

# Organizing the Global Value Chain: a firm-level test

Davide Del Prete<sup>†</sup>  
(IMT Lucca)

Armando Rungi<sup>‡</sup>  
(IMT Lucca)

## Abstract

In this paper, we study the organization of Global Value Chains on a sample of about 4,000 manufacturing parent companies integrating more than 90,000 affiliates in 150 countries. Assuming a technological sequence of production stages, a recent property rights framework (Antràs and Chor, 2013; Alfaro et al., 2015) predicts that vertical integration decisions are crucially based on both the position of a supplier along the chain and on the relative size of demand elasticities faced by the final-good producer and the supplier. In line with this, we find that if final demand is sufficiently elastic (inelastic), *downstream* parents, i.e. final-good producers, integrate production stages that are more proximate to (far from) final demand. However, this result is not valid in the case of *midstream* parents, i.e. producers of intermediate inputs that can integrate either backward or forward along the chain. We document that these companies are at least as common as are *downstream* parents, but the existing theory neglects them. In these cases, we find that demand elasticities do not play a significant role in integration choices. Interestingly, both *midstream* and *downstream* parents tend to integrate affiliates that are more proximate in segments of a supply chain, probably due to technological complementarities in adjacent industries.

**Keywords:** global value chains; vertical integration; outsourcing; property rights theory; multinational enterprises; downstreamness; corporate boundaries

**JEL codes:** F14, F23, D23, G34, L20.

<sup>†</sup> Mail to [davide.delprete@imtlucca.it](mailto:davide.delprete@imtlucca.it). Laboratory for the Analysis of Complex Economic Systems, piazza San Francesco 19 - 55100 Lucca - Italy.

<sup>‡</sup> Mail to [armando.rungi@imtlucca.it](mailto:armando.rungi@imtlucca.it). Laboratory for the Analysis of Complex Economic Systems, piazza San Francesco 19 - 55100 Lucca - Italy.

# 1 Introduction

Since the late 80s, technological progress and a decrease in trade barriers have fostered the fragmentation of production on a global scale.<sup>1</sup> Networks of firms have emerged across national borders along virtually international assembly lines. From the product design to the distribution to consumers, all intermediate stages of production could be ordered in technological sequences and dispersed in several countries. Each production stage along a supply chain can be organized in two alternative ways: either keeping the stage within a corporate boundary, in case of vertical integration, or outsourcing it and engaging in arm's length contracts.

The aim of this paper is to test how firms shape the organization of Global Value Chains (GVCs). According to recent property rights frameworks<sup>2</sup> (Antràs and Chor, 2013; Alfaro et al. 2015), an optimal allocation of ownership rights along the supply chain depends crucially on positions of suppliers along the sequence and on the relative size of elasticities of final demand when compared to the elasticity of substitution across production stages. Reasonably, the theory assumes that a downstream stage cannot commence if an upstream stage does not deliver an input; therefore, contractual frictions may be present along supply chains. In fact, all intermediate producers and the final producer have to rely on a partition of the surplus extracted from the sale of the final good. Outsourcing provides suppliers with better incentives to invest in quality, but integration provides the firm with a better bargaining position by virtue of its residual rights of control. In this framework, a relation-specific investment made by any upstream supplier can affect incentives to invest by downstream suppliers. Eventually, the main prediction is that a firm integrates the production stages that are more proximate to (far from) the final consumer, if the demand for its product is sufficiently elastic (inelastic).

We find that the theory is valid for producers of final goods (i.e., *downstream parents*), starting integration backward from the bottom of a supply chain. However, we document that *midstream parents* are at least as common as are *downstream parents*. Indeed, we may think of many real-world cases of companies whose main output is an intermediate and not a final product, and can integrate other intermediate producers, both backward and forward. In this case, we find that the role of demand elasticities as a driver of integration is less relevant.

More interestingly, we find that both *downstream* and *midstream parents* systematically prefer to integrate production stages that are proximate along segments of the supply chain. The latter finding is robust across different specifications and several robustness checks. We point to a role for technological determinants that may be as important as are contracting frictions in shaping the organization of GVCs, for which extensions of the theory may be needed. We presume that economies of scope exist when companies coordinate production stages that share technological characteristics in adjacent industries, while sourcing dissimilar stages from arm's length suppliers.

To grasp the essential aspects of our findings, let us consider two case studies of *downstream parents* and two cases of *midstream parents* sourced from our data, in the upper and bottom panels, respectively, of Figure 1. For each case, we plot the relative positions on the supply chain of both the parent company and its affiliates, together with the demand elasticity faced by the parent.<sup>3</sup>

---

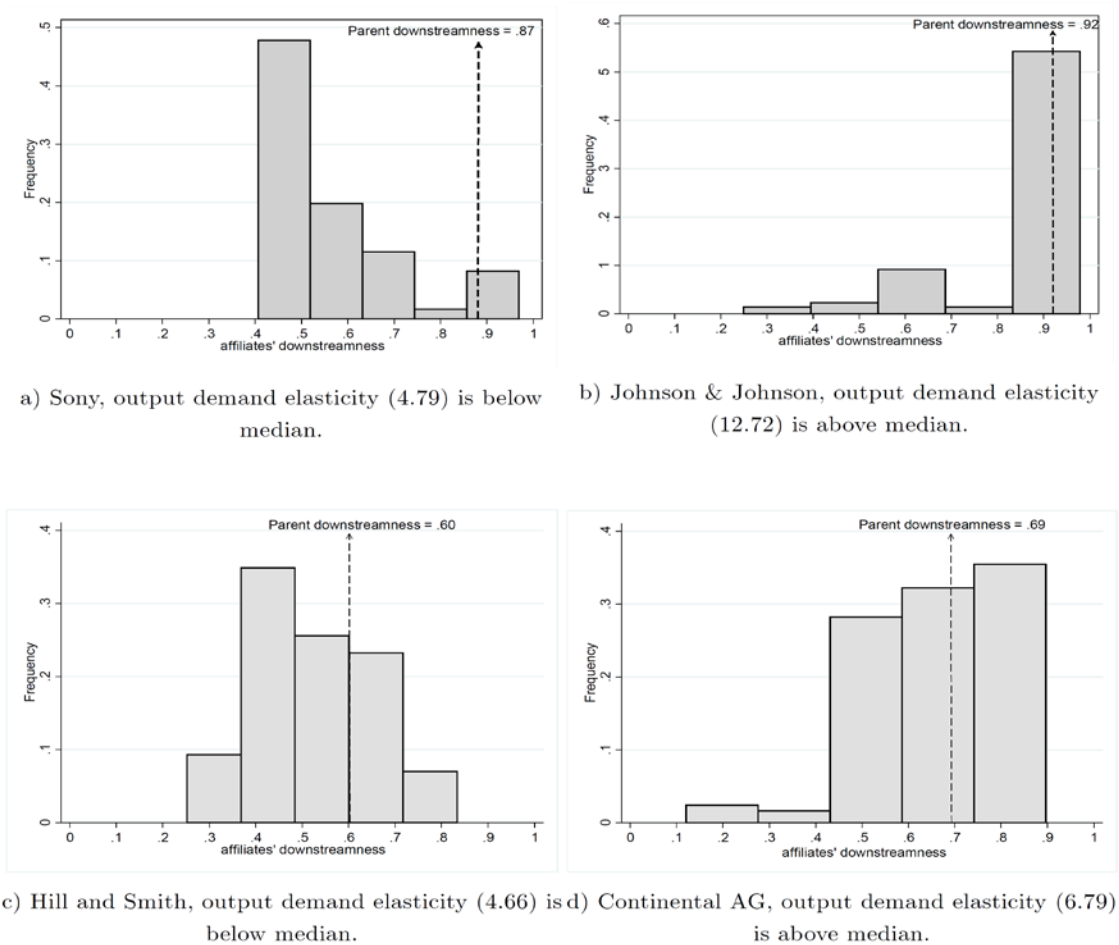
<sup>1</sup> See Baldwin and Lopez-Gonzalez (2014) and, previously, Hummels et al. (2001) for a discussion on the relevance of the phenomenon and on the structural economic changes it entails.

<sup>2</sup> For a review of the successful story of the property rights framework applied to the theory of the firm, see Aghion and Holden (2011). They retrace how Coase (1937) initiated, Williamson (1971, 1975, 1979) followed, and, finally, Grossman and Hart (1986) carved the theory of the firm in economic literature, when using theoretical models of incomplete contracts to understand what determines a firm's boundaries.

<sup>3</sup> Downstreamness metrics are sourced from Antràs and Chor (2013) and demand elasticities from Broda and Weinstein (2006). The boundaries of the firms are taken as at the end of 2012. See Section 3 and the Data Appendix for more details on our data.

Let us start from *downstream parents*. Sony and Johnson & Johnson control 405 and 353 affiliates, respectively. The former is a group that originated in Japan and is primarily focused on electronics manufacturing. The latter is a US multinational corporation producing medical devices and pharmaceutical products. From our data, both exhibit similar positions (a downstreamness of 0.87 and 0.92, respectively) on the supply chain, but they face different demand elasticities. Following the theoretical predictions by Antràs and Chor (2013) and Alfaro et al. (2015), Sony integrates more upstream, having a relatively lower demand elasticity, whereas Johnson & Johnson integrates more downstream, having a relatively higher demand elasticity.

Figure 1: Cases of *downstream* and *midstream parents*: determinants of integration



At the bottom of Figure 1, Hill & Smith Holdings PLC and Continental AG are *midstream parents* that integrate stages in both directions of the technological sequence. The first is a UK-based group active in manufacturing and supplying infrastructure products. The second is a world leading German producer of tyres, brakes, and other components for the automotive industry. They control 127 and 279 affiliates, respectively. Hill & Smith integrates relatively more upstream, whereas Continental seems to integrate relatively more downstream, but the different propensities toward vertical integration are not as remarkable, as is the closeness between parent and average affiliate position. Thus, they seem to be involved in stages that are proximate on the technological sequence.

In particular, take the case of Continental AG. It started in 1871 as a general rubber manufacturer. Over

time, it extended its boundary both backward and forward along the automotive chain. Among others, Continental AG integrated the production of brakes and chassis in 1998, and it concluded a deal with Motorola in 2006 for the control of its automotive electronics unit. Then, the company acquired the VDO brand by Siemens for powertrain and fuel injection systems in 2007. More recently, in 2015, the company extended its scope further upstream with the acquisition of the US firm Vejanca Technologies, which is a supplier of engineered rubber products.

