Production Staging: Measurement and Facts

Thibault Fally*
University of Colorado-Boulder

June 2012

Abstract

What is the average number of production stages for the US? Is production more vertically fragmented now than decades ago? Do upstream or downstream stages contribute to a larger share of value-added? To answer these questions, I develop two simple measures of vertical fragmentation of production chains across plants using input-output tables. Against common belief, I find that production has become less vertically fragmented over the past 50 years, whether I include services or focus only on tradable goods. An important part of this overall reduction in production staging reflects a shift of value-added towards industries that are closer to final demand, while upstream industries contribute less to the value of final goods. Looking at changes within industries, the production of more complex goods appears to be relatively less fragmented but these goods exhibit the largest increase in the number of production stages. Also, I show that international trade has provided new opportunities to fragment production and, to a small extent, dampened the overall decline in vertical fragmentation. Finally, I provide an alternative application of this index to the study of comparative advantage along production chains: I find evidence that goods that involve fewer production stages and goods that are closer to final demand are more likely to be imported to the US from rich countries.

Keywords: fragmentation of production, vertical linkages, vertical specialization.
JEL Classification: F10, F14, L16, L23, O14.

*I specially thank Russell Hillberry for very helpful suggestions. I also thank Pol Antràs, Jeff Bergstrand, Yongmin Chen, Arnaud Costinot, Gene Grossman, Gordon Hanson, David Hummels, Robert Johnson, Wolfgang Keller, Kalina Manova, Jim Markusen, Keith Maskus, Emanuel Ornelas, Scott Savage, Robert Staiger, John Stevens, Jonathan Vogel and seminar participants at Stanford University, Philly Fed Workshop, Bank of France, Graduate Institute, European Trade Study Group, London School of Economics and the University of Colorado-Boulder for helpful comments.
Contact: Department of Economics, University of Colorado-Boulder, 256 UCB Boulder, Colorado 80309-0256, USA. fally@colorado.edu, Ph: 303-492-5562.

1 Introduction

Recent work has documented the increasing complexity of production chains, with the examples of iPods, airplanes or cars. In particular, production has become more fragmented across countries (Hummels, Ishii and Yi, 2001, Johnson and Noguera, 2012), associated with a large growth in intermediate goods trade. Yet, little is known about the fragmentation of production across plants within countries. How long are production chains on average? Is production more fragmented now than it was decades ago? Production staging and the number of plants sequentially involved in production chains (henceforth referred to as vertical fragmentation) matter for several key issues in trade and other economic phenomena. As trade costs decline, gains from trade are magnified when production is, or can be, fragmented: not only can consumers import goods at a lower price, but producers can reduce costs by importing inputs at lower prices as well. Similarly, vertical linkages and the possibility of fragmented production constitute one of the main sources of gains from agglomeration according to Marshall.[1] Economic development has also put a traditional emphasis on the role of vertical linkages (Hirshman, 1955) and formalized more recently with the "O-ring" theory (Kremer 1993, Jones 2010).

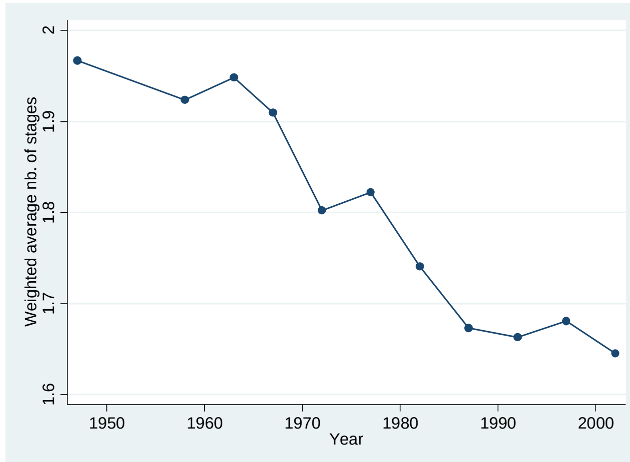
In this paper, I provide new quantitative analyses of the length of production chains, the evolution of production staging over time, and its determinants. I develop two simple measures to reflect: i) the number of production stages embodied in each product;[2] ii) the average number of stages between production and final consumption. These two different indexes provide complementary information on the position of each product along value chains. In particular, the first index corresponds to a weighted average of the number of plants sequentially involved in the production of a certain good, where the weight is the value that has been added at each stage. I show that these indexes have simple structural interpretations and are closely linked to traditional concepts of backward and forward linkages. I also examine aggregation properties of these two indexes and to what extent industry-level data can provide information on fragmentation across plants within and between industries. Moreover, in a closed economy, I find that a weighted average of these two indexes across all sectors equals the ratio of total gross output to value added, thereby offering a novel interpretation of this ratio.

I calculate these measures of vertical fragmentation for the US using benchmark input-output tables from the Bureau of Economic Analysis for periods covering 1947 to 2002 (aggregate sectors) and 1967 to 1992 (6-digit product level). I find that production chains are short on average and that most of the value added comes from later stages: the weighted number of

[1] Recently confirmed by Ellison et al., forthcoming.

[2] Here, stages correspond to plants. This definition may differ from a task-level approach where each stage could be associated with one task. It may also differ from a purely international perspective where "fragmentation" may refer to the fragmentation of production across countries (as in Johnson and Noguera, 2012).

Figure 1: Aggregate measure of vertical fragmentation (tradable goods excluding petroleum)



stages is smaller than 2 on average for the aggregate economy. Both indexes of fragmentation exhibit large variations across industries. In particular, I find that the number of embodied stages is negatively correlated with product specificity, R&D intensity, skill intensity and dependence on external finance, but does not seem to depend significantly on industry concentration (either proxied by the share of the largest firms in industry production or the Herfindahl Index).

The main and most surprising finding of the paper is that the weighted-average number of production stages has been decreasing by more than 10% over the past 50 years. While this decrease can be partly explained by the increasing share of services in total production, I find that the weighted-average number of production stages has also decreased for primary and manufacturing industries ("tradable" goods). Figure 1 plots this evolution, aggregating over all tradable goods excluding petroleum.

Since the main measure of fragmentation captures a value-added-weighted average number of production stages, changes in relative price of intermediate goods versus final goods may potentially explain changes in this measure (holding quantities fixed). Indeed, swings in oil prices may explain short-term changes in observed fragmentation by magnifying the weight put on early stages. Over the long term, however, I show that changes in relative prices of commodities and intermediate goods do not explain the overall observed decline.

I also specifically investigate the role of trade. The decrease in the overall fragmentation of production remains puzzling since it coincides with the reorganization of supply chains across borders.[3] We expect that the large decline in transport costs over the past decades has provided

[3]Note that the production staging index developed here accounts for both foreign and domestic sourcing. A pure substitution between foreign and domestic sources would not affect the index.

new opportunities. I find indeed that increased import penetration induced an increase in vertical fragmentation, suggesting that foreign outsourcing is not just a substitute to domestic outsourcing. This effect is small however, and not robust to instrumenting by transportation costs or tariff declines.

Perhaps the most intuitive way to understand the main finding of the paper is to look at the reallocation of value-added along production chains. I provide evidence of a large and significant shift over time of value-added towards industries that are closer to final demand (i.e. more downstream). In other words, early stages now contribute less to the final value of production, whereas more value is added at later stages. Industry characteristics can shed light on this shift of value-added. In particular, industries that are more intensive in advertising, in skilled labor and less intensive in capital have experienced a larger growth rate and are also relatively more downstream. Overall, such industry characteristics can explain about half of the shift of value-added towards downstream industries.

Furthermore, trade data suggest that this trend is global. I find evidence that the value of multi-lateral trade flows has grown faster in downstream industries relative to upstream industries (even if we omit trade to and from the US). This finding is similar to the shift of US value-added towards final stages and shows that this trend seems common to other countries.

This paper belongs both to the trade and industrial organization literatures. Since the possibility to fragment production affects trade patterns and the gains from trade (Grossman and Rossi-Hansberg, 2008), it is important to measure the extent of the fragmentation of production. Empirical evidence provides various examples of global supply chains (e.g. Feenstra, 1998) and document large trade flows in intermediate goods (Yeats, 2001, Campa and Goldberg, 1997). In comparison, my paper aims at capturing the fragmentation of production across plants instead of fragmentation across borders. There is of course a strong connection: when production can be fragmented within borders it is also more likely to be fragmented across borders. The decision to fragment production within borders remains however largely underexplored.[4]

In this paper, I also discuss and provide alternative uses of the two measures of fragmentation to examine trade patterns. I show that developed and developing countries tend to specialize at different stages along the value chain. In particular, my results suggest that richer countries such as the US have a comparative advantage: i) in goods that involve fewer production stages and ii) in goods that are closer to final demand. Previous indices on vertical specialization describe the use of imported inputs in exported goods or the value-added content in trade (e.g.

[4]A notable exception is Fort (2011) who examines the decision to fragment production (domestically and internationally) in a cross section of US plants in 2007. In all industries, she finds that most firms *do not* fragment their production, even domestically. This supports my results that production is not highly fragmented vertically. The data however do not allow her to examine the evolution of fragmentation over time.

Hummels, Ishii and Yi, 2001, Johnson and Noguera, 2010), but are not informative about the position of traded goods along the value chain and their sorting across countries.[5]

This paper also relates to several trends within the industrial organization literature. Firstly, it contributes to analyses of input-output tables pioneered by Leontief (1941). This literature has traditionally examined inter-industry "linkages" and the propagation of shocks across industries and regions. Instead, I show how input-output matrices can provide very interesting information on the number of plants involved sequentially in production chains and quantify the relative position industries along production chains.[6] To my knowledge, this is the first paper to document a decrease in a weighted-average number of production stages and a shift of value added towards downstream stages.

Secondly, it relates to an extensive amount of work in industrial organization on the make-or-buy decision and the determinants of vertical integration (see Lafontaine and Slade, 2007, for a survey of previous empirical works). Within this literature, many studies take as given the decision to source from a supplier, and focus on the ownership structure, i.e. on the decision to integrate this supplier or not. However, as documented by Syverson and Hortacsu (2011), most domestic shipments occur between two independent firms while plants within the same firm do not trade much among themselves. Hence it may be just as important to examine the decision to source inputs from within the same plant vs. from another plant, as reflected by the index developed here.[7]

The remainder of the paper contains four sections. Section 2 defines the key indexes and describes their properties. Section 3 describes the data. Section 4 presents descriptive statistics and the main empirical results. Section 5 presents several robustness checks and Section 6 concludes.

[5] Please note that my paper preceeds Antràs, Chor, Fally and Hillberry (2012) as well as Antràs and Chor (2012). Additional findings on the role of institutions and focusing on the second index (distance to final demand, also refered to as "upstreamness") are further described in Antràs, Chor, Fally and Hillberry (2012).

[6] The work on "average propagation length" (Dietzenbacher and Romero 2007, Bosma, Dietzenbacher and Romero 2005) also provides a step in this direction.

[7] Various indexes have been used to measure the extent of vertical integration such as firm size (e.g. Brynjolfsson, Malone, Gurbaxani and Kambil, 1994) or the ratio of value added to gross output (Adelman, 1955). The closest index related to this paper is the "Vertical Industry Connection Index" and similar indexes of vertical integration that take higher values when a firm owns a plant producing goods in an industry having strong make-buy relationship according to the input-output table (e.g., automobile manufacturing and steel) as in Maddigan (1981), Hitt (1999), Fang and Lan (2000), Acemoglu, Johson and Mitton (2007), Acemoglu, Aghion, Griffith and Zilibotti (forthcoming) among others. The later approach has several caveats however. The first is that it requires detailed plant-level data with sufficient information on the range of products that are produced. This makes it difficult to study the evolution of an entire economy over an extended period of time. A second caveat is that it is sensitive to the product classification, especially if inputs and outputs are classified in the same category making it impossible to distinguish integrated from disintegrated processes. Another caveat is that this index is based on ownership structure rather than actual shipments of intermediate goods (this index can take a high value even if these plants to not actually trade).

