugar, Except Refining	6.0%
2063	Beet Sugar	5.0%
5812	Eating Places	4.4%
2062	Cane Sugar Refining	4.1%
2051	Bread, Cake, and Related Products	3.5%
5410	Grocery and Convenience Stores	1.7%
2043	Cereal Breakfast Foods	1.7%
2052	Cookies and Crackers	1.6%
5149	Groceries, Not Elsewhere Classified	1.6%
2099	Food Prep., Not Elsewhere Classified	1.6%
2066	Chocolate and Cocoa Products	1.5%
5194	Tobacco and Tobacco Products	1.4%
2045	Prepared Flour Mixes and Doughs	1.0%
2048	Prepared Feeds and Feed Ingredients	0.9%
2033	Canned Fruits and Vegetables	0.7%
2087	Flavoring Extracts and Syrups	0.6%
2024	Ice Cream and Frozen Desserts	0.6%
5191	Farm Supplies	0.5%
2086	Bottled and Canned Soft Drinks	0.5%

Notes: This table shows the flows, according to the algorithm described in Web Technical Appendices B.1-B.4, originating from industry SIC 2061 (Cane Sugar, Except Refining). Under the benchmark definition, the first 15 industries are classified to be downstream of the Cane Sugar industry.

Table A2. Establishment-Level Shares of Internal Shipments: Additional Robustness Checks

	Percentile							
Specification/Sample	50th	75th	90th	95th	Frac. = 0	Frac.= 1	Weighted Mean	Approx. Establishment.-years
1. 10 percent cutoff definition for VI	0.0%	4.4%	28.7%	58.8%	55.1%	0.8%	12.1%	42,800
2. 15 percent cutoff definition for VI	0.0%	4.3%	27.3%	55.3%	55.0%	0.7%	9.8%	34,300
3. 20 percent cutoff definition for VI	0.0%	4.5%	28.2%	55.1%	55.1%	0.7%	9.3%	27,400
4. Interplant transfers	0.0%	0.0%	0.0%	0.0%	97.8%	0.1%	6.1%	766,000
5. Interplant transfers, establishments surveyed in benchmark sample	0.0%	0.0%	13.2%	51.1%	76.6%	1.1%	9.1%	37,000
6. Establishments that are in the CMF	0.0%	6.2%	33.7%	64.5%	50.9%	1.1%	12.2%	37,000
7. Establishments that are in the CMF, shipments to manufacturers	0.0%	0.3%	11.7%	33.5%	71.1%	1.1%	5.6%	37,000
8. Don't require the sending and receiving establishments to be part of a vertical link	3.0%	19.4%	59.2%	84.8%	29.1%	2.1%	20.8%	67,500
9. Original method for defining vertical links	0.0%	1.0%	17.4%	44.5%	67.5%	0.8%	6.4%	29,900
10. Original method for defining vertical links & don't require the sending and receiving establishments to be part of a vertical link.	2.5%	18.9%	57.7%	84.0%	33.7%	2.1%	19.3%	29,900
11. Retail and wholesale are always classified as downstream of other industries.	1.4%	14.6%	52.0%	80.2%	36.8%	1.7%	19.0%	67,500
12. MLE Estimate	0.0%	8.9%	40.4%	65.0%	51.4%	0.2%		67,500
13. MLE Estimate & the number of sampled shipments per surveyed estab. was 2 times as large.	0.5%	8.9%	40.2%	64.7%	45.3%	0.1%		67,500
14. MLE Estimate & the number of sampled shipments per surveyed estab. was 5 times as large.	0.5%	8.9%	40.1%	64.6%	37.3%	<0.1%		67,500
15. MLE Estimate & the number of sampled shipments per surveyed estab.	0.6%	8.9%	40.1%	64.5%	34.5%	<0.1%		67,500

was 25 times as large.

16. Establishment in firm with 10,000+ employees	1.1%	15.2%	57.9%	82.7%	38.3%	1.4%	21.4%	19,000
17. Establishment in firm with 1000-10,000 employees	0.0%	5.3%	30.1%	61.8%	52.3%	1.0%	10.9%	21,600
18. Establishment in firm with 500-1000 employees.	0.0%	3.4%	25.0%	55.4%	57.6%	1.0%	7.0%	5,600
19. Establishment in firm with fewer than 500 employees.	0.0%	4.4%	27.6%	58.8%	55.0%	1.1%	6.9%	21,300

Notes: Each row shows, for a different subsample, the distributions of the shares (by dollar value) of upstream integrated establishments' shipments that are internal to the firm. The criteria for inclusion in and size of each subsample is discussed in the text. For data confidentiality reasons, the reported percentiles are averages of immediately surrounding percentiles, e.g., the median = 0.5*(49^{th} percentile + 51^{st} percentile).