Eventually, such proximity on the technological sequence is systematic in following analyses in both cases of *midstream* and *downstream* parent companies.

The remainder of this paper is organized as follows. In Section 2, we briefly review related works. Section 3 introduces the construction of our sample and first evidence from descriptive statistics. In Section 4, we present empirical analyses and several robustness checks. Section 5 concludes.

## 2 Literature review

Several works have already investigated the determinants of cross-border vertical integration, i.e., the global decision whether to '*make or buy*', in the context of contractual frictions.<sup>4</sup> However, it is only recently that scholars have started to consider the sequential nature of contracting and technology as an implicit feature of GVCs.

Acemoglu et al. (2007) were the first to analyze the consequences when unique headquarters have to commit to contracts with multiple suppliers, thus extending the scope of the one-shot '*make or buy*' decision of previous works. They showed that a greater contractual incompleteness leads to the adoption of less advanced technologies, even more so when intermediate inputs are highly complementary. Further efforts were made by Harms et al. (2012), who considered the role of transportation costs and their non-monotonic effects on an underlying chain of production, and by Costinot et al. (2013), who derived a sequential multi-country model in which mistakes can occur with a given probability along the chain. Hence, countries with relatively higher probabilities of making mistakes, i.e., performing more knowledge-intensive tasks, are better situated upstream in a production chain, where the destruction of value has minimal impact.

Antràs and Chor (2013) focused more specifically on the integration choices of GVCs, when suppliers perform production stages that are sequential in nature, and contracts are all potentially incomplete. For the first time, they introduced a model with technological order in production stages, modelling a dependence of downstream stages from upstream stages. If an upstream stage does not deliver the intermediate input, a downstream stage cannot produce. Eventually, a relation-specific investment made by each supplier in the sequence potentially increases the value of the final product at the end of the chain. In this framework, a firm and its suppliers bargain sequentially on the basis of an expected surplus that is realized only after the sale of the final good. On the same line, Alfaro et al. (2015) proposed a generalization of the previous model, with contractual asymmetries along the supply chain, which comes to a similar conclusion. In this context, the relative size of the elasticity of output with respect to the elasticity of substitution across production stages crucially determines the direction of integration on the supply chain.

In fact, we find evidence for the main prediction by Antràs and Chor (2013) and Alfaro et al. (2015), according to which a company producing a final good (i.e. a *downstream parent*) tends to integrate stages that are more upstream (downstream), when the elasticity demand for the output is relatively lower (higher). We also find a role for contractual asymmetries, in line with the generalization of the model by Alfaro et al. (2015), according

---

<sup>4</sup> For a detailed review of firms' organization strategies and trade, we refer to Antràs and Yeaple (2014).

to which the direction of vertical integration, i.e., whether relatively more upstream or downstream, tends to be where contracts are overall easier to enforce.

However, our results show that demand elasticities are not a significant determinant of integration in the case of *midstream* parents. In fact, we show that cases of producers of intermediate outputs that integrate both backward and forward are very common. They violate the main stringent assumption of existing theoretical models, according to which integration always starts from the bottom of the supply chain. These findings emerge only after we explicitly control for the position of both the parent and its affiliates along the chain. In this, we differ from Alfaro et al. (2015), who instead find general validation of their prediction on the role of the demand elasticity, but aggregating information about the production activities at the parent level.

Our findings are robust to different sample compositions, firm-level heterogeneity, and different institutional environments of countries of both the parent and affiliate. Thus, we also consider country-specific determinants of internalization, as suggested by existing models (see, among others, Nunn, 2007; Corcos et al., 2012)

Interestingly, we find that both *downstream* and *midstream* parent companies prefer to integrate affiliates that are close enough along segments of the supply chain. To the best of our knowledge, the latter finding is neither predicted nor tested by existing contributions on GVCs. We presume that a role is played by economies of scope in technology, because a segment of the supply chain may be easier to manage from headquarters when its industries share some technological characteristics.

### 3 Sample construction and preliminary evidence

#### 3.1 Sample construction

For the scope of our analysis, we focus on manufacturing parents that integrated an affiliate at least once in 2004–2012.<sup>5</sup> Following international standards<sup>6</sup>, we assume an affiliate is controlled by a parent company when the latter has either a direct or an indirect stake of at least 50.01% in equity. Data are sourced from the Orbis and Zephyr databases, both compiled by Bureau Van Dijk.<sup>7</sup>

From Zephyr, we consider a company to be the target of a brownfield direct investment operation by a parent company if the latter reaches either a direct or an indirect equity stake higher than 50% after an M&A deal is completed. Conversely, we consider an affiliate as a greenfield direct investment operation made by a parent company if: i) the first has been established in 2004–2012; ii) the affiliate has never been acquired by another parent company after an M&A deal since incorporation. Therefore, we consider both brownfield and greenfield investment operations as new integration choices made by a unique controlling parent. The parent company may already have a set of affiliates at the moment of new integration choices, which we include in our analyses. See Appendix A for further details on the sample and its stratification.

---

<sup>5</sup> Our choice is motivated by the fact that metrics for testing determinants of integration are available only for recent decades, possibly catching technological advances and modern consumer preferences. Indeed, using them as determinants of past integration choices may introduce an endogeneity/simultaneity bias in our estimates. It should be noted that modern corporate boundaries are often the product of century-old stories of vertical integration. From our data, the average age of companies that never integrated in 2004–2012 is about 71 years. In Appendix Table C5, we include a control group for these cases.

<sup>6</sup> The procedure for the identification of corporate boundaries is the same acknowledged in UNCTAD (2016) as a standard, following the algorithm proposed by Rungi et al. (2016). It is also in line with manuals for national statistics of MNEs proposed by OECD (2005) and UNCTAD (2009). In Appendix A and related figures, we report a descriptive example of our stratified data structure.

<sup>7</sup> Bureau Van Dijk is a consultancy firm that collects standardized firm-level information on a worldwide basis. It produces both the Zephyr database, which collects information on about 1.5 million M&A deals, and the Orbis database, which collects financial accounts for about 200 million firms.

We exclude from our baseline analysis parent companies whose output is not a manufactured product (i.e., financial companies, consultancy firms, utilities), which may not be strictly considered as organizing supply chains. By definition, stand-alone firms, who never integrated any affiliate, are also excluded from our study.

We obtained a sample of 4,192 manufacturing parents, which acquired or established 28,638 new affiliates in 2004–2012. Overall, they control a total of 90,298 affiliates in 151 countries. On average, a representative parent controls 2.63 affiliates. Approximately 63% of our sample is composed of multinational groups, for which a parent reports at least one foreign affiliate.

From our data construction, we derive three control groups that are excluded from baseline estimations but are exploited later for robustness checks: a) a sample of 72,868 direct investment operations made by non-manufacturing parents in 2004–2012, which includes parent companies not explicitly engaged in GVCs (financial companies, utilities, consultancy firms, etc); b) a sample of 20,991 portfolio investment operations, i.e., affiliates in which our parent companies acquired only minority stakes; c) a sample of 110,104 direct investment operations made by manufacturing parents in the period before 2004, when metrics of position, demand elasticities, and financial accounts are not available to us.

In Table 1, we provide a geographic coverage of our baseline sample by representative countries/areas. Details of our three control groups are included in Appendix A.

Table 1: Sample geographic coverage: origin by country of parent companies

Economy	Parents	All affiliates	Affiliates as before 2004–2012	Newly integrated affiliates
		(A+B)	(A)	(B)
<b>OECD</b>	3,356	84,047	58,492	25,555
<b>non-OECD</b>	836	6,251	3,168	3,083
<b>European Union</b>	1,405	40,375	27,986	12,389
<b>United States</b>	1,404	29,462	19,452	10,010
<b>Rest of the world</b>	1,383	20,461	14,222	6,239
<i>of which:</i>				
<b>Japan</b>	285	11,351	9,215	2,136
<b>Latin America</b>	50	543	178	365
<b>Middle East</b>	47	628	406	222
<b>China</b>	153	615	264	351
<b>Africa</b>	24	547	415	132
<b>ASEAN</b>	211	962	585	377
<b>Total</b>	<b>4,192</b>	<b>90,298</b>	<b>61,660</b>	<b>28,638</b>

Manufacturing parents are classified by their country of origin in the second column. First, we report their integration choices in 2004–2012, and then, we describe the number of affiliates they already controlled as a stock before 2004. In line with previous works studying FDI operations, we register in our sample a majority of investors in capital-abundant countries, as the bulk of parent companies originate in OECD countries. Among these, the EU

and the US report the largest shares of parents (around 34% in both cases) and controlled affiliates (45% and 33%, respectively). Among non-OECD economies, China and members of the Association of South-East Asian Nations (ASEAN) are the most represented in our sample.

For the scope of our analysis, we match firm-level primary activities at the 6-digit of the NAICS classification with similarly disaggregated industrial metrics of demand elasticity,<sup>8</sup> sourced from Broda and Weinstein (2006), and positions on supply chains, sourced from Antràs and Chor (2013). In the absence of original information on actual shipments of intermediate inputs, Antràs and Chor (2013)<sup>9</sup> turn to Input-Output tables to locate an industry along a supply chain, which is measured as the technological distance from final consumers. Metrics are normalized on a range (0,1), where 0 is the ideal start of a production line and 1 represents final consumption.

In Section 4.2, we will also use the more recent bilateral measure for position on the chain proposed by Alfaro et al. (2015),<sup>10</sup> which abandons the assumption of a unique technological sequence. The *Relative Upstreamness* is very useful to proxy the distance between each input and each output along a chain; however, it abandons the orientation to final demand. It indicates the relative distance of the input from a specific output, where the latter can also be an intermediate good. For the sake of comparison across metrics, we invert it and obtain a similar range (0,1) in *Relative Downstreamness*. Inputs approaching the value 1 are more proximate to the (final or intermediate) output. However, we will adopt the previous *Absolute Downstreamness* in baseline estimates, while keeping the novel bilateral measure as a robustness check. See Section 4.3 for comments on the complementary information provided by the estimates of both the *Absolute Downstreamness* (Antràs and Chor, 2013) and the *Relative Downstreamness* (Alfaro et al., 2015) in testing the theory.

We complement our dataset with measures that have been identified as systematic determinants for transactions within firm boundaries. First, we add parent-level variables. Labor productivity allows controlling for the fact that the most productive firms have the highest probability to integrate abroad, as suggested by Antràs and Helpman (2004) and Helpman et al. (2004). Capital intensity is usually positively associated with a larger share of intra-firm trade (Antràs, 2003). Firm size, age, number of already established affiliates, and multinational status allow us to control whether larger and older parents can more easily afford the sunk costs of a vertical integration decision (Blomström and Lipsey, 1991), especially when they are already multinational (Greenaway and Kneller, 2007).