2 Definitions and properties

2.1 Embodied production stages: index N

In this section, I define two measures N_i and D_i defined by industry or product[8] to characterize the position along production chains. For each product i, I define:

i) N_i to reflect how many stages on average enter the production of i (average number of stages embodied in good i). This corresponds to a weighted-average number of plants involved sequentially in the production of i.

ii) D_i to measure how many plants on average this product will go through (e.g. by being assembled with other products) before reaching final demand. In other words, it captures the distance to final demand in terms of production stages.[9]

To construct N_i, I rely on information provided by input-output tables. In particular, we need data on the value of inputs from industry j used to produce one dollar of goods in industry i, which we denote by μ_{ij}. I define this index recursively: the average number of production stages embodied in a good depends on how many stages are embodied in each intermediate good. Using these μ's, I implicitly define N_i for each industry i by:

$$N_i = 1 + \sum_j \mu_{ij} N_j \tag{1}$$

This provides one equation for each industry. This system of linear equations generally has a unique solution that characterizes N_i.[10]

If production does not require any intermediate goods, the measure of fragmentation N equals one. If production relies on a particular intermediate good, the measure of production stages N depends on how important intermediate goods are in the production process and on how many production stages are needed to produce these intermediate goods.

Note that, in a special case where $N_j = N_i$ for all inputs j entering the production of good i, the index N_i would be equal to the gross-output-to-value-added ratio. This GO-VA ratio has previously been used as a measure of vertical fragmentation at the industry level.[11] In

[8] While the US input-output classification after 1967 is precise enough to name each category as a "product", I will henceforth refer to i indifferently as an industry or as a product. For convenience, time subscripts are dropped in this section and will be added in the empirical section.

[9] In Antras *et al.* (2012) we refer to this index as a measure of "upstreamness".

[10] As a corollary of the Perron-Frobenius theorems for non-negative matrices, this system has a unique solution if $\sum_j \mu_{ij} < 1$ for all i (this condition is always satisfied in practice). By inverting this system of equations, we obtain the (transposed) matrix of total requirements. This measure of production stages corresponds to the sum of "total requirement" coefficients for a given industry.

[11] See for instance Adelman (1995), Woodrow (1979), Macchiavello (2009).

general, though, N_i differs from the gross-output-to-value-added ratio and better accounts for inter-industry linkages when $N_j \neq N_i$ for a significant fraction of intermediate goods.

Another way to understand the intuition behind this index is to decompose output into slices of value-added. Let us denote by V_i the total value-added of industry i (gross output minus intermediate goods purchase). By construction, we have the following accounting equality: $\frac{V_i}{Y_i} + \sum_j \mu_{ij} = 1$. We can then see that a fraction $v_i^{(1)} = \frac{V_i}{Y_i}$ of the value of output has "gone through" only one stage since it has been added within the plant.

Then, looking more closely at intermediate goods, a fraction $v_i^{(2)} = \sum_j \mu_{ij} \frac{V_j}{Y_j}$ of output value comes from first-tier suppliers and has gone through 2 stages (including the value added by first-tier suppliers within the same industry i). Similarly, a fraction $v_i^{(3)} = \sum_{j,k} \mu_{ij}\mu_{jk} \frac{V_k}{Y_k}$ of the value has gone through 3 stages (i.e. that comes from suppliers of first-tier suppliers), and so forth. We can thus decompose each dollar of output i into different slices of value-added corresponding to different stages along the production chain:

$$1 \; = \; \frac{V_i}{Y_i} \; + \; \sum_j \mu_{ij} \frac{V_j}{Y_j} \; + \; \sum_{j,k} \mu_{ij}\mu_{jk} \frac{V_k}{Y_k} \; + \; ... = \sum_{n=1}^{\infty} v_i^{(n)}$$

where $v_i^{(n)}$ denotes the fraction of output value going through n stages. This fraction $v_i^{(n)}$ can be defined recursively by $v_i^{(n+1)} = \sum_j \mu_{ij} v_j^{(n)}$, with $v_i^{(1)} = \frac{V_i}{Y_i}$. Based on this decomposition, we obtain the following result:

Proposition 1 *If N_i is defined recursively as in equation (1) and $v_i^{(n)}$ is defined as above, then:*

$$N_i = \sum_{n=1}^{\infty} n \, v_i^{(n)}$$

In other words, N_i is the average number of stages to produce good i weighted by the share $v_i^{(n)}$ of value added at each stage n ($n = 1$ being most downstream).

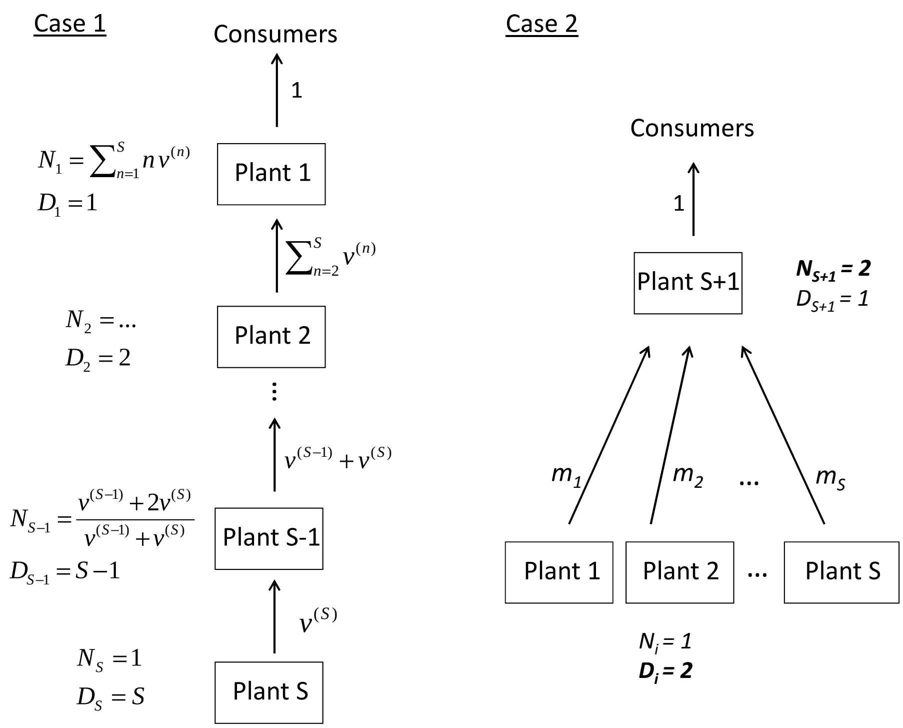
The proof is provided in the appendix section. Hence the index N_i can be reinterpreted as the average number of stages involved in the production chain, weighted by the value added at each stage.[12] Note that an input coming from a different plant (a supplier) counts as a different stage even if this input is classified in the same industry as the output.

To better grasp what N_i is measuring with respect to trade, firm ownership and the type of integration, several comments are in order:

[12] In Section 5.3, I examine an alternative index based on $v_i^{(n)}$ for each good i, inspired from the Herfindahl-Hirschman Index, to measure the dispersion of value added along the chain.

Snakes or spilders? While this measure aims at capturing the sequential nature of production, it obviously does not reflect all dimensions of complexity of production chains. In particular, it does not depend on the number of suppliers producing input j for industry i, as long as the share of inputs j in industry i's total costs remain constant. This point is illustrated in Figure 2, cases 1 and 2.

Figure 2: Vertical vs. horizontal fragmentation: an illustration



In the first case, each plant n contributes to a fraction $v^{(n)}$ of the final value of the product ($\sum_{n=1}^{S} v^{(n)} = 1$). According to Proposition 1, N equals $\sum_{n=1}^{S} n\, v^{(n)}$ for the final product and increases with the number of plants involved sequentially. In the second case, index N does not depend on the number of plants as long as they all contribute to a constant fraction $\sum_j m_j$ of the value of the final product ($\sum_j m_j = 1$ in the example above).[13]

Plants or firms? When the input-output table is constructed at the plant level (as is the BEA input-output matrix for the US), this index reflects the fragmentation of production across plants independently from the ownership structure (i.e. does not depend on whether suppliers are affiliated or not).[14] Note that, according to Hortacsu and Syverson (2011), shipments across

[13]Baldwin and Venables (2010) classify these two cases as "snakes" and "spiders"; my index only captures the length of snakes and is indifferent to the number of a spider's legs.

[14]A similar point has been made by Woodrow (1979) about the value-added-to-gross-output ratio: transactions are recorded in the input-output table even if it involves two plants owned by the same firm. It is however

plants belonging to the same firm account for only a very small fraction of total shipments. It suggests that similar results would be obtained if within-firm transactions were excluded.

Foreign or domestic sourcing? Index N_i does not depend on the share of imported inputs in intermediate goods purchases as long as products of the same classification requires the same number of production stages abroad as domestically.[15] Here I simply assume that production of input j is associated with the same measure N_j whether it is imported or produced domestically, taking the US as the benchmark.[16] Formally, if we differentiate input usage into domestic μ_{ij}^D vs. foreign purchases μ_{ij}^F, the sum of these two coefficients correspond to the observed input-output coefficient $\mu_{ij} = \mu_{ij}^D + \mu_{ij}^F$. Ideally, if we denote by N_i^D and N_i^F the weighted average number of production stages required to produce goods i from domestic and foreign sources respectively, we would like to define N_i^D by the following recursive equation:

$$N_i^D = 1 + \sum_j \mu_{ij}^D N_j^D + \sum_j \mu_{ij}^F N_j^F$$

Assuming that $N_j^F = N_j^D = N_j$, we obtain the same equality as in equation (1):

$$N_i = 1 + \sum_j (\mu_{ij}^D + \mu_{ij}^F) N_j = 1 + \sum_j \mu_{ij} N_j$$

This also means that this index does not differentiate between foreign sourcing (offshoring) and domestic sourcing, as long as both types of transactions occur across plants. If there is only a substitution between domestic and foreign sourcing, there is no effect of trade on index N_i. There is an effect only if foreign sourcing is a substitute to in-house (within-plant) production.

2.2 Distance to final demand: Index D

Whereas N_i reflects the number of stages *before* obtaining good i, an alternative measure D_i can be constructed to reflect the number of production stages between production of good i and final demand. For each product i, now we need to know the share of its production that is used as intermediate goods in industry j. We denote this coefficient by φ_{ij}. In a closed

difficult to track intra-firm transactions between plants.

[15]Input-output tables generally account for both imported and domestically produced inputs. The BEA tables incorporate the use of imports. However, these tables do not provide information on the share of imported inputs.