Table A3. Relative Material Use Intensity Ranks between Establishments in Vertical Ownership Structures and Other Establishments

		Material's intensity rank in non-vertically linked establishments				
		1	2	3	4	5
Material's intensity rank in vertically linked establishments	1	50.5%	13.7%	8.0%	4.5%	3.3%
	2	14.7%	26.1%	15.4%	10.1%	6.0%
	3	8.1%	14.7%	19.1%	13.2%	9.8%
	4	5.6%	10.6%	12.1%	14.8%	11.8%
	5	3.1%	6.5%	9.6%	11.2%	11.3%
	6	3.2%	5.1%	6.5%	7.9%	10.4%
	7	2.2%	4.3%	5.9%	6.4%	7.3%

Notes: This table shows, for a sample of 9,545 industry-material-year cells, the ranks of materials intensity use (by share of materials purchases) for the five most intensively used materials in establishments we define as not in vertical ownership structures. The entries in the table correspond to the fraction of cells where vertical and non-vertical establishments in the same industry share a particular pair of materials intensity rankings. For example, across all industry-years in the sample, the most intensively used (rank 1) material by non-vertical establishments in an industry-year is the most intensively used material by the industry-year's vertical establishments 50.5 percent of the time. Non-vertical establishments' rank 1 material is vertical establishments' second most intensively used material 14.7 percent of the time, and so on. Industries are defined according to the BEA's IOIND classification.

Table A4. Establishment-Level Shares of Internal Shipments, by Industry

	Using Commodity Flow Survey Data					Using Census of Manufacturers Data		
Industry	Fraction of ests. that are in the VI sample	Fraction of ests. w/ share > 0	Mean share for ests. in the VI sample	Mean share, cond. on share > 0	Mean share	Fraction of ests. w/ share > 0	Mean share, cond. on share > 0	Mean share
12, Coal Mining	78.4%	26.0%	13.5%	40.8%	10.6%			
14, Stone	65.9%	27.7%	10.0%	23.7%	6.6%			
20, Food	82.0%	52.4%	13.4%	20.9%	11.0%	3.7%	13.9%	3.5%
22, Fabric	78.1%	50.5%	22.3%	34.4%	17.4%	8.3%	53.1%	20.5%
23, Apparel	55.6%	32.9%	11.9%	19.9%	6.6%	0.8%	46.4%	4.3%
24, Wood	53.7%	31.1%	11.4%	19.5%	6.1%	1.4%	27.3%	4.7%
25, Furniture	39.5%	16.1%	4.1%	10.2%	1.6%	0.9%	10.1%	0.9%
26, Paper	73.0%	40.4%	7.7%	13.9%	5.6%	11.1%	25.1%	10.0%
27, Printing	55.4%	21.6%	4.3%	11.2%	2.4%	0.3%	16.7%	0.6%
28, Chemicals	86.7%	49.2%	9.6%	16.9%	8.3%	6.4%	19.4%	7.5%
29, Petroleum	94.0%	76.8%	30.6%	37.5%	28.8%	8.8%	6.8%	3.0%
30, Plastics	58.7%	28.0%	7.5%	15.8%	4.4%	4.0%	18.0%	3.4%
31, Leather	64.0%	38.0%	17.5%	29.6%	11.2%	1.3%	31.3%	3.7%
32, Glass, Stone	69.1%	38.5%	8.8%	15.9%	6.1%	1.5%	25.0%	4.0%
33, Primary Metals	77.6%	48.9%	10.8%	17.2%	8.4%	7.6%	26.1%	11.0%
34, Fabr. Metals	50.6%	26.7%	10.7%	20.3%	5.4%	2.1%	34.2%	6.1%
35, Ind. Machinery	67.5%	40.8%	7.1%	11.9%	4.8%	1.3%	18.8%	4.4%
36, Elc. Equipment	73.9%	46.4%	9.6%	15.3%	7.1%	3.5%	26.3%	6.5%
37, Trans. Equip.	86.2%	65.6%	13.0%	17.2%	11.2%	4.5%	28.6%	9.4%
38, Instruments	74.8%	43.8%	9.0%	15.4%	6.7%	2.2%	11.4%	3.3%
39, Miscellaneous	35.4%	11.9%	6.2%	18.4%	2.2%	0.5%	13.3%	1.1%
501, Vehicles	75.9%	52.7%	8.8%	12.7%	6.7%			
502, Furniture	39.0%	17.5%	5.6%	12.5%	2.2%			
503, Lumber	53.7%	17.9%	2.8%	8.4%	1.5%			
504, Prof. Equip.	49.5%	31.4%	10.3%	16.1%	5.1%			
505, Metals	59.5%	24.3%	3.4%	8.4%	2.0%			