Finally, we include a measure of *Contractibility* matched with activity of affiliates. It is sourced from Antràs and Chor (2013) and is based on the methodology of Nunn (2007). It ranks industries by easiness in reaching a deal with a supplier, in the spirit of both Rauch (1999) and Nunn and Trefler (2008). Two other country level variables control for the institutional environment of the hosting country: i) *Rule of Law* assesses the extent to which agents have confidence in and abide by country rules (Nunn, 2007); ii) *Entry Cost* measures the cost of starting a business in a country as a percentage of income per capita, both sourced from the World Bank (2015).

---

<sup>8</sup> We source from US BEA correspondence tables from NAICS to 2002 I-O industry codes (IO2002). Some companies can report more than one primary activity, based on their conglomerate nature. This is the case for 17% of parents and 33% of affiliates. In these cases, we use average values by firm. We control later in the analyses for the robustness of findings when excluding such cases.

<sup>9</sup> Metrics of position based on I-O tables were first proposed by Antràs et al. (2012) and inspired by Fan and Lang (2000); no technological sequence was assumed in these metrics. The extension of the US I-O tables to other countries is justified by the assumptions of a common technology frontier, a Leontief production function, or cross-country factor price equalization. Indeed, as in Acemoglu et al. (2009) and Alfaro et al. (2016), we can reasonably expect a strong correlation in input usage across countries. Antràs and Chor (2013) produce two alternative metrics of (absolute) downstreamness (see Appendix B for details). Here, we introduce results of the first (*DuseTuse*), because we do not find structural differences in estimates with the second (*DownMeasure*). Results with the *DownMeasure* are available upon request.

<sup>10</sup> We thank Alfaro et al. (2015) for the provision of the *Relative Upstreamness* and related codes for computation. We converted from SIC rev. 1987 to NAICS rev 2007 using correspondence tables from US Census Bureau. We merged their industry-by-industry measure with our firm-level data structure. See Section 3 and Appendix B for further details. Possible mistakes are our exclusive responsibility.

### 3.2 Parents and affiliates on supply chains

Against this background, we can detect the position of each parent and each affiliate in Figure 2, based on their industry affiliations. In Table 2, we report descriptive statistics for positions and demand elasticities. We assume that the primary activity of the (manufacturing) parent company also indicates the main activity of the corporate boundary, whereas the primary economic activities of affiliates represent production stages that have been integrated.<sup>11</sup> Later in our analyses, we perform robustness checks of the validity of these assumptions.

Figure 2: Position of parent companies and affiliates along the supply chain

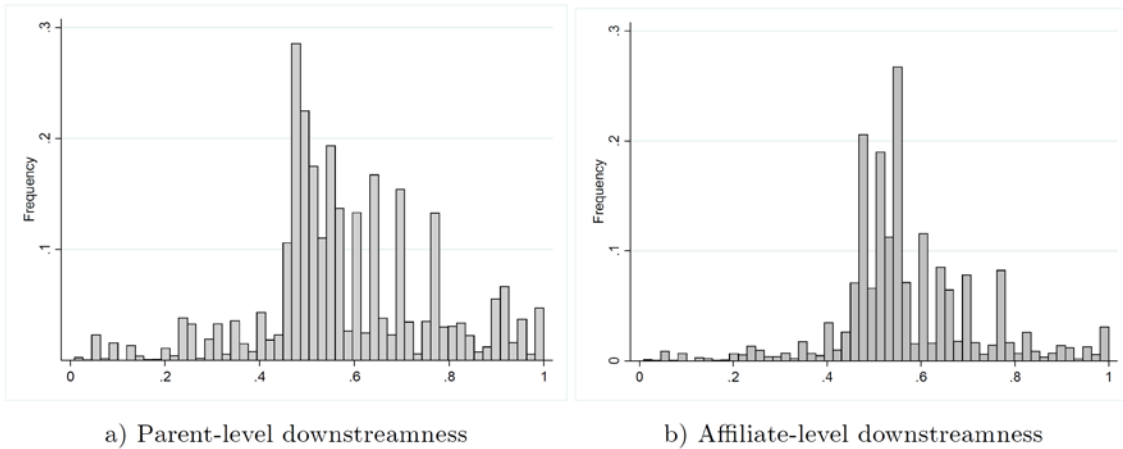


Table 2: Absolute Downstreamness and elasticities of substitution across firms

Metrics		Obs.	Mean	Median	St. dev.	Min	Max
<b>Parents</b>	Downstreamness	4,192	.63	.65	.22	.07	1
	Elasticity of substitution	4,192	8.85	5.97	9.99	1.3	108.5
<b>Affiliates</b>	Downstreamness	90,298	.58	.54	.16	.01	1
	Elasticity of substitution	90,298	7.39	3.76	8.39	1.3	108.5
<b>Parent vs. affiliates (within)</b>	Difference in Downstreamness	90,298	.05	.01	.24	-.91	.94
	Difference in Elasticity of substitution	90,298	1.51	.00	11.60	-95.78	105.46

From the combined reading of Figure 2 and Table 2, we document some important descriptive findings. More often than not, the output of a representative parent company is an intermediate good. The average *Absolute Downstreamness* for parents is 0.63. The representative company starting integration is indeed located *midstream*

<sup>11</sup> More than one affiliate controlled by a parent can perform the same production stage within its corporate boundary. We find about 23% of cases in our sample, when the activity by an affiliate is a duplicate at the industry-country level. Different qualities of inputs and/or geographic proximity to factories may determine multiple integrations of the same stage.



on the technological sequence.

In the case of affiliates, the average *Absolute Downstreamness* is 0.58. If we calculate the average difference in downstreamness<sup>12</sup> of each parent from its affiliates (i.e., the difference within a corporate boundary), we obtain a value of 0.05. That is, a representative parent is slightly more downstream than are its affiliates. At the same time, a parent is, on average, proximate to its integrated affiliates along the supply chain. The median values and the standard deviations support these findings.

This evidence calls for a refinement of existing theoretical models, according to which a parent company is usually a producer of final goods starting integration backwards, as in Antràs and Chor (2013) and Alfaro et al. (2015). Further, proximity on the supply chain has been rather neglected in previous studies on GVCs. Thus, our evidence implies a fragmentation of supply chains by segments, either integrated or not integrated within different corporate boundaries. Table 2 further includes descriptives of demand elasticities faced by parents and affiliates, which do not show significant differences in distributions.

## 4 Empirical results

### 4.1 Baseline and robustness checks

Our aim is to explain the position of an integrated affiliate along a value chain as a function of the position of the parent and its demand elasticity. We estimate the following baseline equation, augmented with firm, industry, and country level information:

$$\begin{aligned} \text{downstreamness}_{i(j)c} = & \beta_0 + \beta_1 \text{downstreamness}_j + \beta_2 \text{complements}_j + \\ & + \beta_3 (\text{complements}_j \cdot \text{downstreamness}_j) + \beta_4 Z_j + \beta_5 W_i + \beta_6 V_c + \varepsilon_{i(j)c} \end{aligned} \quad (1)$$

where  $\text{downstreamness}_{i(j)c}$  is the  $i$ th affiliate absolute downstreamness integrated by the  $j$ th parent and operating in country  $c$ . Among independent variables,  $\text{downstreamness}_j$  stands for  $j$ th parent absolute downstreamness, and  $\text{complements}_j$  is a latent variable for parent demand elasticity, equal to one when above the median and zero otherwise, which is calculated following theory (Antràs and Chor, 2013). We split the sample by industry and use the median elasticity value (5.97) to reproduce cases of sequential substitutes (when elasticity is below the median) and sequential complements (when elasticity is above the median). Further, we interact  $\text{complements}_j$  with the position of the  $j$ th parent along the supply chain. The inclusion of an interaction term is pivotal to derive marginal effects of demand elasticity (taking into account coefficients  $\beta_2$  and  $\beta_3$  with average parent downstreamness) and of parent downstreamness (taking into account  $\beta_1$  for sequential substitutes and  $\beta_1 + \beta_2$  for sequential complements). Additionally,  $Z_j$  collects parent-level controls in logarithmic scale, which are as follows: labor productivity, capital intensity, firm size, age, number of already established affiliates, and multinational status. In  $W_i$ , we collect the affiliate's industry contractibility, and in  $V_c$ , we include host country controls for institutional environments, possibly substituted by country-fixed effects in nested specifications,

---

<sup>12</sup> Hereafter, downstreamness indicates *Absolute Downstreamness*, if nothing else is specified. *Relative Upstreamness* is used from Section 4.2 onward.

allowing us to control for country and industry-specific determinants of internalization, as suggested by Nunn (2007) and Corcos et al., (2012). Results are reported in Table 3.

We find that parents producing final goods integrate suppliers that enter further downstream, when output demand is sufficiently elastic. This result is in line with basic predictions by Antràs and Chor (2013) and, more recently, by Alfaro et al. (2015). If we take the estimates of column 3 of Table 3 as reference, the marginal effect of parent elasticities on affiliate downstreamness is positive for values of parent downstreamness that are higher than 0.57. That is, the prediction is valid for the segment of downstreamness where we will likely find producers of final goods.

Table 3: Baseline estimates

Dependent variable:	(1)	(2)	(3)	(4)	(5)
affiliate downstreamness $i(j)c$	OLS	OLS & firm controls	OLS, firm controls & fe	OLS, firm controls & fe	OLS, firm controls & fe
parent downstreamness $j$	.150*** (.019)	.224*** (.036)	.231*** (.033)	.372*** (.031)	.188*** (-.051)
complements $j$	-.040*** (.015)	-.055** (.028)	-.062** (.026)	.023 (.027)	-.069* (.038)
complements $j$ *parent downstreamness $j$	.060** (.026)	.102** (.044)	.108*** (.041)	-.043 (.047)	.138** (.059)
(log of) labor productivity $j$		.009* (.005)	.007* (.004)	.007 (.008)	.010** (.005)
(log of) capital intensity $j$		-.011** (.005)	-.008 (.005)	-.015** (.007)	.014* (.007)
(log of) employment $j$		.006* (.003)	.006** (.003)	.009** (.004)	.009** (.004)
(log of) number of affiliates $j$		-.007* (.004)	-.006 (.003)	-.005 (.004)	-.018*** (.005)
(log of) age $j$		-.005 (.005)	-.004 (.004)	.005 (.006)	-.012*** (.005)
multinational group $j$		-.034*** (.011)	-.042*** (.010)	-.061*** (.013)	.002 (.015)
contractibility $i$		-.190* (.099)	-.205*** (.028)	-.309*** (.031)	-.038 (.041)
rule of law $c$		-.012 (.013)			
contractibility*rule of law $ic$		-.001 (.122)			
entry cost $c$		-.004*** (.001)			
Constant	.487*** (.012)	.518*** (.044)	.206*** (.053)	.357*** (.049)	.530*** (.047)
Observations	90,274	30,340	39,808	19,482	15,111
Adjusted R-squared	.082	.158	.174	.224	.113
Affiliate country f.e.	NO	NO	YES	YES	YES
BEC category of parent	All	All	All	Intermediates	Final goods

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* represent p-value < 0.01, p-value < 0.05, and p-value < 0.10, respectively.