[16]This "mirror" assumption might generate a bias if imports are systematically correlated with the number of production stages. Results from Table 1 and Table 9 show that this is not the case: there is no significant correlation.

economy, this coefficient φ satisfies:

$$\varphi_{ij} = \frac{Y_j \mu_{ji}}{Y_i}$$

where Y_i stands for both the demand for good i and the supply of good i. In an open economy, part of the local demand is met by imports while a fraction of the local production is exported. Assuming that the share of production that is purchased by industry j is the same whether the good is internationally traded or not, then φ_{ij} should satisfy:

$$\varphi_{ij} = \frac{Y_j \mu_{ji}}{Y_i + M_i - X_i}$$

where Y_i stands for the value of production of good i, M_i for imports and X_i for exports. The denominator $Y_i + M_i - X_i$ is total demand (absorption) of good i in the country, and thus φ_{ij} is the fraction of this demand that corresponds to intermediate input demand from industry j.[17]

We can now use these coefficient φ_{ij} in the same way as for input-output coefficients μ_{ij}. For each product i, we define the "distance to final demand" D_i by:

$$D_i = 1 + \sum_j \varphi_{ij} D_j \tag{2}$$

Again, it defines one equation for each industry. This system of linear equations generally has a unique solution.

The intuition behind this index D mirrors the intuition for N. While N reflects the number of production stages embodied in production, D reflects the number of stages that have yet to be achieved before reaching final demand. In the extreme case where the entire production of this good is used as final consumption, this measure of distance to final demand is one. If part of the production is used as an intermediate good, this index is greater than 1 and depends on the share of production used as intermediate good and as well as the number of stages separating the corresponding downstream industry from final demand.

We should also note that it improves on a simple classification of parts versus final goods. As noted by Hummels *et al.* (2001), goods such as tires can be used as both intermediate goods and final goods. Index D_i does not suffer from this drawback since it more precisely account for the share of output being purchased by final consumers and producers.

A production chain such as case 1 of Figure 2 provides a simple example to grasp the meaning behind the index D. In this example, plants are index from 1 to S depending on their position on the chain (with with one being the closest to consumers). We obtain that $D_n = n$

[17] Note that this open-economy adjustment is consistent with situations where countries specialize at different stages of production. More details on open-economy adjustments are provided in Antràs, Chor, Fally and Hillberry (2012) where we further examine specialization patterns across a broad range of countries.

or each plant n. Note that, in this example, the measure of production stages for the last stage N_n equals the average of the distance to final demand D_i across all plants i weighted by the contribution of each plant to value added. This result is a corollary of Proposition 1 and also holds for the aggregate economy (Proposition 2 of section 2.4 below).

As we show in Antràs, Chor, Fally and Hillberry (2012), this intuition can be generalized to more complicated cases where the whole production in a particular stage is not necessarily sold to a unique plant. In particular, we can decompose output in a similar way as for Proposition 1 above. The above definition of D_i is equivalent to constructing a weighted average of the number of stages between an industry's output and final demand:

$$D_i = \sum_{n=1}^{\infty} n \, s_i^{(n)}$$

weighted by the share of output $s_i^{(n)}$ of industry i that goes through n stages before reaching final demand. In particular, $s_i^{(1)}$ corresponds to $1 - \sum_j \varphi_{ij}$, the fraction of output of industry i that goes to final demand. $s_i^{(2)}$ corresponds to the $\sum_j \varphi_{jk}(1 - \sum_j \varphi_{kj})$, i.e. the fraction of output of industry i that is purchased as inputs for the production of goods that are then sold to final consumers, etc. The fraction $s_i^{(n)}$ can be formally defined by $s_i^{(n+1)} = \sum_j \varphi_{ij} s_j^{(n)}$, with $s_i^{(1)} = 1 - \sum_j \varphi_{ij}$.

2.3 Structural interpretations

While the index N_i has an intuitive interpretation, the link with previous models and more structural interpretations is not straightforward and depends on the structure of production. This section motivates this index from a more structural standpoint, linking N_i to: i) cumulative trade costs along production chains; ii) the elasticity of prices to productivity; iii) the elasticity of output to productivity; iv) the gains from trade in a Ricardian framework.

i) Cumulative transport costs: As shown by Yi (2010), vertical specialization and multiple border crossings along production chains magnifies the effect of transport costs on trade. A similar effect can apply to domestic trade between plants.

To illustrate the role of fragmentation and the relevance of index N_i, let us examine cumulative trade costs T_i that are being paid for the production of output Y_i. If we assume that an iceberg transport τY is being paid for each shipment Y between two plants (for this illustration I assume that τ is constant across industries), then the cumulative transports costs paid for output i can be defined by the following equality:

$$T_i \,=\, \tau Y_i \,+\, \sum_j Y_i \mu_{ij} \, \frac{T_j}{Y_j}$$

where τY_i reflects transport costs for the shipment of output Y_i, where $Y_i \mu_{ij}$ is the amount of intermediate goods j that enters the production of i, and where $\frac{T_j}{Y_j}$ is the cumulative transport cost for the production of one dollar of input j. From this equality, we can easily verifies that $\frac{T_j}{\tau Y_j}$ satisfies the same recursive definition as N_i (equation 1). Total transport costs embodied in one dollar of output i are thus equal to $\tau \, N_i$. This implies, for instance, that a 1% increase in transport costs τ would have a larger effect on high-N industries.

ii) Price multiplier: Now, let us consider an economy with J goods, characterized by the following production functions:

$$Q_i = ZF_i(Q_{i1}, Q_{i2}, ..., Q_{iJ}, L_i)$$

where Z is a economy-wide productivity term, and F_i is a good-specific production function with constant returns to scale, Q_{ij} the quantity of good j used in the production of good i, and L_i the amount of labor used for i. In this general setting, after normalizing wages to unity, we obtain that the elasticity of prices to economy-wide productivity shocks corresponds to the fragmentation index N_i:

$$\frac{\partial \log P_i}{\partial \log Z} = -N_i$$

(see proof in appendix). In the spirit of the O-ring theory (Kremer, 1993) and a more recent model of Costinot, Vogel and Wang (2012), we could further assume that mistakes are made at each stage of production and that mistakes destroy both production and inputs used in production, so that productivity is determined as $Z = \mathrm{e}^{-\lambda}$ where λ is the Poisson rate of arrival of mistakes. In this setting, the semi-elasticity of prices to the rate of mistakes λ equals $-N_i$. The intuition is straightforward: the larger number of production stages, the larger the effect of mistakes on production. While this setup is slightly different from Costinot et al (2012) who do not explicitly model vertical linkages between different industries, this result shows that the number of production stages (as measured by index N_i) plays a key role in understanding production chains. As shown in Costinot et al (2012), we can thus expect more productive countries to have a comparative advantage in high-N industries.

We can further examine how a change in productivity affects welfare. In this general framework, we obtain that the effect of productivity on welfare depends on the average of index N_i

weighted by the share of each good in final consumption:

$$\frac{\partial \log e}{\partial \log Z} = -\frac{\sum_i C_i N_i}{\sum_i C_i} \tag{3}$$

where e denotes the expenditure function for a given level of utility.[18]

iii) Output multiplier: While the role of index N as a multiplier for prices holds in a general framework, the link between productivity and output depends on the structure of the economy and the shape of production functions. In the appendix, I illustrate the role of N_i and D_i in two cases: with Cobb-Douglas and with Leontief production functions.

In a first simple case where production functions and preferences are Cobb-Douglas functions of goods i, the elasticity of output in industry i to economy-wide productivity shocks Z correspond to the index N_i:

$$\frac{\partial \log Q_i}{\partial \log Z} = N_i$$

This simple case formalizes the link with "total output multipliers" that are well-known in the input-output literature (Chenery and Watanabe 1958, Rasmusen 1956).

In a second case where production functions and preferences are Leontief functions of goods i, the elasticity of output in industry i to economy-wide productivity shocks Z now depends on the index D_i:

$$\frac{\partial \log Q_i}{\partial \log Z} - \frac{\partial \log Q_j}{\partial \log Z} = D_i - D_j$$

In other words, a change in productivity (or in the rate of mistakes as in Costinot et al 2012), the effect on output is the largest for the more upstream goods, i.e. goods that are the "furthest" from final demand. In both cases, we can see that the position of an industry on the production chain determines the sensitivity of output to productivity shocks.

iv) Welfare gains multiplier: As motivated in the introduction, the fragmentation of production magnifies the gains from trade and economic integration. This intuition can be formalized by taking the same approach as in Arkolakis, Costinot and Rodriguez-Clare (2010).

For simplicity, let us assume that we have several industries i and that production in each industry is as in Eaton and Kortum (2002): markets are perfectly competitive, productivity draws for each variety follow a Frechet distribution, labor is the only factor of production, trade flows satisfy a gravity equation, and demand is CES (see Arkolakis et al 2010, for more details on the underlying assumptions of the competitive case). If there is only one production stage, and if the wage at home is normalized to unity, Arkolakis et al (2010) show that the change in

[18]Since wages are normalized to unity, a decrease in e reflects a increase in welfare.

the price index is given by:

$$\hat{P}_i = \frac{\hat{\lambda}_i^{dom}}{\theta_i}$$

where λ_i^{dom} refers to the fraction of goods that are *not* imported (in the consumption of goods in industry i) and where θ_i is both the coefficient of dispersion of the Frechet distribution of production in industry i and the elasticity of trade to trade costs in this industry.

If we extend their model by allowing for inter-industry linkages, assuming Cobb-Douglas production functions with coefficients μ_{ij} for the share of input j in the production of good i, the expression above becomes:

$$\hat{P}_i = \frac{\hat{\lambda}_i^{dom}}{\theta_i} + \sum_j \mu_{ij} \hat{P}_j$$

If we further assume that the change in import penetration is the same in all industries ($\hat{\lambda}_i^{dom} = \hat{\lambda}^{dom}$) and that $\theta_i = \theta$ is also constant across industries, then we obtain that the change in the price index P_i is proportional to the average number of production stages as measured by N_i:

$$\hat{P}_i = \frac{\hat{\lambda}^{dom}}{\theta} \,.\, N_i$$

The intuition is simple. When a country opens to trade, not only consumers can have access to cheaper foreign goods but domestic producers can also reduce their costs by importing cheaper intermediate goods. This magnifies the gains from trade, especially in industries with multiple production stages.

2.4 Index for the aggregate economy

Before turning to the data and computing these indices, I show that these two indices satisfy two key aggregation properties. First, the weighted average of these two indices equal the ratio of gross output to value added for the aggregate economy. Then, I investigate under which conditions the computation of these two indices does not generate measurement errors when available data are partially aggregated.

While both measures N_i and D_i are defined for each industry, we need to characterize the aggregate economy. For aggregation purposes, the key is to consider the appropriate weights to compute averages. With these two indices in hand, we can compute:

i) The number of production stages embodied in final goods (using index N_i), averaged across all goods purchased by final consumers. For this purpose, a natural weight is the total value of good i used for final consumption. As shown in equation (3), this would be also a natural weight to examine welfare implications.

ii) The average number of stages between production and final consumption (distance to final demand), making use of index D_i. For this purpose, a natural weight is the value added by industry i.

I denote by C_i the value of final consumption of good i. It satisfies: $C_i = Y_i - \sum_j \mu_{ji} Y_j + M_i - X_i$. It corresponds to total production minus the amount used as intermediate goods by domestic plants, plus net imports. Similarly, I denote by V_i the value added by industry i, which equals production of good i minus the total use of intermediate goods for the production of good i: $V_i = (1 - \sum_j \mu_{ij}) Y_i$.