506, Electrical	57.4%	34.4%	6.1%	10.0%	3.5%
507, Hardware	56.5%	25.3%	5.0%	10.9%	2.8%
508, Machinery	48.2%	29.1%	6.0%	10.0%	2.9%
509, Miscellaneous	28.6%	8.7%	3.8%	13.1%	1.1%
511, Paper	61.8%	34.8%	4.0%	7.2%	2.5%
512, Drugs	77.1%	26.5%	1.7%	5.0%	1.3%
513, Apparel	43.5%	27.6%	8.0%	12.8%	3.5%
514, Groceries	62.8%	32.3%	10.8%	21.0%	6.8%
515, Farm Products	63.3%	33.1%	19.9%	38.0%	12.6%
516, Chemicals	44.8%	20.4%	4.0%	8.8%	1.8%
517, Petroleum	73.1%	52.9%	23.5%	32.5%	17.2%
518, Beer & Wine	47.2%	11.1%	2.5%	11.0%	1.2%
519, Miscellaneous	49.0%	22.0%	9.0%	20.1%	4.4%

Note: Each row shows, for a different SIC industry, the fraction of establishments that have positive internal shipments, as well as the average share of internal shipments. Industries in the mining and manufacturing sectors are averaged over 2-digit industries. Industries in the wholesale sector are averaged over 3-digit industries. Tobacco (SIC 21) is combined with Food (SIC 20). All calculations are sales-weighted (i.e, 10.6 percent of the shipment value in the coal-mining industry is within firm).

Table A5. Establishment Attributes by Vertical Ownership Structure: Five Percent Cutoff Rule

	Output per hour	TFP	Output	Capital-labor ratio
A. Within-industry differences				
Indicator for vertical estabs.	0.377*	0.009*	1.515*	0.460*
	(0.002)	(0.001)	(0.005)	(0.003)
Approximate N	970,000	879,000	991,000	937,000
Approximate N[vertical estabs.]	144,000	137,000	147,000	142,000
B. Differences among new establishments				
Indicator for vertical estabs.	0.320*	0.024*	1.288*	0.363*
	(0.006)	(0.004)	(0.012)	(0.008)
Approximate N	240,000	213,000	248,000	233,000
Approximate N[vertical estabs.]	23,000	21,000	24,000	23,000
C. Comparing unintegrated establishments: to-be-vertical vs. remaining non-vertical				
Indicator for to-be-vertical estabs.	0.243*	0.007*	1.265*	0.295*
	(0.006)	(0.004)	(0.012)	(0.008)
Approximate N	453,000	415,000	462,000	439,000
Approximate N[to be vertical]	13,000	12,000	13,000	13,000
D. Changes upon entering vertical ownership				
Newly vertical indicator	0.043*	-0.010*	0.005	0.033*
	(0.006)	(0.004)	(0.008)	(0.009)
Approximate N	397,000	345,000	407,000	375,000
Approximate N[newly vertical]	13,000	12,000	13,000	12,000

Notes: This table shows establishment "type" comparisons between establishments in (or to-be-in) vertical ownership structures and their non-vertical counterparts. Unlike Table 3, industry *I* is defined to be upstream of industry *J* if greater than five percent—not one percent—of industry *I*'s output is sent to industry *J*. Panel A compares across all establishments for which type measures are available. Panel B compares new establishments. Panel C compares *prior period* types among non-vertical establishments that will become part of vertical ownership structures by next period to those remaining non-vertical. Panel D compares changes in type for establishments that become part of vertical ownership structures to changes for unintegrated establishments that remain so. All regressions include industry-year fixed effects; industries are defined according to the BEA's IOIND classification. Samples are comprised of non-administrative-record manufacturing establishments. See text and Web Technical Appendix A on construction of type measures and additional details. An asterisk denotes significance at the five percent level.