Apparently, the marginal effect could become negative below that threshold, where we should find only parents producing intermediate goods, i.e., *midstream parents*. However, the impact of demand elasticity is not significant for integration decisions by *midstream parents* in column 4 of Table 3, when we explicitly differentiate the output on the basis of the Broad Economic Categories (BEC) classification, between producers of intermediate and final goods. That is, the prediction of the theory in these cases is no more valid, as no statistical significance is found for the variable ‘complements’ and its interaction term. On the other hand, separate estimates for *downstream parents* for producers of final goods in column 5 report a marginal effect of ‘complements’ on affiliate downstreamness equal to 0.5, which is comparable to general results in column 3. Indeed, we find no *downstream parent* below this threshold.

More interestingly, we find that parents tend to integrate production stages that are proximate on the supply chain. A positive correlation exists between the downstreamness of a parent and the downstreamness of each of its integrated affiliates. Such a proximity is higher in the case of sequential complements, because the interaction term is positive. We believe that proximity on the technological sequence is probably due to the existence of economies of scope in similar technologies,<sup>13</sup> when it is easier for a parent company to manage activities on the same segment, because they share some technological features. Conversely, it may be more costly to coordinate production stages when they are dissimilar in technology from the core activity of the company.

To the best of our knowledge, such a proximity on the supply chain in vertical integration has been neglected in previous studies on GVCs, although the possibility to integrate both backward and forward has not been excluded in previous works on vertical integration (see, for example, Acemoglu et al., 2007, and Acemoglu et al., 2010).<sup>14</sup>

In Table 4, we test whether our findings are robust to different sample compositions. In the Appendix Table C4 and Table C5, we repeat the same exercises for *downstream* and *midstream* parents only, respectively.

Starting from the first column, we include fixed effects by origin of the parent to control for country-specific characteristics, especially relevant in the case of multinational enterprises. We alternatively compute ‘complements’ as a binary variable equal to 1 when  $(\rho > \alpha)$ , where  $\alpha$  is the median elasticity of affiliates, in the spirit of the original theoretical setups. In column 3, we include only the top 5% of affiliates classified in terms of employment, because Ramondo et al. (2016) found that only a few large affiliates are responsible for the bulk of US intrafirm trade. In column 4, we exclude affiliates engaged in service industries, for possible errors of measurement in downstreamness. Then, we exclude affiliates from the same sector of the parent to eliminate possible horizontal integration strategies. In column 6, we include only the top 100 inputs specific for each parent output, as from IO tables, and parents with only one primary activity, to eliminate possible cases of conglomerates. In column 7, we focus on affiliates that are more upstream than the parent, i.e., excluding the case of buyers of the parent output. In column 8, we select only affiliates integrated by the same parents, but done so after 2004. In column 9, we exclude affiliates that are indirectly controlled by the parent through other affiliates (i.e., affiliates of affiliates, see Appendix A).

Across all columns in Table 4, the role of demand elasticities is always statistically significant. In fact, the marginal effect of the ‘complements’ variable is positive only for parent downstreamness that is either higher than

---

<sup>13</sup> The proximity of parents and affiliates in terms of industrial activities is consistent with findings by Alfaro and Charlton (2009). Their explanation is that many inputs can come from the same sector, even after narrow and disaggregate definitions of the latter. However, our results are robust also when we exclude affiliates from the same sector of the parent, in column 5 of Table 4.

<sup>14</sup> At the moment we are finalizing our paper, two on-going works that are worth mentioning. These works may help extend the theory according also to our own findings. MengXiao (2017) models and tests the cases of firms integrating both backward and forward in a context of contractual imperfections. The author finds that excluding these cases from previous studies may bias estimates. On the other hand, Berlingieri et al. (2017) introduce technological determinants as complementary to contractual imperfections in vertical integration decisions.

0.52, which is the least conservative case in column 7, or higher than 0.65, which is the most conservative case in column 5. Looking at similar exercises reported in Appendix Table C4, and in accordance with results reported in column 3 of Table 3, we conclude that the results are valid for *downstream parents*, while no statistical significance is detected for the role of output demand faced by the parent company when integration choices are made by *midstream parents* (Appendix Table C5).

Table 4: Variants on sample composition

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
affiliate downstreamness <sub><i>i(j)c</i></sub>	parent country FE	output & input elasticities	top 5% affiliates	only manuf affiliates	different sector	top 100 inputs	only upstream to parent	only new affiliates	only direct control
<b>parent downstreamness<sub><i>j</i></sub></b>	.234*** (.031)	.231*** (.031)	.357*** (.037)	.317*** (.034)	.069** (.029)	.312*** (.037)	.529*** (.030)	.171*** (.032)	.229*** (.039)
<b>Complements<sub><i>j</i></sub></b>	-.055** (.023)	-.062** (.025)	-.104*** (.031)	-.056* (.029)	-.082*** (.024)	-.061** (.029)	-.045* (.026)	-.094*** (.024)	-.041* (.022)
<b>complements*parent downstreamness<sub><i>j</i></sub></b>	.102*** (.037)	.100*** (.038)	.161*** (.046)	.094** (.044)	.126*** (.038)	.110** (.045)	.085** (.042)	.161*** (.038)	.081* (.046)
<b>(log of) labor productivity<sub><i>j</i></sub></b>	.008* (.004)	.006 (.004)	.004 (.005)	.010* (.005)	.001 (.005)	.008* (.005)	-.000 (.004)	.009*** (.003)	.011** (.005)
<b>(log of) capital intensity<sub><i>j</i></sub></b>	-.008 (.005)	-.007 (.005)	-.005 (.005)	-.006 (.006)	-.006 (.005)	-.009* (.006)	-.012** (.005)	-.011*** (.004)	-.011** (.005)
<b>(log of) employment<sub><i>j</i></sub></b>	.006** (.003)	.007** (.003)	.007** (.003)	.010*** (.004)	.005* (.003)	.003 (.003)	.004 (.003)	.005* (.003)	.008** (.003)
<b>(log of) number of affiliates<sub><i>j</i></sub></b>	-.005 (.003)	-.005 (.003)	-.009** (.004)	-.006 (.004)	-.004 (.003)	-.004 (.004)	-.007** (.004)	-.007** (.003)	-.010*** (.004)
<b>(log of) age<sub><i>j</i></sub></b>	-.003 (.004)	-.004 (.004)	-.004 (.005)	-.007 (.005)	.001 (.005)	-.005 (.005)	-.007 (.005)	-.004 (.004)	.001 (.005)
<b>multinational group<sub><i>j</i></sub></b>	-.044*** (.010)	-.044*** (.010)	-.021 (.013)	-.048*** (.011)	-.037*** (.011)	-.029*** (.011)	-.002 (.014)	-.006 (.011)	-.014 (.012)
<b>contractibility<sub><i>i</i></sub></b>	-.204*** (.027)	-.202*** (.028)	-.110*** (.029)	-.145*** (.029)	-.265*** (.027)	-.189*** (0.043)	-.083*** (.021)	-.207*** (.026)	-.188*** (.036)
<b>Constant</b>	.758*** (.080)	.198*** (.050)	.160*** (.050)	.623*** (.049)	.767*** (.060)	.185*** (.060)	.501*** (.063)	.241*** (.051)	.202*** (.045)
<b>Observations</b>	39,808	39,808	3,021	31,124	32,701	21,337	17,014	11,732	4,069
<b>Adjusted R-squared</b>	.187	.173	.332	.213	.092	.262	.389	.181	.195
<b>Parent country f.e.</b>	YES	NO	NO	NO	NO	NO	NO	NO	NO
<b>Affiliate country f.e.</b>	YES	YES	YES	YES	YES	YES	YES	YES	YES
<b>BEC category of parent</b>	All	All	All	All	All	All	All	All	All

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* represent p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Once again, we can confirm a positive and significant correlation between the position of the parent and its affiliates. The most conservative case from Table 4 is when we exclude affiliates from the same sector of the parent. In this case, the marginal effect of parent downstreamness on affiliate downstreamness is 0.07 for sequential substitutes (when output elasticity is below the median) and 0.20 for sequential complements (when output elasticity is above the median). The least conservative case from Table 4 is when we include upstream

affiliates only. In this case, marginal effects for sequential substitutes and sequential complements are 0.53 and 0.61, respectively.

Among other controls, we emphasize the negative coefficient that is always present on affiliate-level contractibility, which implies that upstream suppliers are, on average, more contractible. In Section 4.3, we will explicitly and positively test for asymmetries in contractibilities, following the generalization in theory by Alfaro et al. (2015). In this case, we find that parents prefer integrating in the direction in which stages are more contractible.

Finally, in Appendix Table C6, we check whether results are still valid when introducing control groups for firms that were initially excluded from our sample (See Section 3.1 and Appendix A for details).

As expected, the ‘complements’ variable, alone and in interaction, is no longer significant in the cases of (1) service parents, because they may include firms that are not integrating affiliates for the organization of GVCs (e.g., financial companies, utilities, consultancy firms, etc.); (2) portfolio investment operations, when manufacturing parent companies have only a minority stake, because the company’s interest may not be in the coordination of that production stage; and (3) manufacturing parents that never integrated after 2004, which is the period for which we have both metrics for determinants and firm-level controls. In the latter case, simultaneity/endogeneity bias may be present for tests against *ex post* regressors. Nonetheless, a correlation between positions of a parent and its affiliates persists in all cases.

## 4.2 Relative downstreamness

In this section and the next, we use a more recent measure for positioning along the value chain developed by Alfaro et al. (2015). The new indicator originally measures *Relative Upstreamness* of an  $m$ th input, with respect to the  $n$ th output, based on 1992 U.S. I-O Tables.