Closed economy

In a closed economy, net imports equal zero and $C_i = Y_i - \sum_j \mu_{ji} Y_j$. Using accounting equalities and the definition of the index (see proof in the appendix), it turns out that the weighted average of both measures of fragmentation equal the ratio of gross output to value added:

Proposition 2 *For a closed economy, the average of the number of production stages N_i across all industries weighted by their contribution to final demand C_i equals the average distance to final demand D_i weighted by value added V_i, and both equal the ratio of total gross output over GDP:*

$$\frac{\sum_i C_i N_i}{\sum_i C_i} = \frac{\sum_i V_i D_i}{\sum_i V_i} = \frac{\sum_i Y_i}{\sum_i V_i}$$

This result provides an interesting interpretation of the gross-output-to-value-added ratio in an economy: it equals the average number of production stages and reflects the fragmentation of production in the economy (note that this is not the case at the industry level).

Open economy

In an open economy, there is no longer equality between supply and demand for intermediate goods by domestic industries (net imports $M_i - X_i$ no longer equal zero). In this case, the weighted average of the number of production stages is no longer equal to the ratio of gross output to GDP, and no longer equal to the average distance to final demand weighted by value added. Interestingly, the differences between each index and the GO/VA ratio can be expressed as a correlation term between net imports and each index across products:

Proposition 3 *For the aggregate economy, the average of the number of production stages N_i across all products i weighted by final consumption C_i and the average number of stages between* *production and final demand D_i weighted by value added V_i satisfy:*

$$\frac{\sum_i C_i N_i}{\sum_i C_i} = \bar{N} + \frac{\sum_i (M_i - X_i)(N_i - \bar{N})}{\sum_i C_i} \quad (4)$$

$$\frac{\sum_i V_i D_i}{\sum_i V_i} = \bar{N} - \frac{\sum_i (M_i - X_i)(D_i - 1)}{\sum_i V_i} \quad (5)$$

where $\bar{N}$ denotes the gross-output-to-value-added ratio.

When net trade $(M_i - X_i)$ is not correlated with either fragmentation index N_i or D_i, then the equality to the gross-output-to-value-added ratio continues to hold even in an open economy. When net imports are positively correlated to the number of production stages N_i, the gross-output-to-value-added ratio underestimates the weighted average number of production stages as it does not account for the larger number of production stages embodied in imports. Conversely, the gross-output-to-value-added ratio underestimates the average number of stages to final demand when a country tends to export goods that are relatively further from final demand.

Cross-border production sharing and the VAX ratio

In this analysis, the measure of fragmentation captures the number of plants involved sequentially in production whether these plants are within the same country or whether production is organized across countries. Johnson and Noguera (2012) instead define fragmentation as cross-border production sharing. Their measure of fragmentation for the aggregate world economy is the ratio of total value-added content of exports to the total gross value of exports ("VAX_world"). In keeping with Johnson and Noguera's notation, this is:

$$VAX_{world} = \frac{\sum_{i \neq j} \sum_s va_{ij}(s)}{\sum_{i \neq j} \sum_s x_{ij}(s)}$$

where $x_{ij}(s)$ denotes bilateral gross trade flows between countries i and j in sector s and where $va_{ij}(s)$ the value-added content of trade between i and j.

As one could expect, there is a close link between these two measures of fragmentation, the VAX ratio and the gross-output-to-value-added ratio. In particular, Propositions 1 and 2 can shed light on the interpretation of the VAX ratio. To see the correspondence, one could consider each country as one plant. In line with this interpretation, the equivalent measure of gross output would be total exports for the world in Johnson and Noguera's case (within-country transactions are not counted in the measure of total exports, such as within-plant transactions in the measure of gross output) and the equivalent measure of value added would be total value-added content of trade in in Johnson and Noguera's case.[19] Using Propositions 1 and 2, we can conclude that the inverse of the VAX ratio corresponds to the number of embedded border crossings in each dollar of the final good, weighted by the contribution of each country to total value-added content of trade (details are provided in the appendix section). Formally:

$$\frac{1}{VAX_{world}} = \frac{\sum_n \sum_{i \neq j} n.va_{ij}^{(n)}}{\sum_n \sum_{i \neq j} va_{ij}^{(n)}}$$

where $va_{ij}^{(n)}$ denotes the part of the value added by country i that is going to cross n borders before reaching final demand in country j. Hence the *inverse* of the VAX ratio is the analogous of the gross-output to value added ratio when focusing on cross-border transactions instead of transactions between plants.

2.5 From varieties to industries

For the calculation of N_i and D_i, the unit of observation would be ideally the plant or the product variety. Unfortunately, calculating this index at the plant or variety-level would require plant-level input-output matrices (with data on transactions matched between buyers and suppliers) that are not available.

In this section I derive conditions under which the index measured at the industry level (equation 1) equals the average of an ideal index at the plant level weighted by the value of production by each plant that is sold to final consumers. If production techniques are homogenous across plants within each industry, this question would be irrelevant. However, Fort (2011) documents substantial heterogeneity within each industry in terms of fragmentation of production and sourcing strategies.

Some additional notation is needed for this subsection only. Let us assume that each industry i is composed of a set of varieties $\omega \in \Omega_i$. These sets Ω_i offer a partition of the set of all varieties produced in the economy. If we denote by $y(\omega)$ the value of production of variety ω, gross output Y_i of industry i can be defined as $Y_i = \int_{\Omega_i} y(\omega)d\omega$.

Without loss of generality, I assume that each variety is either sold to final consumers or sold to a unique downstream industry j.[20] I denote by Ω_{ij} the set of varieties in industry i that are sold as intermediate goods to industry j, and I denote by Ω_{iF} the set of varieties in industry

[19] The part of value-added that corresponds to final consumption within the same country is not counted in the total value-added content of trade.

[20] While in practice the same type of product (e.g. tires) can be sold as an intermediate good to a downstream industry (e.g. the auto industry) and as a final good to consumers, for accounting purposes we can simply consider these products as different varieties that require the same production process (e.g. tires sold to final consumers vs. other tires).

i that are sold as final goods. For a given industry i, the sets Ω_{ij} and Ω_{iF} offer a partition of Ω_i. In particular, Ω_{ii} refers to the set of varieties of industry i that are used as intermediate goods by industry i (e.g. chemicals used as inputs for other chemicals).

Now let us assume that $N(\omega)$ is the "true" index of production stages at the variety level which could be measured if we had plant-level input-output matrices, i.e. data on the full supply chain for each variety ω. Under the following conditions, the industry-level index equals a weighted average of the variety-level index in each industry:

Proposition 4 *If* $(\int_{\Omega_{ij}} y(\omega)N(\omega)d\omega)/(\int_{\Omega_{ij}} y(\omega)d\omega)$ *does not depend on the downstream industry* j*, for all* $j \neq i$ *or* $j = F$*, then:*

$$N_i = \frac{\int_{\Omega_{iF}} y(\omega)N(\omega)d\omega}{\int_{\Omega_{iF}} y(\omega)d\omega}$$

is the solution to equation (1) which characterizes index N_i *at the industry level.*

In other words, the industry-level index defined by equation (1) provides an unbiased measure of the average of the "true" index at the variety level (weighted by final consumption) provided that the number of production stages does not depend on the buying industry j. Formally, it requires that:

$$\frac{\int_{\Omega_{ij}} y(\omega)N(\omega)d\omega}{\int_{\Omega_{ij}} y(\omega)d\omega} = N_i$$

whatever the downstream industry $j \neq i$. While plants may be heterogeneous in terms of production processes, such heterogeneity matters in terms of aggregation only if there is a systematic link between supply and demand across industries. For instance, if more productive firms are more likely to fragment their production, this would affect the measure of the industry-level index only if those firms are more likely to sell goods to a particular downstream industry rather than another.

Note also that these conditions do not impose any constraint on within-industry linkages and we may have:

$$\frac{\int_{\Omega_{ii}} y(\omega)N(\omega)d\omega}{\int_{\Omega_{ii}} y(\omega)d\omega} \neq N_i$$

In particular, if all varieties are aggregated into a unique industry (representing the whole economy), the measured index of production stages for an aggregate closed economy (GO/VA) is unbiased and equals the average of the index across all varieties that are sold to final consumers.

In order to mitigate the aggregation bias, more aggregation might be an answer instead of an issue. Indeed, if fragmentation depends on the buying industry, aggregating industries into larger industries might actually eliminate such patterns. For instance, if the production of auto parts is more or less fragmented depending on whether buyers are final consumers or plants in the auto industry, then aggregating auto parts with the rest of the auto industry would eliminate the bias that arises between the observed index of production stages and the true average across varieties of the number of production stages.

In Section 5.2, I show that the measure at a more aggregated level does not differ from the weighted average of the index measured at a more disaggregated level. I show that aggregation yields very little bias in the construction of the fragmentation index when I use an artificially aggregated input-output matrix (i.e. after aggregating the US input-output matrix at the 2-digit instead of 6-digit level). The new measure is very close to the weighted average of the most precise one ($< 1\%$ error on average). This suggests that the measure of the number of production stages using equation (1) is robust to using aggregated data.

Similar properties can be derived for the distance to final demand D_i. Let $v(\omega)$ denote the value added in the production of variety ω and $\mu_j(\omega)$ denote the use of inputs from industry j in the production of variety ω. We obtain the following conditions for unbiased aggregation:

Proposition 5 *If:* $(\int_{\Omega_i} y(\omega)\mu_j(\omega)D(\omega)d\omega)/(\int_{\Omega_i} y(\omega)\mu_j(\omega)d\omega) = (\int_{\Omega_i} v(\omega)D(\omega)d\omega)/(\int_{\Omega_i} v(\omega)d\omega)$ *for all downstream industries* $j \neq i$*, then:*

$$D_i = \frac{\int_{\Omega_i} v(\omega)D(\omega)d\omega}{\int_{\Omega_i} v(\omega)d\omega}$$

is the solution to equation (2) which defines index D_i *at the industry level.*

In other words, the measure of the number of stages to final demand is unbiased at the industry level if there are no systematic differences in the distance to final demand depending on the use of inputs.

In the appendix section, I investigate additional aggregation properties. I examine how the aggregation of two sub-industries into one affects the aggregate measure for these two sub-industries and other industries in the economy.

3 Data

The main data sources are the US input-output matrices developed by the Bureau of Economic Analysis (see Horowitz and Planting, 2009, for a description of the methodology). The US input-output matrices are unique in the world: they cover the longest time span (since 1947) and are available at a very detailed level (6-digit classification since 1967). Input-output tables for other countries are generally not available at such disaggregated level or only for a much shorter time span.[21]

I use the BEA input-output tables for benchmark years, which are available online.[22] Unfortunately, industry classifications are not always homogenous across periods. The 1997 and 2002 IO tables follow the NAICS classification (430 product categories); the 1967, 72, 77, 82, 87 and 92 IO tables are based on the SIC classification (6-digit level, up to 540 product categories); the 1963 table also follows the SIC classification but is defined at the 4-digit level; previous tables (1947 and 1958) are aggregated to 85 industries.

When I construct the vertical fragmentation index for the aggregate economy I can thus cover 55 years. When more disaggregated data are required for cross-industry comparisons, I rather focus on the period 1967 to 1992 which provides a panel of 377 harmonized product categories.[23] No very precise concordance table is available for NAICS to SIC and so I do not consider the 1997 and 2002 IO tables in my regressions by industry.[24]

Note that the industry classification is more precise for manufacturing goods and commodities, with 305 disaggregated industries in the manufacturing sector. Some services sectors (such as retail and wholesale trade) are not described at a detailed level. Also, I complete these data by a set of various covariates that are used throughout Section 4. The source and construction of these variables are described in the appendix section. Given the greater availability of data for manufacturing industries, regressions performed at the industry level mostly focus on the manufacturing sector. The manufacturing sector is composed of 305 consolidated input-output industries, 266 of which having information on all variables.