Table A6. Establishment Type Differences Controlling for Firm Size: Five Percent Cutoff Rule

	Output per hour	TFP	Output	Capital-labor ratio
VI indicator	0.070*	-0.002	0.300*	0.069*
	(0.004)	(0.002)	(0.007)	(0.005)
Approximate N	231,000	220,000	235,000	227,000
Approximate N[VI Indicator]	131,000	125,000	134,000	129,000

Notes: This table shows the results from regressing establishment-level type measures on an indicator for vertically integrated establishments, a set of industry-year fixed effects, and control variables for firm size; industries are defined according to the BEA's IOIND classification. The sample consists of establishments in multi-industry firms. Unlike Table 4, industry *I* is defined to be upstream of industry *J* if greater than five percent—not one percent—of industry *I*'s output is sent to industry *J*. The firm size control variables include quintics of several measures of the establishment's owning-firm size: (log) employment, the (logarithm of the) number of establishments, and the (logarithm of the) number of industries. These firm size measures are computed by summing over the *other* plants in the firm of the establishment in question. An asterisk denotes significance at the five percent level.

Table A7. Logit Regression: Probability that Establishment *i* Produces a Given 7-digit Product in Year *t*

I(estab. produced 6-digit product in *t*-5)	1.215* (0.037) 0.086	0.921* (0.038) 0.052	1.058* (0.039) 0.063
I(estab. produced 7-digit product in *t*-5)	2.313* (0.036) 0.469	2.366* (0.036) 0.422	2.189* (0.037) 0.399
I(in t-5, another estab. from the acquired firm produced the 6-digit product)		0.774* (0.041) 0.055	0.321* (0.059) 0.018
I(in t-5, an estab. from the acquiring firm produced the 6-digit product)		0.619* (0.038) 0.041	0.113* (0.054) 0.006
I(in t-5, another estab. from the acquired firm produced the 7-digit product)			0.608* (0.054) 0.052
I(in t-5, an estab. from the acquiring firm produced the 7-digit product)			0.702* (0.051) 0.053
Approx. N	140,000	140,000	140,000
Approx. number of establishment-by-4-digit industry groups	7,600	7,600	7,600
Pseudo R^2	0.353	0.363	0.368
Average probability that *i* produces the 7-digit good in year *t*	13.1%	13.1%	13.1%

Notes: Each column gives the results from a separate logit regression. For each variable, coefficient estimates, standard errors, and marginal effects are reported. The dependent variable equals 1 provided establishment *i* produces 7-digit product, *p*, in year *t*. The sample includes all *i*-*p* pairs for which a) *i* was purchased between *t*-5 and *t*-1, and b) product *p* was produced at least such acquired establishment in year *t*. Control variables for total sales in year *t* of the 7-digit product (minus sales of the product by establishment *i*) are included, but not reported. $t \in \{1992, 1997\}$. All regressions include establishment-by-4-digit-industry fixed effects. An asterisk denotes significance at the five percent level.

Table A8. Logit Regression: Probability that Establishment *i* Ships to Zip Code *z* in 1997