For our scope, this measure is complementary to the *Absolute Downstreamness* we have used in previous estimates. It is very useful, because it abandons the strong assumption of a unique technological sequence for all outputs, catching the technological distance between any two industries that have an input-output relationship, thereby making the technological distance specific and oriented toward each output. In other words, each output now has an upstream technological sequence, with inputs located at varying distances.

However, this means that the orientation to final demand is lost and an arbitrary  $n$ th output can be just another intermediate stage, either a part or a component, used further downstream by either another intermediate or a final producer. That is, the *Relative Upstreamness* may measure the distance between two intermediate stages of production, once assuming that one sources from the other. From our point of view, this measure better proxies the proximity between outputs and inputs, rather than the distance of production stages from final consumers.

*Relative Upstreamness* originally ranges in an interval  $(0, +\infty)$ , such that a higher level of the  $m$ th input, given the  $n$ th output, implies that the two industries are more distant. For the sake of comparison with previous analyses, we consider its inverse, hence falling in a standardized range  $(0, 1)$ . In this case, a higher level of *Relative Downstreamness* of the  $m$ th input implies that the latter is more proximate to the  $n$ th output.

After matching with parent and affiliate data on industries, we can now consider the *Relative Downstreamness* of an affiliate with respect to its parent as a dependent variable in the following equation:

$$relative\_downstreamness_{i(j)c} = \beta_0 + \beta_1 complements_j + \beta_2 Z_j + \beta_3 W_i + \beta_6 V_c + \varepsilon_{i(j)c} \quad (2)$$

where any  $i$ th affiliate producing an  $m$ th input is integrated by the  $j$ th parent producing the  $n$ th output. In this case, we do not need to introduce the position of the parent company as a further covariate, because this information is already part of the dependent variable. Moreover,  $Z_j$ ,  $W_i$  and  $V_c$  collect parent, affiliate and country controls, respectively, as in eq. (1). Nested results are reported in Table 5 in the case of producers of final goods, i.e., *downstream* parents.

Table 5: Relative downstreamness for downstream parents

<b>Dependent variable:</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>
<b>relative downstreamness of affiliate <math>i(j)c</math></b>	<b>OLS</b>	<b>OLS and firm controls</b>	<b>OLS, firm controls &amp; fe</b>
<b>complements <math>j</math></b>	.037** (.015)	.065*** (.021)	.051*** (.018)
<b>(log of) labor productivity <math>j</math></b>		.030* (.017)	.030** (.015)
<b>(log of) capital intensity <math>j</math></b>		.013 (.015)	.009 (.014)
<b>(log of) employment <math>j</math></b>		.017** (.009)	.013 (.008)
<b>(log of) number of affiliates <math>j</math></b>		-.024*** (.008)	-.019** (.008)
<b>(log of) age <math>j</math></b>		-.007 (.013)	-.014 (.012)
<b>multinational group <math>j</math></b>		-.060* (.035)	-.079** (.031)
<b>contractibility <math>i</math></b>		-.144 (.258)	-.154** (.075)
<b>rule of law <math>c</math></b>		-.150*** (.035)	
<b>contractibility*rule of law<math>ic</math></b>		.038 (.300)	
<b>entry cost <math>c</math></b>		.005 (.003)	
<b>Constant</b>	.483*** (.008)	.396*** (.086)	.721*** (.089)
<b>Observations</b>	21,316	6,796	9,643
<b>Adjusted R-squared</b>	.009	.084	.086
<b>Affiliate country f.e.</b>	NO	NO	YES
<b>BEC category of parent</b>	Final goods	Final goods	Final goods

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* represent p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

In fact, no assumptions could be made for *midstream* parents, because the use of *Relative Downstreamness*<sup>15</sup> prevents us from making reference to distance from the final demand, as from the existing

<sup>15</sup> It should be noted that, by construction, the relative upstreamness/downstreamness requires an *ex ante* assumption on the orientation between an  $m$ th industry and an  $n$ th industry, regarding which is the input and which is the output. Once the metrics match the parent-affiliate data, we implicitly assume that the output is produced by the parent and its affiliates are possible upstream suppliers (never buyers). In this, the ‘absolute downstreamness’ allows us more flexibility for including cases of forward integration by parents. However, in Appendix Table C7, we also report the case of *midstream parents*, where a reversal of sign is detected for the ‘complements’ variable,

theory. Further, in this case, the results confirm the main theoretical prediction, according to which if *downstream* parent companies face a sufficiently elastic demand, they tend to own suppliers located relatively close to their final output.

### 4.3 Non-integrated stages

Finally, we control whether our results are robust when we explicitly consider production stages that were never integrated. In fact, thus far, we tested integration by only observing *ex post* which affiliates and industries a parent company chose to integrate. In other words, in absence of data on external suppliers, we relied on data for integrated affiliates.

Now, we modify our data structure, in line with Alfaro et al. (2015), to consider also the production stages that could have been possibly integrated, but which, eventually, were not. We lose the parent and affiliate level dimension to switch to a parent-industry dimension. Our unit of observation is now a pair comprising a parent company (taken with its output) and any possible input industry, as derived from input-output tables.<sup>16</sup>

We test a conditional logit model<sup>17</sup> with parent-level fixed effects, in the following equation:

$$INT_{mj} = \beta_0 + \beta_1 downstreamness_{mj} + \beta_2 complements_j + \beta_3 (downstreamness_{mj} \cdot complements_j) + \beta_4 upto_{mj} + \beta_5 (upto_{mj} \cdot complements_j) + \beta_6 N_j + \eta_j + \varepsilon_{mj} \quad (3)$$

where  $INT_{mj}$  is a binary variable equal to one when the parent  $j$  has integrated at least one affiliate active in the input industry  $m$ . If the parent has not integrated any affiliate in the input industry, that input is assumed to be outsourced. Among regressors,  $downstreamness_{mj}$  is the  $m$ th input's *Relative Downstreamness* of parent  $j$  active in the  $n$ th industry.  $complements_j$  is a binary variable that takes the value 1 for sequential complements, when elasticity of output is above the median of all inputs, either integrated or not integrated by a parent firm ( $\rho > \rho_{med}$ ), and 0 for sequential substitutes, when elasticity of output is below the median ( $\rho < \rho_{med}$ ). The variable  $upto_{mj}$  is the cumulative contractibility for all the ordered upstream stages included up to the  $m$ th input, as in Alfaro et al. (2015).  $N_j$  is a binary variable that is equal to one if the input belongs to the same industry as the parent. Finally, we introduce a full set of parent-level fixed effects ( $\eta_j$ ). Predicted probabilities for nested specifications are reported in Table 6.

First, the results confirm that an input is more likely integrated if it is proximate to the parent output along

---

although no assumption can be made about either the nature of this correlation or the role of demand elasticity from existing theory.

<sup>16</sup> In line with Alfaro et al. (2015), we select the top 100 inputs for each output, as from coefficients of I-O tables. Thus, we exclude possible activities that are loosely related to the output of the parent, as in the case of conglomerates, for example. Considering only primary activities, on average, a representative parent integrates 3 input industries out of the top 100. We further test whether the inclusion of both primary and secondary activities at the parent level and affiliate level, as originally done by Alfaro et al. (2015), does not change our results. There, the parent integrates an average of 4 input industries out of the top 100. Results are similar in magnitude and statistical significance.

<sup>17</sup> Results are robust to different specifications, including the linear probability model originally tested by Alfaro et al. (2015), the probit model, and the logit model. These results are available upon request. We report findings of the conditional logit, because it fits better the multinomial case of *ex ante* alternatives. See McFadden (1974) for more details and Head et al. (1995) for an application.

the technological sequence, possibly due to some shared economies of scope in technologies.<sup>18</sup>

Second, the role of demand elasticity is found to be significant overall, but this result is no longer valid for *downstream parents* when we introduce a measure of cumulate contractibility up to the input  $m$  (Table 6, columns 2 and 4). This is in line with the result of Alfaro et al. (2015), for which asymmetries in contractibility along the technological sequence can be a further determinant of integration choices. In this framework, a greater degree of contractibility of upstream inputs increases the propensity of a firm to integrate in that direction, where input  $m$  can be found. Note how these findings are also consistent with previous estimates, starting from baseline in Table 3, where we find that (integrated) upstream suppliers are, on average, more contractible.

Unfortunately, no assumption can be made about the role of elasticity of demand for *midstream parents* and their direction of integration, as reported in Appendix Table C8, because the use of *Relative Downstreamness* prevents us from making reference to distance from the final demand, as from the existing theory. See also the comments and the note to Table 5.

Table 6: Fixed effects conditional logit – integrated and non-integrated stages by downstream parents

Dependent variable:	(1)	(2)	(3)	(4)
<b>Input <math>m</math> integrated by parent <math>j</math> (Yes/No)</b>				
<b>relative downstreamness</b> $m_j$	.383*** (.027)	.430*** (.034)	.393*** (.023)	.443*** (.029)
<b>complements*relative downstreamness</b> $m_j$	.166*** (.039)	.065 (.050)	.154*** (.033)	.037 (.042)
<b>contractibility up-to-input</b> $m_j$		.007 (.005)		.005 (.004)
<b>contractibility up-to-input*complements</b> $m_j$		.016** (.007)		.021*** (.006)
<b>Dummy: Self-NAICS</b> $m_j$	.740*** (.022)	.770*** (.025)	.671*** (.019)	.703*** (.021)
<b>Observations (parent per stage)</b>	101,082	93,910	113,035	105,546
<b>Pseudo R-squared (McFadden's)</b>	.186	.198	.144	.155
<b>Parent-level f.e.</b>	YES	YES	YES	YES
<b>Activities integrated</b>	Primary	Primary	Primary and secondary	Primary and secondary
<b>BEC category of parent</b>	Final goods	Final goods	Final goods	Final goods

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

## 5 Conclusions

Assuming a technological sequence of ordered inputs leading to a final output, recent theory predicts that decisions of (backward) integration are based crucially on the role of demand elasticity of the final producer, in the presence of contractual frictions.

<sup>18</sup> Please note the change in the interpretation of coefficients. In the case of simple *downstreamness*, a correlation between a parent and an affiliate reveals an *ex post* proximity on the supply chain, as in Section 4.1 and 4.2. After the adoption of a conditional logit model, *relative downstreamness* reveals a positive and significant probability that (industries of) a parent and (industries of) its affiliates are proximate.