Several remarks are in order about the construction of these data. First, the US input-output matrices are based on data on establishments, or plants. As defined by the Census Bureau, an establishment is "a business or industrial unit at a single physical location that produces or distributes goods or that performs services." (Horowitz and Planting, 2009, p39). Hence, each input-output matrix should reflect transactions between plants even if these plants are classified in the same industry. In the construction of indexes N_i and D_i, these within-industry transactions do matter in order to measure the degree of vertical linkages not just across industries but also across plants within industries. Specifically, these within-industry

[21] This is particularly the case for input-output tables that have been homogenized across several countries, e.g. OECD IO Tables (constructed for 40 industries since 1992), IDE-JETRO IO Tables and GTAP IO Tables (about 80 industries). Among specific countries, Denmark probably has the best coverage (about 200 industries since 1966), which still does not compare to the coverage provided by US IO tables.

[22] http://www.bea.gov/industry/io_benchmark.htm

[23] Some sectors are more disaggregated for certain years but I consolidate these industry classifications to obtain a homogenous classification across all years. The final one is close to 1987 SIC.

[24] See Pierce and Schott (2009) for a discussion. My attempts to include these two years generally confirm my results for 1967-1992.

transactions are reflected in the diagonal terms μ_{ii} in the IO matrix. Finally, we should note that, given the level of disaggregation of the US tables, these diagonal terms are not large: only 10% of intermediate goods purchases are recorded from within the same industry (between 9.8% and 10.9% each year). This fraction is typically much larger in other input-output tables (e.g. OECD tables) where product classifications are much more aggregated.

4 Empirical Findings

4.1 Descriptive statistics

Evolution of production staging, 1947-2002

The first striking fact is that the weighted-average number of production stages for the US is below 2 except for 1947 and 1958. This is shown in Figure 3 with the average index of production staging proxied by the gross-output-to-value-added ratio for all products. Production is not as disintegrated as we could expect. In other words, value added embodied in production goes through less than two plants (two stages) on average before reaching final demand.

Figure 3: Weighted average number of production stages

Moreover, the fragmentation of production in the US has been decreasing over time. This decrease in the fragmentation of production has been quite smooth over time except for years 1977 and 1982. An obvious candidate explanation for the peak in 1977 and 1982 is the increase in oil prices.[25] When I thus reconstruct my index by excluding petroleum-related industries (crude petroleum and refining), the 1977 and 1982 peak almost disappears and the overall decline in the fragmentation of production is confirmed.[26]

[25]See sections 4.3 and 4.4 for more precise analyses of the role of prices.

[26]The negative trend is statistically significant even after correcting for auto-correlation.

One simple potential explanation is the increasing role played by services in the US economy. Services now account for more than two thirds of GDP but generally require fewer production stages. Moreover, we need to carefully interpret the fragmentation measure using services as the input-output matrix is much more aggregated for these sectors.[27] In comparison, data on manufacturing sectors are more finely detailed.

In the right panel of Figure 3, I compute the aggregate index of fragmentation using only tradable goods and tradable inputs (manufacturing goods and commodities, excluding services and petroleum-related industries). Even if we exclude services, the downward trend is confirmed. The average number of embodied stages for tradable goods declined from 2 to 1.6 over the past 50 years. We can further restrict our attention to manufacturing industries but the picture remains similar.

The figures above are based on the gross-output-to-value-added ratio, adjusting value added for the use of excluded industries such as petroleum. This amounts at considering the US as closed economy. In an open economy, aggregate measures of fragmentation may differ, as shown in Proposition 3. In particular, the aggregate number production stages $\frac{\sum_i N_i C_i}{\sum_i C_i}$ (weighted by final consumption) can differ from the aggregate number of stages to final demand $\frac{\sum_i D_i V_i}{\sum_i V_i}$ (weighted by value added). Using industry-level trade data from 1967 to 1992, I compute the difference between each of these aggregated indexes and the GO/VA ratio, as described in Proposition 3.

Table 1: Aggregation biases in open economy

Year	Import Penetration	GO/VA Ratio	Δ Number of stages	Δ Distance to final demand
1967	0.033	1.937	0.006	-0.002
1972	0.064	1.805	0.012	-0.005
1977	0.073	1.814	0.013	0.011
1982	0.094	1.728	0.020	0.023
1987	0.140	1.665	0.011	-0.021
1992	0.157	1.658	0.012	0.005

Notes: GO/VA is the ratio of gross output to value added calculated for the aggregate economy. The terms ΔN and ΔD corresponds to the difference between each aggregate index and the GO/VA ratio.

Results are shown in Table 1. While trade has grown very rapidly during this period (import penetration rose from 3.3% in 1967 to 15.7% in 1992), not adjusting for trade creates very little bias in the computation of the aggregate measure of fragmentation. Deviations are smaller

[27] For instance, wholesale trade and retail correspond to only two industries in the input-output table.

than 0.02, i.e. a 1% error at most. Figure 3 would thus remain identical after correcting the fragmentation index for international trade. Basically, deviations reflect the correlation between fragmentation measure and net imports. The small magnitude of these deviation terms is surprising given that we would expect trade patterns to be systematically related to the position on the value chain. This issue is further discussed in Section 4.3.[28]

Indexes of production staging by industry in 1992

I now turn to cross-indutry variations in the production staging indexes and describe industries with the largest values of embodied production stages N_i. I find that food industries typically involve long production chains withi little value added at each stage (see Table 2a). Among the top-5 industries with the largest values for N_i, we find meat packing, sausages, cheese and butter industries (poultry is next). Among the top 25 industries, 17 are related to food. Non-food industries in the top 25 are metal-intensive industries (e.g. cans), leather tanning, petroleum refining, video and audio equipment, wood preserving and the car industry. If we only look at tradable intermediate goods (manufacturing goods and commodities, excluding services and petroleum-related industries), the ranking among top industries is almost the same. In line with case studies (e.g. Helper, 1991), the car industry appears to be quite disintegrated, though not as disintegrated as the food industry. The weighted average number of stages is 2.8, and it is 2.4 for auto parts.

Note that the fragmentation index N_i differs from the gross-output-to-value-added ratio Y_i/V_i at the industry level. Fragmented industries generally exhibit a large GO/VA ratio but the difference between the two indices can also be large (first vs. last column) and the ranking is not preserved.

In turn, if we look at the index on the number of stages between production and final demand (distance to final demand), primary goods exhibit the largest values. The largest is obtained for basic metal products (Table 2b).

Conversely, industries with the smallest number of production stages are generally service industries (see Table 3). If we only consider tradable goods, industries with the smallest number of production stages correspond to primary goods. Similarly, industries that are closest to final demand are generally services industries. In 1992, 8 products are not used as intermediate goods: "Residential care", "Hospitals", "Cigarettes", "House slippers", "Doctors and dentists", "Owner-occupied dwellings", "Child day care services", "Ordnance and accessories, n.e.c".

These figures show that no industry exhibit extreme values for both indexes N_i and D_i

[28]In Section 4.3 we confirm that import penetration is not significantly correlated with the number of production stages. We find, however, that fragmentation has increased relatively more in sector with larger import penetration.

Table 2: Industries with the largest index values

Table 2a: **Index N_i (embodied production stages)**

Production stages	All inputs	Tradables	GO/VA
Top-5 industries:			
Meat packing plants	3.49	2.67	8.74
Sausages and other prepared meat products	3.39	2.65	4.88
Leather tanning and finishing	3.15	2.43	3.93
Natural, processed, and imitation cheese	3.15	2.35	5.55
Creamery butter	3.13	2.35	5.12
Motor vehicle industries:			
Motor vehicles and passenger car bodies	2.79	2.03	6.09
Motor vehicle parts and accessories	2.40	1.78	3.15
Truck and bus bodies	2.41	1.82	2.83
Truck trailers	2.59	1.91	3.75

Table 2b: **Index D_i (distance to final demand)**

Stages to final demand	All inputs	Tradables
Nonferrous metal ores, except copper	7.17	6.48
Copper ore	5.10	4.37
Oil and gas field machinery and equipment	4.45	3.22
Primary smelting and refining of copper	4.39	3.65
Iron and ferroalloy ores mining	4.32	3.59
Primary aluminum	4.28	3.66
Electrometallurgical products, except steel	4.25	3.56
Logging	4.18	3.04
Chemical and fertilizer minerals	4.09	3.31
Lime	3.90	2.78

(production stages and distance to final demand). We should also note that these two indexes are only weakly correlated across all commodities and manufacturing industries. The correlation is negative until 1982: -7.5% in 1967, -4.3% in 1972, -1.9% in 1977. Then it is between -1% and 1% after 1982. This small correlation shows that these two indexes capture different dimensions of the fragmentation of production and can be both informative to characterize the position of an industry along supply chains.

An overall comparison between commodities, manufacturing goods and services confirms the intuition above (Table 4). Manufacturing industries embody more production stages than commodities and commodities more than services. Commodities are further from final demand than manufacturing industries, while services are closer to final demand than manufacturing industries on average. The comparison between manufacturing goods and commodities carries

Table 3: Industries with the smallest number of production stages

Production stages	All inputs	Production stages	Tradables
Owner-occupied dwellings	1.23	Carbon black	1.03
Other Federal Government enterprises	1.25	Greenhouse and nursery products	1.08
Greenhouse and nursery products	1.33	Manufactured ice	1.14
U.S. Postal Service	1.34	Miscellaneous crops	1.15
Real estate	1.44	Forestry and fishery products	1.16

Table 4: Averages for groups of industries

Inputs from:	All industries		Tradables excl. oil	
Index:	Production stages	Stages to final demand	Production stages	Stages to final demand
Manufacturing	2.19	2.11	1.60	1.53
Commodities	2.06	3.01	1.38	2.45
Services	1.75	1.79	/	/
Petroleum	2.33	3.48	/	/

over if we only consider tradable inputs and exclude petroleum-related products.

Now I show that, among manufacturing industries, there are systematic differences between industries and these differences depend on various industry characteristics. The choice of these industry characteristics is primarily motivated by the literature on firm boundaries (see Lafontaine and Slade, 2007). Even if these measures of fragmentation only capture within-plant integration (boundaries of the plant), it may well be influenced by factors determining ownership (boundaries of the firm). Hortacsu and Syverson (2011) show that shipments that occur within the firm account for a very small portion of all shipments across plants. This suggests that the decision to integrate production within the same firm often goes along within-plant production.

The literature on the boundaries of the firm has identified various determinants of vertical integration. First, innovative industries rely less intensively on outsourcing whereas mature industries are more likely to outsource components (see Acemoglu, Aghion and Zilibotti, 2007). We can thus expect a negative correlation between R&D intensity and vertical fragmentation. Skill intensity and the complexity of tasks may also affect externalization decisions, with more complex tasks more likely to be performed within the firm (see Costinot, Oldenski and Rauch, 2009).[29] Following Antràs (2003) model based on the property-right approach, the internal-

[29]Here I focus on a measure skill intensity. I obtain similar results with the measure of non-routine vs. routine

ization decision can also depend on capital intensity. Capital-intensive industries rely more intensively on investment decisions taken by headquarters and are thus more likely to be integrated, whereas decisions taken by suppliers are relatively more important in labor-intensive industries leading to more outsourcing in these industries (a similar argument applies to R&D intensive industries vs. mature industries as in Antràs, 2005). Other factors affecting integration include competition and market thickness (McLaren, 2000) and financial constraints (Acemoglu, Johnson and Mitton, 2007, Carluccio and Fally, *forthcoming*). We proxy competition by the fraction of output produced by the 4 largest companies in the industry[30] and financial constraints by an index of external finance dependence (Rajan and Zingales, 1998).