I(estab. *i* sold to zip code *z* in 1993)	2.357* (0.017) 0.178	2.226* (0.018) 0.156	2.215* (0.018) 0.154	2.212* (0.018) 0.153	2.176* (0.039) 0.155	2.223* (0.020) 0.153
I(in 1997, an estab. from the merged firm has a physical location in *z*)	1.141* (0.030) 0.047	0.988* (0.031) 0.0377	0.986* (0.031) 0.037	0.982* (0.031) 0.037	1.292* (0.050) 0.059	0.794* (0.039) 0.027
ln(distance)	-0.127* (0.017) -0.003	-0.114* (0.017) -0.003	-0.112* (0.017) -0.003	-0.112* (0.017) -0.003	0.007 (0.037) 0.000	-0.152* (0.019) -0.003
I(in 1993, another estab. from the acquired firm shipped to *z*)		1.1299* (0.024) 0.046	0.802* (0.044) 0.027	0.801* (0.044) 0.027	0.587* (0.090) 0.019	0.872* (0.051) 0.03
I(in 1993, an estab. from the acquiring firm shipped to *z*)		0.638* (0.017) 0.02	0.435* (0.022) 0.012	0.432* (0.022) 0.012	0.480* (0.045) 0.014	0.417* (0.025) 0.011
I(in 1993, another estab. in the same 2-digit SIC, from the acquired firm shipped to *z*)			0.454* (0.051) 0.027	0.155* (0.068) 0.008	0.298* (0.126) 0.014	0.114 (0.082) 0.006
I(in 1993, an estab. in the same 2-digit SIC, from the acquiring firm shipped to *z*)			0.420* (0.029) 0.017	0.187* (0.034) 0.007	0.186* (0.064) 0.007	0.182* (0.041) 0.007
I(in 1993, another estab. in the same 4-digit SIC, from the acquired firm shipped to *z*)				0.406* (0.061) 0.027	0.422* (0.109 0.028	0.401* (0.074) 0.027
I(in 1993, an estab. in the same 4-digit SIC, from the acquiring firm shipped to *z*)				0.526* (0.040) 0.027	0.659* (0.071) 0.039	0.454* (0.049) 0.022
Include establishments with (or without) internal shipments?	Both	Both	Both	Both	Internal Share>0	Internal Share=0
Approx. N	1.45 million	1.45 million	1.45 million	1.45 million	0.31 million	1.14 million
Number of establishment-by-destination counties	46,500	46,500	46,500	46,500	10,500	36,000
Pseudo R^2	0.178	0.189	0.190	0.191	0.203	0.188
Average probability that *i* ships to *z* in 1997	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%

Each column gives the results from a separate logit regression. For each variable, coefficient estimates, standard errors, and marginal effects are reported. The dependent variable equals 1 provided establishment *i* ships to zip code *z* in 1997. The sample includes all *i*-*z* pairs for which *i* was purchased between 1992 and 1996, and *z* was a destination zip code for at least one such acquired establishment in 1997. Control variables for total sales in zip code *z* (minus sales from establishment *i*) are included, but not reported. All regressions include establishment-destination county fixed effects. An asterisk denotes significance at the five percent level.

Table A9. Logit Regression: Probability that Establishment *i* Produces a Given 7-digit Product in Year *t*: Robustness Checks

I(estab. produced 6-digit product in *t*-5)	1.086* (0.051) 0.070	1.022* (0.061) 0.060	1.066* (0.042) 0.064	1.009* (0.103) 0.064	1.046* (0.034) 0.057
I(estab. produced 7-digit product in *t*-5)	2.093* (0.049) 0.391	2.326* (0.058) 0.427	2.181* (0.040) 0.399	2.241* (0.101) 0.418	2.379* (0.032) 0.427
I(in *t*-5, another estab. from the acquired firm produced the 6-digit product)	0.306* (0.074) 0.018	0.336* (0.097) 0.019	0.317* (0.059) 0.018		0.343* (0.051) 0.018
I(in *t*-5, an estab. from the acquiring firm produced the 6-digit product)	-0.054 (0.072) -0.003	0.343* (0.083) 0.020	0.090 (0.058) 0.005	0.224 (0.142) 0.013	0.065 (0.058) 0.003
I(in *t*-5, another estab. from the acquired firm produced the 7-digit product)	0.599* (0.069) 0.053	0.631* (0.087) 0.056	0.612* (0.055) 0.052		0.644* (0.047) 0.052
I(in *t*-5, an estab. from the acquiring firm produced the 7-digit product)	0.752* (0.068) 0.054	0.651* (0.077) 0.056	0.675* (0.055) 0.050	0.879* (0.139) 0.073	0.787* (0.055) 0.055
Year of merger	*t*-5 to *t*-3	*t*-2 to *t*-1	*t*-5 to *t*-1	*t*-5 to *t*-1	*t*-5 to *t*-1
Multi-unit/single unit in *t*-5?	Either	Either	Multi	Single	Either
Use Ownership Change Database to define mergers?	No	No	No	No	Yes
Approx. N	83,000	57,000	119,000	21,000	215,000
Approx. number of establishment-by-4-digit-industry groups	4,700	2,900	6,600	1,000	10,600
Pseudo R^2	0.353	0.391	0.375	0.322	0.385
Average probability that *i* produces the 7-digit good in *t*	13.2%	13.0%	13.3%	11.8%	11.8%

Notes: Each column gives the results from a separate logit regression. For each variable, coefficient estimates, standard errors, and marginal effects are reported. The dependent variable equals 1 provided establishment *i* produces 7-digit product, *p*, in year t. The sample includes all *i*-*p* pairs for which a) *i* was purchased between *t*-5 and t-1, and b) product *p* was produced at least such acquired establishment in year *t*. Control variables for total sales in year *t* of the 7-digit product (minus sales of the product by establishment *i*) are included, but not reported. See text for details. $t \in \{1992, 1997\}$. All regressions include establishment-by-4-digit-industry fixed effects. An asterisk denotes significance at the five percent level.