On the one hand, we provide empirical evidence at the firm level that supports the main theoretical prediction (Antràs and Chor, 2013; Alfaro et al., 2015), according to which producers of final goods engage in (backward) vertical integration, preferably more downstream, when the elasticity of final output is sufficiently high. On the other hand, we argue that cases of *midstream parents* are neglected by existing works on GVCs. The latter are companies that specialize in intermediate inputs, possibly integrating both backward and forward suppliers, starting from the middle of the supply chain. We report that they are at least as common as are cases of *downstream parents*. We may think of many real-world cases of companies that integrate intermediate inputs and do not directly sell to final consumers. In these cases, we find no significant role for demand elasticities as drivers of integration choices.

Instead, we find that both *downstream* and *midstream parents* systematically prefer to integrate production stages that are proximate along segments of the supply chain. Our findings point to a role played by technological determinants, which may be as important as contracting frictions in shaping firm-level organization of GVCs. In fact, we presume that it is less costly to coordinate production stages from headquarters when these stages share some technological characteristics, as coming from adjacent industries, while sourcing dissimilar stages from arm's length suppliers.

We believe that exploring technological complementarities among suppliers on a technological sequence is both a theoretical and an empirical challenge for future works, which would help to better understand patterns of vertical integration within and across national borders.

## 6 Acknowledgements

We would like to thank Pol Antràs, Laura Alfaro, Davin Chor, Paola Conconi, Daniel Trefler, Joanne Oxley, Kunal Dasgupta, Gianmarco Ottaviano, Dalia Marin, Carlo Altomonte, Giorgia Giovannetti, Anna Giunta, Giuseppe De Arcangelis, and Pierluigi Montalbano, in addition to participants at the 'World Bank/CEPR Second Conference on Global Value Chains, Trade and Development 2016', 'Royal Economic Society 2016', 'Warsaw International Economic Meeting 2015', 'European Trade Study Group 2014', 'Explaining Economic Change 2014 - Rome', and also at various seminars, for useful comments. The authors claim financial support by CrisisLab PNR Project at IMT School for Advanced Studies, Lucca - Italy.

## References

- Acemoglu, D., P. Antràs, and E. Helpman (2007). "Contracts and Technology Adoption". *American Economic Review* 97(3), 916–943.
- Acemoglu, D., R. Griffith, P. Aghion, and F. Zilibotti (2010). "Vertical integration and technology: theory and evidence". *Journal of the European Economic Association* 8(5), 989–1033.
- Acemoglu, D., S. Johnson, and T. Mitton (2009). "Determinants of vertical integration: financial development and contracting costs". *The Journal of Finance* 64(3), 1251–1290.
- Aghion, P. and R. Holden (2011). "Incomplete contracts and the theory of the firm: What have we learned over the past 25 years?". *The Journal of Economic Perspectives* 25(2), 181–197.
- Alfaro, L., P. Antràs, D. Chor, and P. Conconi (2015). "Internalizing Global Value Chains: A Firm-Level Analysis". *NBER Working Paper No. w21582*.
- Alfaro, L. and A. Charlton (2009). "Intra-industry Foreign Direct Investment". *American Economic Review* 99(5), 2096–2119.
- Alfaro, L., P. Conconi, H. Fadinger, and A. Newman (2016). "Do Price Determine Vertical Integration?". *Review of Economic Studies* 83, 1–35.
- Altomonte, C. and A. Rungi (2013). "Business Groups as Hierarchies of firms: Determinants of Vertical Integration and Performance ". *ECB Working Paper No. 1554*.
- Alvarez, V., J. Cravino, and A. A. Levchenko (2016). "The Growth of Multinational Firms in the Great Recession". *Journal of Monetary Economics* 85, 50–64.
- Antràs, P. (2003). "Firms, Contracts, And Trade Structure". *The Quarterly Journal of Economics* 118(4), 1375–1418.
- Antràs, P. and D. Chor (2013). "Organizing the Global Value Chain". *Econometrica* 81(6), 2127–2204.
- Antràs, P., D. Chor, T. Fally, and R. Hillberry (2012). "Measuring the Upstreamness of Production and Trade Flows". *The American Economic Review* 102(3), 412–416.
- Antràs, P. and E. Helpman (2004). "Global Sourcing". *Journal of Political Economy* 112(3), 552–580.
- Antràs, P. and S. Yeaple (2014). "Multinational Firms and the Structure of International Trade". *Handbook of International Economics* 4(55).
- Baldwin, R. and J. Lopez-Gonzalez (2014). "Supply-chain Trade: A Portrait of Global Patterns and

Several Testable Hypotheses". *The World Economy*.

Berlingieri, G., F. Pisch, and C. Steinwender (2017). "Technological Determinants of the Make-Or-Buy Decision". *Mimeo*.

Blomström, M. and R. E. Lipsey (1991). "Firm size and foreign operations of multinationals". *The Scandinavian Journal of Economics*, 101–107.

Broda, C. and D. Weinstein (2006). "Globalization and the Gains From Variety". *Quarterly Journal of Economics* 121(2), 541–585.

Coase, R. H. (1937). "The nature of the firm". *Economica* 4(16), 386–405.

Corcos, G., D. M. Irac, G. Mion, and T. Verdier (2013). "The determinants of intrafirm trade: Evidence from French firms". *Review of Economics and Statistics* 95(3), 825–838.

Costinot, A., J. Vogel, and S. Wang (2013). "An elementary theory of global supply chains". *The Review of Economic Studies* 80(1), 109–144.

Cravino, J. and A. A. Levchenko (2016). "Multinational Firms and International Business Cycle Transmission". *The Quarterly Journal of Economics*, forthcoming.

Fan, J. P. H. and L. H. P. Lang (2000). "The measurement of relatedness: An application to corporate diversification". *The Journal of Business* 73(4), 629–660.

Greenaway, D. and R. Kneller (2007). "Industry differences in the effect of export market entry: learning by exporting?". *Review of World Economics* 143(3), 416–432.

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

Harms, P., O. Lorz, and D. Urban (2012). "Offshoring along the production chain". *Canadian Journal of Economics/Revue canadienne d'économie* 45(1), 93–106.

Head, K., J. Ries, and D. Swenson (1995). "Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States". *Journal of International Economics* 38 (3-4), 223-247.

Helpman, E., M. Melitz, and S. Yeaple (2004). "Export versus FDI with heterogeneous firms". *American Economic Review* 94.

Hummels, D., J. Ishii, and Kei-mu. Yi (2001). "The nature and growth of vertical specialization in world trade". *Journal of International Economics* 54, 75–96.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2011). "The worldwide governance indicators: methodology and analytical issues". *Hague Journal on the Rule of Law* 3(02), 220–246.

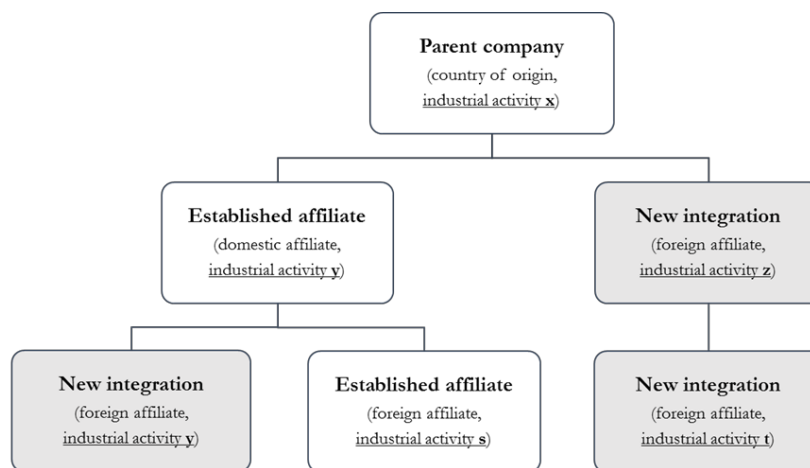
- MengXiao, L. (2016). "Global Supply Chains without Principals". *Mimeo*.
- McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior". In *Frontiers in Econometrics*, New York:Academic Press
- Nunn, N. (2007). "Relationship-specificity, incomplete contracts, and the pattern of trade". *Quarterly Journal of Economics* 122(2), 569–600.
- Nunn, N. and D. Trefler (2008). "The Boundaries of the Multinational Firm: An Empirical Analysis". In E. Helpman, D. Marin, and T. Verdier (Eds.), *The Organization of Firms in a Global Economy*. Cambridge University Press.
- OECD (2005). "Guidelines for Multinational Enterprises". *OECD, Paris*.
- Pierce, J. and P. Schott (2012). "A concordance between ten-digit US Harmonized System Codes and SIC/NAICS product classes and industries". *Journal of Economic and Social Measurement* 37(1), 61–96.
- Ramondo, N., V. Rappoport, and K. J. Ruhl (2016). "Intrafirm trade and vertical fragmentation in US multinational corporations". *Journal of International Economics* 98, 51–59.
- Rauch, J. E. (1999). "Networks versus markets in international trade". *Journal of International Economics* 48(1), 7–35.
- Rungi, A., G. Morrison, and F. Pammolli (2017). "Global Ownership and Corporate Control Networks". *EIC working paper series, IMT Institute for Advanced Studies Lucca*.
- UNCTAD (2009). "Training Manual on Statistics for FDI and the Operations of TNCs Vol. II.". *United Nations*.
- UNCTAD (2016). "World Investment Report 2016". *United Nations*.
- Williamson, O. E. (1971). "The vertical integration of production: market failure considerations". *The American Economic Review* 61(2), 112–123.
- Williamson, O. E. (1975). "Markets and hierarchies". *New York*, 26–30.
- Williamson, O. E. (1979). "Transaction-cost economics: the governance of contractual relations". *The journal of law & economics* 22(2), 233–261.
- World-Bank (2015). "World Development Indicators". *The World Bank*.

## A – Appendix: Corporate boundaries

Data were obtained from ORBIS and ZEPHYR, two large cross-country databases maintained by Bureau van Dijk. The ORBIS database includes information on financial accounts and ownership. ORBIS is integrated with ZEPHYR, which collects information on M&A deals. They source original information from several national registries and international providers. For our scope, ORBIS permits the identification of an *ultimate owner* and its links with affiliates worldwide. We specify that the *ultimate owner* should be a parent owning at least 50% of an affiliate. We keep companies in which parents have a minority stake as the first control group for robustness checks later in the analyses. We exclude from our sample all stand-alone firms, which do not have a parent company and may be owned by individuals. Eventually, our definition of corporate boundaries is the same as that acknowledged in international standards, see, for example, UNCTAD (2016), following the algorithm by Rungi et al. (2017), but also OECD (2005) and UNCTAD (2009) as manuals for statistics on multinational enterprises. See also Cravino and Levchenko (2016), Alviarez et al. (2016) and Altomonte and Rungi (2013), for a similar use of Orbis data.