Another important factor to be considered is product specificity. Nunn (2007) suggests that sourcing is more difficult or costly for specific product, especially when contracts are difficult to enforce (see also Hanson, 1995). The claim is not specifically made about the choice between outsourcing and integration, but applies to supplier-buyer relationships in general. As in Nunn (2007), I use Rauch (1999) classification to identify specific products (goods sold on thin markets). We can expect a negative correlation between specificity and vertical fragmentation.

Table 5: Pairwise correlations with industry characteristics

Variable:	Production Stages (N_i)	Distance to final demand (D_i)
Specificity	-0.266*	-0.498*
R&D	-0.259*	-0.038
Capital intensity	0.091	0.524*
Skill intensity	-0.219*	-0.167*
Advertising intensity	-0.083	-0.267*
Productivity	-0.025	0.030
Financial Dep	-0.185*	0.322*
Share of top 4 firms	-0.040	0.075

Notes: Variables for year 1992. A star denotes significance at 1%

Pairwise correlations between each index and these industry characteristics are shown in Table 5 (See Appendix for more details on data and variable definitions). The first column shows that high-tech industries generally embody a smaller average number of production stages. These results are in line with the literature on vertical integration. In particular, there is a negative and significant correlation for N_i with product specificity, R&D intensity, skill intensity and dependence in external finance. We may expect hi-tech industries to be

task developed by Costinot, Oldenski and Rauch (2009). The latter is however initially defined following the NAICS classification, which is difficult to match with the SIC classification.

[30] Alternatively, we can use the Herfindahl-Hirschman Index. Results are qualitatively the same.

more complex and combine multiple inputs, but complex inputs are more difficult to source from other plants.[31] We find however no significant correlation between N_i and either capital intensity, productivity or industry concentration.

Turning to the second column (index D_i), we find that industries that are further from final demand have lower values of product specificity and skill intensity. In particular, these industries are less intensive in the use of advertisements, which is quite intuitive (advertising industries are those that are closer to final consumers). These industries are also more intensive in capital and rely more heavily on external finance.[32]

4.2 Within-between decompositions of aggregate changes

Since the degree of vertical fragmentation varies sensibly across industries, I now examine whether the decrease in the overall fragmentation of production can be explained by composition effects. Is there a continuous shift towards industries with fewer production stages? Or can we only explain the overall decrease by changes within each industry?

Composition effects can occur along two dimensions. First, consumption may be shifting towards goods that require fewer production stages. Second, value added can *shift* towards industries that are closer to final demand, meaning that downstream industries contribute to a larger fraction of final goods value. Following Proposition 2, both shifts can contribute to the aggregate decrease in fragmentation. Hence, using these two indexes provides two different angles to look at these composition effects.

To examine these questions quantitatively, I decompose the change in the fragmentation of production into "between" and "within effects". Between two periods, the change in the aggregate index can be expressed as (Decomposition 1):

$$\Delta \bar{N}_t = \underbrace{\left[\sum_i \frac{(N_{i,t} + N_{i,t-1})}{2} \, . \, \Delta c_{i,t} \right]}_{Between} + \underbrace{\left[\sum_i \Delta N_{i,t} \, . \, \frac{(c_{i,t} + c_{i,t-1})}{2} \right]}_{Within}$$

with Δ denoting simple differences between periods t and $t-1$, and $c_{i,t} \equiv C_{i,t}/[\sum_j C_{j,t}]$ the share of consumption in section i at time t. Decomposition 1 is based on the number of production stages. Alternatively, we can use the distance to final demand weighted by value

[31] As mentioned before, the measure of vertical fragmentation N_i depends does not depend on how many different inputs are assembled, conditionally on the share of outsourced inputs in the value of the final good.

[32] Very similar results are obtained with multivariate regressions.

added (Decomposition 2):

$$\Delta \bar{D}_t = \underbrace{\left[\sum_i \frac{(D_{i,t} + D_{i,t-1})}{2} \cdot \Delta v_{i,t}\right]}_{Between} + \underbrace{\left[\sum_i \Delta D_{i,t} \cdot \frac{(v_{i,t} + v_{i,t-1})}{2}\right]}_{Within}$$

where $v_{i,t} \equiv V_{i,t}/[\sum_j V_{j,t}]$ denotes the share of value added in section i at time t. In each decomposition, the first term reflects a change in the composition (between effect) whereas the second term reflects changes within industries. As documented in Table 1, aggregate indexes $\bar{N}_t$ and $\bar{D}_t$ are almost equal to each other, and very close to the ratio of gross output to value added.[33] Hence, these two approaches can be seen as two alternative decompositions of the evolution of the aggregate average number of production stages.

I first decompose the change in the index calculated for all industries, including all inputs (Table 6, Panel A). Panel A shows similar results for both decompositions. In both decompositions, the within and between effects are equally large. Summing across all years, the between effect actually dominates. This negative trend for both indices can be explained by a shift of demand and production towards services. Services require fewer stages and are also closer to final demand. While the between effect is consistently negative in Decomposition 1, the between effect in Decomposition 2 is positive for the transition period between 1972 and 1977. This can be explained by the increase in basic commodity prices such as petroleum, which increases the share of industries that are further from final demand. For other years, the between effect is negative though. Similarly, increases in commodity prices can explain the positive within effect in the first decomposition (see Table 8 in the next section).

Then, I decompose the change in fragmentation by considering tradable goods only (manufacturing and commodities excluding petroleum). Panel B shows that the between effect in Decomposition 1 is much smaller for tradable goods, and a large part of the evolution across years is explained by the within effect. This confirms that part of the results from Panel A are driven by the shift towards services and shows that, among tradable goods, there has been no shift of consumption towards less fragmented goods. Hence, changes in consumption patterns across tradable goods do not explain the decline in production staging.

The between effect in Decomposition 2 remains large compared to Decomposition 1. Except for 1967, the variations in aggregate distance to final demand are mostly driven by the between effect. Except for 1972-1977 period, value-added has been shifting towards manufacturing industries that are closer to final demand.

[33] In theory, the weighted average of the number of production stages may differ from the weighted average of the distance to final demand in an open economy. However Table 1 show that, in practice, these two measures are almost equal to each other for the US.

Table 6: Within-between decompositions

	Average number of production stages: (decomposition 1 along N_i)			Average distance to final demand: (decomposition 2 along D_i)		
Year	Aggregate change	Between effect	Within effect	Aggregate change	Between effect	Within effect
Panel A: **All industries**						
67-72	-0.087	-0.028	-0.059	-0.100	-0.023	-0.078
72-77	0.070	-0.009	0.078	0.049	0.032	0.016
77-82	0.013	-0.033	0.045	0.026	-0.016	0.042
82-87	-0.086	-0.007	-0.079	-0.097	-0.041	-0.055
87-92	-0.031	-0.030	-0.001	-0.014	-0.010	-0.004
Panel B: **Tradeable goods**						
67-72	-0.127	0.022	-0.148	-0.136	-0.002	-0.134
72-77	0.011	-0.024	0.035	0.025	0.042	-0.017
77-82	-0.079	-0.030	-0.049	-0.074	-0.055	-0.019
82-87	-0.072	-0.002	-0.070	-0.107	-0.055	-0.052
87-92	-0.006	-0.006	0.001	0.019	0.009	0.010

Notes: Panel A: all industries are included except petroleum; Panel B: primary and secondary industries are included except petroleum. See text for within and between decomposition. It is applied to the number of production stages in columns 3 and 4 and to the number of stages to final demand in columns 5 and 6. The values in column 2 (difference in aggregate GO/VA between two years) equal the sum of columns 3 and 4 and also the sum of columns 5 and 6.

In what follows, I will first examine the evolution of the number of production stages N focusing on the within effect of decomposition 1 (section 4.3). Then I will turn to distance to final demand by providing additional evidence on the shift of value-added towards final stages (section 4.4). The latter provides simple and intuitive insights on the aggregate decrease in the weighted number of production stages.

4.3 Determinants of "Within" changes

As we have shown previously (decomposition 1), the aggregate decrease in fragmentation in the manufacturing sector mostly corresponds to "within" effects rather than a shift of consumption towards goods that require fewer production stages. As motivated previously, whether production chains are more vertically fragmented across plants may depend on the complexity of tasks, on the need for capital, on the thickness of upstream or downstream markets, etc. Now, are there empirical regularities that could explain the *change* in vertical fragmentation by industry?

Table 7 explores the change in fragmentation by industry depending on various industry characteristics. The dependent variable is increase in the index of fragmentation: $\Delta N_i = N_{i,1992} - N_{i,1967}$. Results in column (1) show that the change in fragmentation is positively related to product specificity (measured by Rauch 1999 index), R&D intensity and capital intensity, and negatively related to skill intensity and financial dependence. All-in-all, these industry characteristics can account for about 12% of the variance in the change in fragmentation (R-squared). The positive correlation with variables characterizing hi-tech industries (such as R&D intensity) can be consistent with a product-cycle interpretation: innovative industries become more fragmented as they mature (see e.g. Antràs 2005). However, while this interpretation might help understand why some industries are becoming more fragmented than others, it does not shed light on the overall decline fragmentation.

Could the decline in fragmentation be explained by *changes* in these industry characteristics? As R&D, skill and capital intensity are important determinants of differences vertical fragmentation across industries, one might suspect that changes in R&D, capital and skill intensities might be driving the decrease in fragmentation. However, as shown in Table 8, column (2), I find that the increase in the number of production stages by industry does not significantly depend on the increase in R&D, skill and capital intensity.[34]

The number of stages to produce goods in a certain industry may not just depend on the characteristics of this industry but may also depend on the characteristics of the upstream industry. To examine how upstream industry characteristics are related to vertical fragmentation, I replace each variable by a weighted average of the corresponding upstream values. To be more precise, I follow the Nunn (2007) methodology and construct a set of variables $x_{it}^{k,up}$ such that:[35]

$$x_{it}^{k,up} = \frac{\sum_j \mu_{ijt} x_{it}^{k}}{\sum_j \mu_{ijt}} \tag{6}$$

As shown in column (3) of Table 7, the results based on upstream industry characteristics are qualitatively similar to column (2). Coefficients are larger but the standard deviation of average upstream characteristics is smaller and thus the implied beta coefficient for most variables is smaller than for column (2).

[34]Data on industry characteristics over time are available for skill and capital intensity for the full period (1967-1992) and data on R&D intensity changes for the period 1982-1992. One may be also interested in the evolution of product specificity, but such information is difficult to obtain.

[35]Industry characteristics are not defined for all upstream industries. I restrict the sum over upstream industries for which the corresponding variable is available.

Table 7: Within-industry changes

Dependent variable:	ΔN	ΔN	ΔN
Specificity	2.121	2.018	3.465
	[0.665]***	[0.669]***	[1.984]*
R&D intensity	0.565	0.539	1.367
	[0.153]***	[0.184]***	[0.435]***
Capital intensity	1.490	1.395	2.641
	[0.475]***	[0.480]***	[1.334]**
Skill intensity	-6.285	-6.239	-10.286
	[3.052]**	[3.047]**	[5.477]*
Advertising intensity	-0.018	-0.022	-0.021
	[0.079]	[0.079]	[0.075]
Productivity growth	-2.666	-2.458	-5.649
	[2.052]	[2.252]	[6.551]
Financial Dependence	-0.293	-0.299	-0.621
	[0.168]*	[0.175]*	[0.328]*
Top 4 share	-0.014	-0.018	-0.060
	[0.013]	[0.014]	[0.032]*
Δ R&D intensity		0.259	1.320
		[0.521]	[1.284]
Δ K-intensity		0.024	-0.163
		[0.029]	[0.049]***
Δ skill intensity		-0.155	-0.546
		[0.186]	[0.374]
Characteristics	Same industry	Same industry	**Upstream** industry
Number of industries	266	266	266
R-squared	0.12	0.12	0.15

Notes: OLS regressions. Dependent variable: increase in the number of production stages by industry between 1967 and 1992. Data on industry characteristics are described in the appendix. Robust standard errors into brackets; * significant at 10%; ** significant at 5%; *** significant at 1%.