Table A10. Logit Regressions: Probability that Establishment *i* Ships to Zip Code *z* in 1997: Robustness Checks

I(Estab. *i* sold to zip code *z* in 1993)	2.184* (0.023) 0.155	2.249* (0.027) 0.152	2.201* (0.019) 0.151	2.304* (0.057) 0.146	1.489* (0.027) 0.132	2.263* (0.015) 0.174
I(in 1997, an estab. from the merged firm has a physical location in *z*)	0.931* (0.038) 0.035	1.066* (0.053) 0.040	1.003* (0.031) 0.038	0.600* (0.142) 0.016	1.192* (0.055) 0.095	0.991* (0.030) 0.041
ln(distance)	-0.100* (0.022) -0.002	-0.126* (0.026) -0.003	-0.100* (0.018) -0.002	-0.228* (0.052) -0.004	-0.104* (0.021) -0.005	-0.096* (0.015) -0.002
I(in 1993, another estab. from the acquired firm shipped to *z*)	0.842* (0.058) 0.030	0.746* (0.068) 0.024	0.805* (0.044) 0.027		0.777* (0.124) 0.051	0.558* (0.037) 0.018
I(in 1993, an estab. from the acquiring firm shipped to *z*)	0.458* (0.030) 0.013	0.406* (0.032) 0.011	0.434* (0.022) 0.012	0.408* (0.085) 0.010	0.595* (0.046) 0.036	0.462* (0.024) 0.014
I(in 1993, another estab. in the same 2-digit SIC, from the acquired firm shipped to *z*)	0.080 (0.090) 0.004	0.262* (0.105) 0.013	0.161* (0.068) 0.008		0.187 (0.206) 0.018	0.384* (0.052) 0.019
I(in 1993, an estab. in the same 2-digit SIC, from the acquiring firm shipped to *z*)	0.162* (0.046) 0.006	0.217* (0.051) 0.008	0.187* (0.036) 0.007	0.184 (0.129) 0.005	-0.025 (0.092) -0.002	0.158* (0.038) 0.007
I(in 1993, another estab. in the same 4-digit SIC, from the acquired firm shipped to *z*)	0.538* (0.080) 0.038	0.216* (0.094) 0.013	0.402* (0.061) 0.027		0.015 (0.173) 0.002	0.450* (0.047) 0.033
I(in 1993, an estab. in the same 4-digit SIC, from the acquiring firm shipped to *z*)	0.637* (0.053) 0.036	0.381* (0.062) 0.017	0.513* (0.042) 0.026	0.642* (0.133) 0.030	-0.013 (0.121) -0.001	0.552* (0.045) 0.032
Year of merger	92-94	95-96	92-96	92-96	92-96	92-96
Multi-unit/single unit in 1992?	Either	Either	Multi	Single	Either	Either
Manufacturing/wholesale?	Manuf.	Manuf.	Manuf.	Manuf.	Whole.	Manuf.
Use Ownership Change Database to define mergers?	No	No	No	No	No	Yes
Approx. N	869,000	589,000	1.31m	147,000	255,000	1.98m
Approx. number of establishment-by-destination counties	28,000	18,000	42,000	4,700	11,000	65,000
Pseudo R^2	0.192	0.190	0.193	0.179	0.138	0.183
Average probability that *i* ships to *z* in 1997	4.0%	4.0%	4.0%	4.0%	7.8%	4.1%

Notes: Each column gives the results from a separate logit regression. For each variable, coefficient estimates, standard errors, and marginal effects are reported. The dependent variable equals 1 provided establishment *i* ships to zip code *z* in 1997. The sample includes all *i*-*z* pairs for which *i* was purchased between 1992 and 1996, and *z* was a destination zip code for at least one such acquired establishment in 1997. Control variables for total sales in zip code *z* (minus sales from establishment *i*) are included but not reported. All regressions include establishment-by-

destination-county fixed effects. An asterisk denotes significance at the five percent level.