In Figure A1, we briefly describe a fictional corporate boundary, which may be present in our sample.

Figure A1: Corporate boundaries



The parent on top directly controls two affiliates and, thus, indirectly controls three other affiliates. That is, in Figure A1, three firms can be considered as ‘affiliates of affiliates’, located on different hierarchical levels. Based on information on M&A deals from ZEPHYR and dates of incorporation from ORBIS, we can virtually separate new and past integration choices. In fact, in Figure A1, white boxes indicate affiliates that were already integrated before the new integrations occurred. We exclude from our sample parent companies that never integrated any affiliate in 2004–2012, for which we have information on metrics for determinants of integration and firm-level controls. The inclusion of the latter may induce a simultaneity bias in estimates with *ex post* regressors. We keep these as the second control group for later robustness checks.

For both parents and affiliates, we have information on the country and on the 6-digit primary industries (NAICS rev. 2007) in which it operates. To distinguish between domestic and foreign affiliates, we consider the

country of the parent company as the origin of the entire group. We further assume that the primary activity of the parent defines the main output of the group, whereas primary activities of the affiliates are its inputs. However, it is possible that either parents or affiliates may report more than one primary activity, which is consistent with the conglomerate nature of some companies. Throughout the analyses, we perform different robustness checks for the validity of assumptions based on the identification of industries. We exclude from our sample non-manufacturing parents (but we keep service affiliates), which may be not interested in the organization of GVCs (i.e., financial companies, utilities, consultancy firms, etc.). We keep them as the third control group for robustness checks later in the analyses.

Originally, we find 284,628 parent companies controlling more than 1 million and a half affiliates at the end of 2012. Among these, only 47,631 have primary manufacturing activities. In 2004–2012, there were 32,470 parent companies that reportedly integrated at least one further affiliate. Of these, only 4,192 are manufacturing parents, controlling 90,298 affiliates for which we have complete information on the countries and the industries in which they operate. They integrated 28,638 new affiliates in 2004–2012.

## B – Appendix Data

**Absolute and Relative Downstreamness.** The *Absolute Downstreamness* is sourced from Antràs and Chor (2013). They propose two alternative metrics based on 2002 U.S. I–O Tables. The first is the ratio of the aggregate direct use to the aggregate total use ( $DUseTUse$ ) of a particular industry  $i$ 's goods, where the direct use for a pair of industries is the value of goods from industry  $i$  directly used by firms in industry  $j$  to produce final goods, while the total use is the value of goods from industry  $i$  used, either directly or indirectly, in producing industry  $j$ 's output for final goods. Thus, a high value of  $DUseTUse$  suggests that most of the contribution of input  $i$  tends to occur relatively downstream, close to final demand. The second measure (*DownMeasure*) introduces a penalty, or a weight, for the average position at which an industry's output is used.

We also make use of the more recent *Relative Upstreamness*, sourced from Alfaro et al. (2015), which is a weighted-average of the number of stages it takes for input  $i$  to enter in the production of  $j$ , where the weights correspond to the share of total requirements' coefficients  $ij$  that enters at the corresponding upstream stage of production. For the sake of comparison across metrics, we take the inverse of *Relative Upstreamness*, such that a larger *Relative Downstreamness* means that a greater share of the total input use value of  $i$  is accrued closer in the production process for  $j$ .

**Demand Elasticity.** This is sourced from Broda and Weinstein (2006). We convert from HS10 codes, using tables provided by Pierce and Schott (2012). Following Antràs and Chor (2013), we identify industries that have elasticity above the median (5.37 for the 437 manufacturing sectors present in our data) as sequential complements and industries below the median as sequential substitutes. In a robustness check, we alternatively identify sequential complements ( $\rho - \alpha$ ) as being those cases in which the ratio between the elasticity faced by the parent company and the median value of the elasticity faced by the affiliates ( $\rho / \alpha$ ) is higher than unity.

**Capital intensity.** This is computed at the parent level from our sample, sourced from the Orbis database. It is a ratio between fixed assets and number of employees.

**Age.** This is computed at the parent level from our sample, sourced from the Orbis database. It is the age since incorporation of the company and until when choices of integration were undertaken.

**Size.** This is computed at the parent level from our sample, sourced from the Orbis database. It is the number of (consolidated) employees.

**Productivity.** This is computed at the parent level from our sample, sourced from the Orbis database. It is the ratio between value added and number of employees.

**Number of affiliates.** This is computed at the parent level from our sample, sourced from the Orbis database. It is the number of affiliates in which the parent has a (direct or indirect) majority stake.

**Multinational Group.** This is computed at the parent level from our sample, sourced from the Orbis database. It is a binary variable equal to 1 if the parent owns at least one foreign affiliate, i.e., in a country different from that of the parent.

**Contractibility.** This is computed at the industry level and then matched with the 6-digit industries of the affiliates. It is sourced from Antràs and Chor (2013), who computed the measure from the 2002 U.S. I–O Tables, following the methodology of Nunn (2007) and Rauch (1999).

**Entry cost.** This is sourced from the World Bank - Doing Business (World Bank, 2015). It is built on the number of procedures, number of days, and cost (as a percentage of income per capita) required to start a business in a country. We averaged over 2003–2005 for each affiliate  $i$  and host country  $c$ .

**Rule of law.** This is sourced from the Worldwide Governance Indicators (Kaufmann et al., 2011). The annual index is linearly rescaled from its original range  $[-2.5, 2.5]$ , to lie in a range  $[0, 1]$ , and averaged over 2004–2010 for each affiliate’s country of origin  $c$ . This index reflects perceptions in terms of the extent to which agents have confidence in and abide by the rules of society, especially in the quality of contract enforcement, property rights, the role of police and courts, and the likelihood of crime and violence.



## C – Appendix Tables

Table C1: Highest and lowest values of absolute downstreamness from sample

NAICS	Industry Label	Downstreamness
<i>Lowest 10 values</i>		
331314	Secondary Smelting and Alloying of Aluminum	0.0000
325110	Petrochemical Manufacturing	0.0599
331411	Primary Smelting and Refining of Copper	0.0741
336112	Light Truck and Utility Vehicle Manufacturing	0.0814
325211	Electric Housewares and Household Fan Man.	0.1205
325910	Printing Ink Manufacturing	0.1325
311119	Other Animal Food Manufacturing	0.1385
333220	Plastics and Rubber Industry Machinery Man.	0.1420
331311	Alumina Refining	0.1447
331312	Primary Aluminum Production	0.1447
<i>Highest 10 values</i>		
337122	Nonupholstered Wood Household Furniture Man.	0.9922
339116	Dental Laboratories	0.9942
332992	Small Arms Ammunition Manufacturing	0.9956
332993	Ammunition (except Small Arms) Man.	0.9956
316211	Rubber and Plastics Footwear Man.	0.9967
316212	House Slipper Manufacturing	0.9967
316213	Men's Footwear (except Athletic) Man.	0.9967
316214	Women's Footwear (except Athletic) Man.	0.9967
316219	Other Footwear Manufacturing	0.9967
336111	Automobile Manufacturing	0.9997

Table C2: Descriptive statistics of firm, industry, and country controls

	Obs.	Mean	Median	St. dev.	Min	Max
<b>Parents</b>						
(Log of) labor productivity	2,425	4.32	4.43	1.1	-3.82	11.72
(Log of) capital intensity	2,608	4.89	4.92	1.28	-3.18	12.02
(Log of) size	3,121	6.81	6.83	2.34	0	12.81
(Log of) Age	4,121	3.25	3.22	.9	0	5.9
(Log of) Number of affiliates	4,192	2.63	2.4	1.78	0	6.91
Multinational Group	4,192	.64	1	.48	0	1
<b>Global Parents only</b>						
(Log of) labor productivity	1,757	4.47	4.52	.99	-3.30	11.72
(Log of) capital intensity	1,857	5.09	5.06	1.13	-3.18	12.02
(Log of) size	2,094	7.55	7.65	2.15	0	12.81
(Log of) Age	2,675	3.28	3.18	1.75	0	6.91
(Log of) Number of affiliates	2,615	3.40	3.37	0.91	0	5.90
<b>Affiliates</b>						
Contractibility	46,627	.03	0	.12	0	1
Rule of Law	73,563	.73	.81	.16	.13	.89
Entry Cost	73,064	1.6	2.15	1.57	-6.91	7.63

Table C3: Correlation matrix of control variables

	(log of) labor productivity $j$	(log of) capital intensity $j$	(log of) employ ment $j$	(log of) number of affiliates $j$	(log of) age $j$	multinati onal group $j$	contracti bility $i$	rule of law $i$	entry cost $i$
(log of) labor productivity $j$	1.000								
(log of) capital intensity $j$	0.702	1.000							
(log of) employment $j$	-0.012	0.086	1.000						
(log of) number of affiliates $j$	0.186	0.289	0.717	1.000					
(log of) age $j$	0.053	0.015	0.280	0.241	1.000				
multinational group $j$	0.150	0.169	0.346	0.368	0.152	1.000			
contractibility $i$	0.045	0.089	-0.029	-0.001	0.004	-0.020	1.000		
rule of law $c$	0.094	0.031	-0.118	-0.053	-0.005	0.064	-0.047	1.000	
entry cost $c$	-0.025	0.013	0.115	0.070	0.045	0.011	0.003	-0.590	1.000