Adjusting for prices

Since the measure of fragmentation developed here is based on value-added weights, a natural question is whether changes in these weights (and changes in the overall measure) is not simply reflecting changes in relative prices along value chains. For instance, if competition among suppliers has eroded their bargaining power compared to final goods producers, we could expect the relative price of intermediate goods to decrease, thus reducing the share (in value) of intermediate goods in final goods production. Such an effect would be reflected in the index N as a decrease. To disentangle such price effects, I propose a further decomposition of the within effect computed in Table 6 (decomposition 1).

Let $Q_{ij,t}$ denote the quantity of intermediate good j used in the production of good i at time t. The input-output coefficient could then be rewritten as:

$$\mu_{ij,t} = \frac{P_{j,t}Q_{ij,t}}{P_{i,t}Q_{i,t}}$$

where $P_{i,t}$ denotes the price index of goods produced by industry i, and where $Q_{i,t}$ denotes total output quantity of industry i. Looking at the evolution across years, we can decompose the change in input-output coefficient $\Delta\mu_{ij,t} \equiv \mu_{ij,t} - \mu_{ij,t-1}$ in two components:

$$\Delta\mu_{ij,t} = \underbrace{\left(\frac{P_{j,t}}{P_{i,t}} - \frac{P_{j,t-1}}{P_{i,t-1}}\right) \cdot \frac{1}{2}\left(\frac{Q_{ij,t}}{Q_{i,t}} + \frac{Q_{ij,t-1}}{Q_{i,t-1}}\right)}_{\substack{\Delta\mu^{P}_{ij,t} \\ \text{Price effect}}} + \underbrace{\frac{1}{2}\left(\frac{P_{j,t}}{P_{i,t}} + \frac{P_{j,t-1}}{P_{i,t-1}}\right) \cdot \left(\frac{Q_{ij,t}}{Q_{i,t}} - \frac{Q_{ij,t-1}}{Q_{i,t-1}}\right)}_{\substack{\Delta\mu^{Q}_{ij,t} \\ \text{Quantity effect}}}$$

Using data on relative price indices, quantity ratios are simply obtained by dividing the input-output coefficient by the relative price ratio $\frac{Q_{ij,t}}{Q_{i,t}} = \frac{\mu_{ij,t}P_{i,t}}{P_{j,t}}$.

To see how the change in the input-output coefficient $\Delta\mu_{ij,t}$ impacts the within-industry change in the fragmentation index, we can write:

$$\Delta N_{i,t} = \sum_j \Delta\mu_{ij,t}\left(\frac{N_{j,t} + N_{j,t-1}}{2}\right) + \sum_j \left(\frac{\mu_{ij,t} + \mu_{ij,t-1}}{2}\right)\Delta N_{j,t}$$

We can see this equality as a linear equation in $\Delta N_{i,t}$, the change in the fragmentation index for each industry. Inverting this equation, we can write the change in the index as a function of the change in input-output coefficients:

$$\Delta N_{i,t} = \sum_k a_{ik,t}\left[\sum_j \Delta\mu_{kj,t}\left(\frac{N_{j,t} + N_{j,t-1}}{2}\right)\right]$$

where $a_{ik,t}$ denotes the coefficients of the matrix $(I-M)^{-1}$ where I is the identity matrix and M is the matrix with coefficients $\frac{\mu_{ij,t}+\mu_{ij,t-1}}{2}$ (i.e. an average of the two input-output matrices for period t and $t-1$). Using the price-quantity decomposition of the change in direct coefficients, we thereby obtain a decomposition of the changes in the fragmentation index for each industry:

$$\Delta N_{i,t} = \underbrace{\sum_k a_{ik,t}\left[\sum_j \Delta\mu^{P}_{kj,t}\left(\frac{N_{j,t} + N_{j,t-1}}{2}\right)\right]}_{\text{Price effect}} + \underbrace{\sum_k a_{ik,t}\left[\sum_j \Delta\mu^{Q}_{kj,t}\left(\frac{N_{j,t} + N_{j,t-1}}{2}\right)\right]}_{\text{Quantity effect}}$$

Table 8: Price vs. quantity decomposition - Tradeable goods

Year	Within effect	Price effect	Quantity effect
67-72	-0.148	0.006	-0.154
72-77	0.035	0.076	-0.041
77-82	-0.049	0.008	-0.057
82-87	-0.070	-0.025	-0.045
87-92	0.001	-0.011	0.012

Notes: The within effect is the same as in Table 6, panel B, and equal the sum of the quantity and price effects.

To proxy for price ratios, I use data on producer price indices from the NBER CES database (for manufacturing industries) and from the BLS (for other commodities).

In Table 8, I compute this decomposition to isolate the role of prices in explaining the within effect in the decomposition of the fragmentation index. Interestingly, price effects are very small except for transition period between 1972 and 1977, where the evolution of prices (increase in the relative price of intermediates) can explain a large increase in the fragmentation index. The quantity effect is however negative for 1972-1977, like other years, suggesting that the index of vertical fragmentation would have decreased during this period if relative prices had remained stable. This table shows that the negative trend in the index N cannot be simply explained by price changes.

While these results suggest that changes in relative price do not explain the overall decrease in vertical fragmentation, one must remain careful about potential price measurement errors. As measured by the BLS, price indices do not fully account for the introduction of new varieties.[36]

Trade and vertical fragmentation

Trade can have two opposite effects. As trade barriers fall, production chains increasingly involve parties located in different countries (Yi, 2003). International trade provides new opportunities to reduce costs by shifting part or entire production abroad. It is thus natural to expect a positive effect of trade on the fragmentation of production. Note however that trade does not affect our measure of fragmentation if there is simply a substitution between domestic outsourcing and foreign outsourcing. As described in Section 2, the measure of fragmentation is based on the total use of inputs and does not differentiate shipments from another plant in the US and shipments from overseas. Hence, if trade is found to have a positive impact, it

[36] This is known as the "outlet substitution bias" in the consumer price index literature. A similar issue arises with international trade and the availability of new imported varieties (Houseman *et al.*, 2011).

Table 9: Import penetration and the measure of production stages

Dependent variable:	N	D	ΔN	ΔN	ΔN	ΔN
Import penetration	-0.032	-0.475				
	[0.128]	[0.307]				
Increase in imports			0.218	0.168		
(same industry)			[0.060]***	[0.076]**		
Increase in imports					0.260	0.240
(upstream industry)					[0.137]*	[0.204]
Controls	No	No	No	**Yes**	No	**Yes**
Nb of industries	305	305	305	266	305	266
R-squared	0.01	0.01	0.03	0.16	0.01	0.13

Notes: OLS regressions. Dependent variables: Average number of production stages (N) and stages to final demand (D) in 1992; ΔN increase in N between 1967 and 1992. Independent variables: average import penetration (col. 1 and 2); increase in import penetration in the same industry (col. 3 and 4); increase in average import penetration in upstream industries (col. 5 and 6). Controls include all industry characteristics shown in column (1) of Table 7. Robust standard errors into brackets; * significant at 10%; ** significant at 5%; *** significant at 1%.

would suggest that it substitutes to tasks that were previously performed within the plant.

There may be also a negative effect of trade on this measure of fragmentation. If trade reduces the relative price of intermediate goods, there is a possibility that it also reduces the amount spent on these goods, and therefore reduces the share of value added associated with upstream stages.[37] Note that this can only occur if there is a very low substitution between outsourced intermediate goods (domestically or internationally) and intermediate goods produced within the plant, otherwise a negative effect of trade on relative prices would also imply an increase in the share of outsourced intermediate goods.

A first question is whether the number of production stages or the position on the value chain is correlated with import penetration across industries (in the cross section). In Table 1 on aggregate measures of fragmentation corrected for trade, results show that there is only a small correlation between either net imports and production stages or net imports and the distance to final demand. In Table 9, I confirm this result by regressing the number of production stages (column 1) and the distance to final demand (column 2) on import penetration across industries (all variables are averaged across periods). Import penetration is defined as the ratio of imports to production plus imports minus exports in each industry. I find no significant

[37]The results above suggest that price effects are small but these price indices do not perfectly account for new varieties of traded inputs.

correlation (OLS regression with robust standard errors). Interestingly, I even find a negative (but not significant) correlation between import penetration and D_i, suggesting that import competition has become relatively tougher in downstream industries than upstream industries. Given this result, it appears also unlikely that import competition has induced a decrease in the relative price of upstream goods vs. compared to downstream goods. This is in line with the fact that price indexes have not decreased relatively faster in upstream industries.

From the small correlation between trade and import penetration in a cross-section analysis, we should however not conclude that trade doesn't affect vertical fragmentation. I now examine whether changes in import penetration are related to changes in the fragmentation of production. For this purpose, I regress the change in the measure of production stages (ΔN_i) by industry on the increase in import penetration between 1967 and 1992, by industry. In columns (3) and (4), I find a positive and significant effect which could suggest that trade indeed creates new opportunities to fragment production. Controlling for other industry characteristics does not greatly affect the main coefficient (column 4).

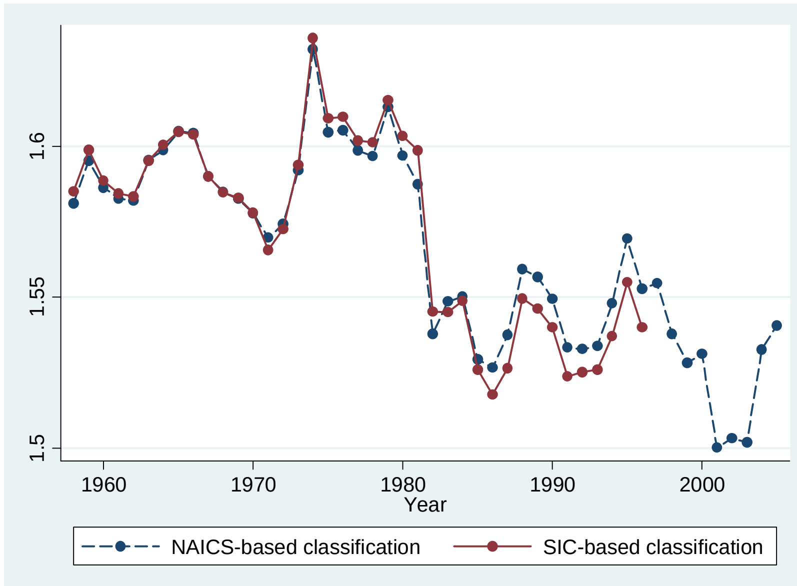
More importantly, we would like to know whether imports of *inputs* are associated with an increase in the number of production stages. For this purpose, I compute the change in average import penetration among upstream industries (as I did in equation 6 for other industry characteristics), and use it instead of the change in import penetration within the same industry. In columns (5), I find a larger coefficient but the associated beta coefficient is smaller (0.11 against 0.20) and less significant. As shown in column (6), the coefficient for imports is no longer significant when additional controls are added. Similarly, I find no significant effect after instrumenting the change in import penetration by transport costs and tariff decreases in upstream industries. Hence, while opening to trade seems to be positively associated with an increase in vertical fragmentation across plants, its effect is not large and robust.