Table C4: Variants of sample composition for downstream parents

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
affiliate downstreamness $i(j)c$	parent country FE	output & input elast	top 5% affiliates	only manuf affiliates	different sector	top 100 inputs	only upstream to parent	only new affiliates	only direct control
parent downstreamness $j$	.207*** (.045)	.243*** (.052)	.347*** (.062)	.320*** (.059)	.037 (.050)	.222*** (.048)	.360*** (.042)	.139*** (.051)	.130** (.053)
Complements $j$	-.065* (.036)	-.010 (.043)	-.053 (.055)	-.030 (.048)	-.070* (.038)	-.100*** (.038)	-.223*** (.043)	-.058 (.039)	-.048 (.047)
complements $j$ *parent downstreamness $j$	.129** (.053)	.039 (.062)	.110 (.076)	.075 (.068)	.135** (.057)	.179*** (.057)	.332*** (.061)	.135** (.060)	.115* (.068)
(log of) labor productivity $j$	.011** (.006)	.010* (.005)	-.007 (.007)	.010 (.006)	.006 (.006)	.016** (.007)	-.003 (.005)	.005 (.006)	.014* (.008)
(log of) capital intensity $j$	.013* (.007)	.014* (.007)	.018** (.007)	.024*** (.007)	.013* (.007)	.009 (.007)	.001 (.007)	.004 (.007)	.006 (.008)
(log of) employment $j$	.011*** (.004)	.009** (.004)	.009* (.005)	.012*** (.004)	.007* (.004)	.007* (.004)	.008** (.004)	.009* (.005)	.006 (.005)
(log of) number of affiliates $j$	-.020*** (.005)	-.017*** (.005)	-.014** (.007)	-.020*** (.006)	-.015*** (.005)	-.019*** (.005)	-.013*** (.005)	-.019*** (.006)	-.015** (.006)
(log of) age $j$	-.008 (.005)	-.012** (.005)	-.009 (.007)	-.018*** (.006)	-.009* (.006)	-.011** (.005)	-.009* (.005)	-.010** (.005)	-.004 (.006)
multinational group $j$	.001 (.015)	-.006 (.014)	-.009 (.019)	.000 (.017)	.018 (.017)	.041** (.018)	.024 (.017)	.028* (.016)	.005 (.017)
contractibility $i$	-.043 (.041)	-.019 (.043)	-.011 (.041)	-.011 (.043)	-.131*** (.036)	.108 (.069)	-.033 (.031)	-.035 (.043)	-.041 (.070)
Constant	-.149 (.107)	.502*** (.039)	.250*** (.071)	.443*** (.045)	.601*** (.047)	.431*** (.076)	.471*** (.096)	.155** (.065)	.656*** (.068)
Observations	15,111	15,111	1,194	11,360	12,595	7,375	8,047	4,687	1,619
Adjusted R-squared	.130	.105	.257	.158	.051	.210	.245	.108	.120
Parent country f.e.	YES	NO	NO	NO	NO	NO	NO	NO	NO
Affiliate country f.e.	YES	YES	YES	YES	YES	YES	YES	YES	YES
BEC category of parent	Final goods	Final goods	Final goods	Final goods	Final goods	Final goods	Final goods	Final goods	Final goods

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* represent p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table C5: Variants of sample composition for midstream parents

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
affiliate downstreamness <sub><i>i(j)c</i></sub>	parent country FE	output & input elast	top 5% affiliates	only manuf affiliates	different sector	top 100 inputs	only upstream to parent	only new affiliates	only direct control
parent downstreamness <sub><i>j</i></sub>	.381*** (.029)	.346*** (.035)	.460*** (.052)	.428*** (.035)	.165*** (.032)	.480*** (.030)	.672*** (.042)	.346*** (.033)	.410*** (.051)
complements <sub><i>j</i></sub>	.032 (.022)	.001 (.027)	-.034 (.039)	.021 (.031)	-.019 (.027)	.041 (.027)	.072** (.035)	-.004 (.026)	.047 (.036)
complements <sub><i>j</i></sub> *parent downstreamness <sub><i>j</i></sub>	-.048 (.039)	-.008 (.048)	.043 (.063)	-.039 (.055)	-.008 (.049)	-.058 (.046)	-.118* (.062)	-.001 (.041)	-.086 (.060)
(log of) labor productivity <sub><i>j</i></sub>	.004 (.008)	.007 (.008)	.019** (.008)	.010 (.010)	-.002 (.009)	.008 (.009)	.007 (.009)	.008 (.005)	.003 (.008)
(log of) capital intensity <sub><i>j</i></sub>	-.011 (.008)	-.015** (.007)	-.018** (.008)	-.014 (.009)	-.011 (.008)	-.015* (.008)	-.017** (.008)	-.013** (.005)	-.013 (.010)
(log of) employment <sub><i>j</i></sub>	.008** (.003)	.008** (.004)	.006 (.005)	.011** (.005)	.006 (.004)	.005 (.005)	.006 (.005)	.007** (.003)	.009** (.004)
(log of) number of affiliates <sub><i>j</i></sub>	-.006* (.003)	-.005 (.004)	-.012** (.006)	-.006 (.005)	-.003 (.004)	-.004 (.006)	-.011** (.005)	-.005* (.003)	-.007 (.005)
(log of) age <sub><i>j</i></sub>	.001 (.005)	.004 (.006)	.005 (.008)	.006 (.007)	.012* (.007)	.003 (.008)	-.006 (.008)	.003 (.006)	.003 (.007)
multinational group <sub><i>j</i></sub>	-.054*** (.014)	-.060*** (.013)	-.015 (.019)	-.069*** (.015)	-.066*** (.017)	-.052*** (.015)	-.026 (.024)	-.029** (.013)	-.024 (.020)
contractibility <sub><i>i</i></sub>	-.310*** (.028)	-.308*** (.030)	-.184*** (.040)	-.252*** (.038)	-.314*** (.037)	-.328*** (.031)	-.110*** (.030)	-.322*** (.024)	-.253*** (.039)
Constant	.682*** (.068)	.380*** (.047)	.477*** (.061)	.336*** (.071)	.348*** (.073)	.310*** (.058)	.117* (.070)	.247*** (.050)	.615*** (.064)
Observations	19,482	19,482	1,443	15,934	15,575	11,189	6,850	5,389	1,901
Adjusted R-squared	.249	.224	.388	.240	.124	.331	.469	.265	.282
Parent country f.e.	YES	YES	YES	YES	YES	YES	YES	YES	YES
Affiliate country f.e.	YES	NO	NO	NO	NO	NO	NO	NO	NO
BEC category of parent	Inter	Inter	Inter	Inter	Inter	Inter	Inter	Inter	Inter
	mediates	mediates	mediates	mediates	mediates	mediates	mediates	mediates	mediates

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* represent p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table C6: Control groups

<b>Dependent variable:</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>
<b>affiliate downstreamness <math>i(j)c</math></b>	non-manuf parents	portfolio investment	no integration in 2004–2012
<b>parent downstreamness <math>j</math></b>	.286*** (.031)	.126*** (.027)	.468*** (.035)
<b>complements <math>j</math></b>	-.006 (.023)	-.015 (.020)	-.031 (.028)
<b>complements <math>j</math> *parent downstreamness <math>j</math></b>	.025 (.037)	.028 (.032)	.067 (.041)
<b>(log of) labor productivity <math>j</math></b>	-.003 (.003)	-.003 (.004)	-.004 (.004)
<b>(log of) capital intensity <math>j</math></b>	-.003 (.002)	.006 (.003)	-.005* (.002)
<b>(log of) employment <math>j</math></b>	.001 (.002)	.001 (.002)	.005** (.002)
<b>(log of) number of affiliates <math>j</math></b>	-.003* (.002)	.002 (.003)	-.006 (.004)
<b>(log of) age <math>j</math></b>	.006* (.003)	-.003 (.004)	-.004 (.003)
<b>multinational group <math>j</math></b>	-.017*** (.006)	.006 (.014)	-.001 (.007)
<b>contractibility <math>i</math></b>	-.220*** (.020)	-.259*** (.022)	-.158*** (.020)
<b>Constant</b>	.655*** (.028)	.871*** (.041)	-.328*** (.045)
<b>Observations</b>	72,628	20,991	110,104
<b>Adjusted R-squared</b>	.166	.087	.360
<b>Affiliate country f.e.</b>	YES	YES	YES

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* represent p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table C7: Relative downstreamness for midstream parents

<b>Dependent variable:</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>
<b>relative downstreamness of affiliate <math>i(j)c</math></b>	<b>OLS</b>	<b>OLS and firm controls</b>	<b>OLS, firm controls &amp; fe</b>
<b>complements <math>j</math></b>	-.048*** (.009)	-.071*** (.015)	-.079*** (.015)
<b>(log of) labor productivity <math>j</math></b>		-.009 (.012)	-.016 (.013)
<b>(log of) capital intensity <math>j</math></b>		.034*** (.012)	.034*** (.011)
<b>(log of) employment <math>j</math></b>		.006 (.005)	.003 (.005)
<b>(log of) number of affiliates <math>j</math></b>		-.007 (.006)	-.001 (.006)
<b>(log of) age <math>j</math></b>		-.006 (.006)	-.005 (.006)
<b>multinational group <math>j</math></b>		-.019 (.020)	-.049* (.027)
<b>contractibility <math>i</math></b>		-.176* (.102)	.064* (.033)
<b>rule of law <math>c</math></b>		-.109*** (.033)	
<b>contractibility*rule of law<math>ic</math></b>		.313** (.152)	
<b>entry cost <math>c</math></b>		.005* (.003)	
<b>Constant</b>	.509*** (.007)	.460*** (.054)	.356*** (.066)
<b>Observations</b>	18,348	6,700	9,259
<b>Adjusted R-squared</b>	.019	.060	.093
<b>Affiliate country f.e.</b>	NO	NO	YES
<b>BEC category of parent</b>	Intermediates	Intermediates	Intermediates

Errors clustered by parent in parentheses. \*\*\*, \*\*, \* represent p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table C8: Fixed effects conditional logit model. Integrated and non-integrated stages by midstream parents

Dependent variable:	(1)	(2)	(3)	(4)
<b>Input <math>m</math> integrated by parent <math>j</math> (Yes/No)</b>				
<b>relative downstreamness</b> $m_j$	.444*** (.019)	.421*** (.030)	.453*** (.018)	.429*** (.028)
<b>complements*relative downstreamness</b> $m_j$	.155*** (.029)	.156*** (.043)	.137*** (.026)	.132*** (.039)
<b>contractibility up-to-input</b> $m_j$		.017*** (.004)		.016*** (.004)
<b>contractibility up-to-input*complements</b> $m_j$		-0.001 (.005)		0.000 (.005)
<b>Dummy: Self-NAICS</b> $m_j$	.602*** (.018)	.671*** (.022)	.575*** (.016)	.646*** (.020)
<b>Observations (parent <i>per</i> stage)</b>	123692	116239	131436	123676
<b>Pseudo R-squared (McFadden's)</b>	.179	.183	.143	.147
<b>Parent-level f.e.</b>	YES	YES	YES	YES
<b>Activities integrated</b>	Primary	Primary	Primary and secondary	Primary and secondary
<b>BEC category of parent</b>	Intermediates	Intermediates	Intermediates	Intermediates