4.4 A shift of value added towards downstream industries

As shown in Proposition 1, the index N can be interpreted as the weighted average number of production stages, weighted by value being added at each stage. Hence, equivalently, a decrease in N can be interpreted as a shift of value towards final stages. While plant-level data between buyers and suppliers are not available, we can still examine the shift of value added towards industries that are closer to final demand. This shift corresponds to the between effect associated with distance to final demand (decomposition 2) in Table 6.

A similar way to illustrate this shift is to examine the value-added-weighted average distance to final demand, using panel data on value added but using a reference value for the index of

Figure 4: VA-weighted distance to final demand



Notes: Distance index measured with the 1992 (SIC-based) and 2002 (NAICS-based) input-output tables. Value-added data are from the NBER-CES database.

distance to final demand for each sector. To be more precise, I compute for each year:

$$\widetilde{D}_{v,t} = \sum_i v_{it} D_{i,1992}$$

where $D_{i,1992}$ is the distance index associated with industry i in year 1992 (or an alternative year) and v_{it} is the share of value added from sector i at time t. Hence, keeping the distance index constant, the observed change in $\widetilde{D}_{v,t}$ would solely reflect a change in the industry composition. Moreover, we are no longer restricted to "benchmark" years since data on value added are available from other sources. Here, I use data on manufacturing value-added from the NBER-CES database available on a SIC-based classification between 1958 and 1996 (this dataset does not cover primary industries). To also examine what happened in subsequent years, I also compute the distance index using the 2002 input-output matrix (based on the NAICS classification) to be combined with NBER-CES data available on a NAICS basis until 2005 included.

Figure 4 illustrate the evolution of $\widetilde{D}_{v,t}$. We can indeed observe an overall shift of production towards downstream sectors during these five decades except between 1973 and 1981 when the price of oil and other basic commodities have dramatically increased.

Table 10 provides yet another way to examine the shift of value added. In columns (1) to

Table 10: Shift of value-added towards final stages

Dependent variable:	VA Growth	VA Growth	VA Growth	Increase in VA/GO	Increase in VA/GO	Price Growth
Stages to final demand	-2.927	-2.951	-2.365	-0.403	-0.270	-0.122
	[1.108]***	[1.119]***	[1.430]*	[0.151]***	[0.207]	[0.597]
Number of stages		0.530	0.837			
		[2.931]	[3.245]			
Specificity			-5.985		-0.400	
			[3.025]**		[0.379]	
R&D intensity			0.323		-0.213	
			[0.688]		[0.084]**	
Capital intensity			-5.330		-0.704	
			[2.088]**		[0.253]***	
Skill intensity			8.126		3.110	
			[9.921]		[1.621]*	
Advertising intensity			0.712		0.003	
			[0.299]**		[0.077]	
Productivity growth			26.580		0.456	
			[11.381]**		[1.027]	
Financial Dependence			1.173		0.210	
			[0.522]**		[0.083]**	
Top 4 share			-0.065		0.003	
			[0.043]		[0.007]	
Import penetration			-36.838		-2.022	
			[8.848]***		[1.104]*	
Number of industries	305	305	266	305	266	305
R-squared	0.02	0.02	0.20	0.02	0.12	0.00

Notes: OLS regressions. Dependent variables: growth of value added by industry between 1967 and 1992 (columns 1 to 3); increase in the value-added-to-gross-output ratio (columns 4 and 5); growth of industry price index (column 6). Independent variables: averages between 1967 and 1992; data on industry characteristics are described in the appendix. Robust standard errors into brackets; * significant at 10%; ** significant at 5%; *** significant at 1%.

(3), I test whether value added has grown significantly more in industries that are closer to final demand (OLS regressions with robust standard errors). The dependent variable is the growth in VA by industry between 1967 and 1992, while the independent variable is the distance to final demand by industry (1967-1992 average).[38] The coefficient is negative and significant; the beta coefficient equals -0.221.

This result clearly confirms the shift of value added towards downstream industries, which is consistent with the negative between effect found in Table 6 (Panel B, Decomposition 2). In column (2), we control for the number of production stages. The coefficient is not significant,

[38]Using 1992 data on the distance index (instead of averages) gives even more significant results.

which reflects the small between effect found in decomposition 1.[39] In column (3), we control for other industry characteristics: Product specificity, R&D intensity, capital and skill intensity, advertising intensity, productivity growth, financial dependence and industry concentration. The coefficient for distance to final demand remains significant but is now smaller. In particular, part of the negative correlation between value-added growth and distance to final demand can be explained by a larger growth in advertising-intensive industries (which are closer to final demand). We also control for import penetration, which has a strongly negative coefficient.

Interestingly, the ratio of value added to gross output (by industry) exhibits a similar pattern. In columns (4) and (5), the dependent variable is the increase (simple difference) in VA/GO between 1967 and 1992, regressed on the distance to final demand by industry. The coefficient is also significantly negative; the beta coefficient equals -0.242 in column (4). In this regression, the constant equals +1.30. We can test and verify that VA/GO has significantly increased for industries that are the closest to final demand, while it has significantly decreased for industries with a measure of distance to final demand equal to 3. These results remain fairly unaltered after controlling for other industry characteristics and import penetration.

Since the growth in value-added is affected by changes in prices, a natural question is whether the shift of value added does not simply reflect an erosion of the relative prices of intermediate goods, e.g. driven by an increase in competition among suppliers and an erosion of their bargaining power relative to downstream producers. To examine this hypothesis, I use industry price data from the NBER-CES database and the Bureau of Labor and Statistics (BLS) over the same time period.[40] In column (6), I regress the change in the industry-level price index on the measure of distance to final demand. The coefficient is however very small and not statistically significant. Additional evidence on relative prices is provided in the robustness section, showing that the price of basic commodities and intermediate goods (compared to final goods) has not decreased over the past decades.

A straightforward explanation for the shift towards downstream industries is that value-added growth has been driven by other factors (e.g. shift towards high-tech industries) and that these factors are themselves related to the distance to final demand. In particular, value added has grown faster in industries that are intensive in R&D, in skills, in advertising, in external finance, and less intensive in physical capital. In turn, these industries are generally closer to final demand (see Table 5) which can explain why value-added growth is negatively correlated with distance to final demand. To examine this explanation quantitatively, I perform

[39] Alternatively, we can use the growth of consumption as the dependent variable. The coefficient for the number of production stages is also not significant.

[40] For a easier concordance with SIC-based input-output classifications, I use the NBER data for most industries (manufacturing) and complete with raw data from the BLS.

the following exercise:

i) First, I regress value-added growth on industry characteristics (all control variables from column 3 of Table 10) excluding the two measures of fragmentation: product specificity, R&D intensity, skill intensity, capital intensity, advertising intensity, productivity growth, dependence in external finance, industry concentration. The regression coefficients are almost identical to those in column 3 of Table 10 for the corresponding variables.

ii) Then, I use the predicted value-added growth by industry from step 1 and regress the constructed variable on distance to final demand.

The resulting coefficient is -1.50 (significant at 1%). It is more than half of the coefficients from Table 10, column 1. This result suggests that these industry characteristics can explain half of the negative correlation between value-added growth and distance to final demand, which itself can shed light on the aggregate decrease in vertical fragmentation. These findings on the shift of value-added also demonstrate that the overall decline in fragmentation is not counter-intuitive if we see it from this angle: US activities that have grown the fastest are those at the last stages of production chains, which also implies that intermediate goods and early-stage production is becoming relatively less important.

A global shift towards downstream industries

While the previous results document the fragmentation of production in the US, and particularly the shift of value-added towards downstream industries, one can ask whether similar results can be observed for other countries and for trade flows.

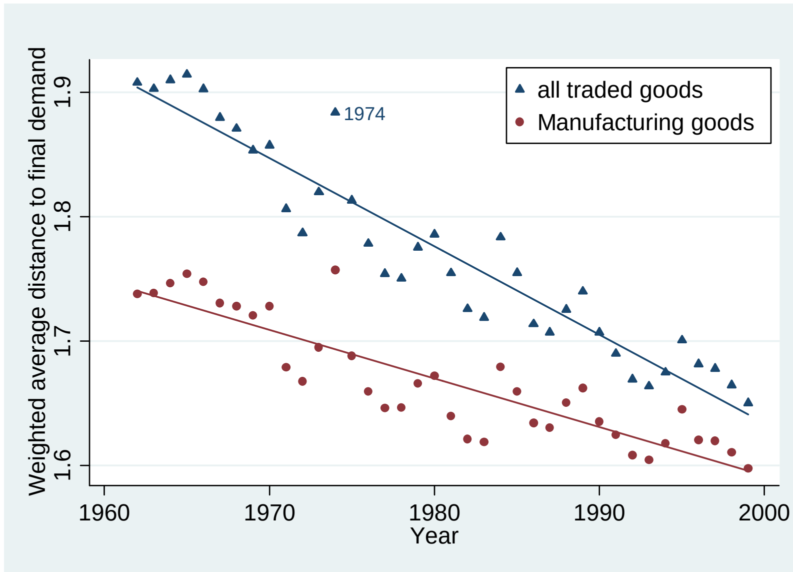
As shown by Hummels, Ishii and Yi (2001), and more recently by Johnson and Noguera (2012), production has become more fragmented *across borders*. This can be shown (see Johnson and Noguera, 2012) as an overall decrease of the ratio of value-added content of trade to total trade (VAX ratio) which corresponds, as I show in Section 2.4, to the inverse of the average number of border crossings embodied in traded goods.

Surprisingly, I find however that trade flows have shifted towards *downstream* industries, in parallel to the shift of value-added in the US.[41] To document this fact, I construct the average of distance to final demand across industries weighted by the total value of world trade:

$$\widetilde{D}_{x,t} = \sum_i x_{it}^{world} D_{i,1992}$$

[41] A similar finding has been pointed out by Hummels, Ishii and Yi (2001). Looking at trade across Broad Economic Classifications (distinguishing goods into capital, consumption and intermediate goods), the share of intermediate goods trade has been decreasing from 50% in 1970 to 40% in 1992. As discussed earlier, the BEC classification has some drawbacks while D_i better accounts for the position on the value chain.

Figure 5: Trade-weighted distance to final demand (based on 1992 IO table)



where $D_{i,1992}$ is the distance index associated with industry i in year 1992 (or an alternative year) and x_{it}^{w} is the share of total trade of product i in world trade, at time t. To compute x_{it}^{w}, I use multilateral trade data from the UN-NBER database between 1962 and 1996.[42]

The evolution of $\widetilde{D}_{x,t}$ is shown in Figure 5. The decline in average distance to final demand is even starker than for US value-added. The year 1974 is an outlier although petroleum-related trade flows have been dropped for the calculation of the weighted-average. One may think that this shift simply reflect an increasing share of manufacturing goods relative to basic commodities, but a similar trend is obtained if we just look at trade flows in manufacturing industries.

Note that the fact that trade has shifted towards more downstream industries is not inconsistent with an overall increase in *international* fragmentation (as shown for instance by the decrease in the VAX ratio, Johnson and Noguera, 2012). For instance, let us consider two industries: a downstream industry and an upstream industry. If international trade is initially more concentrated in the upstream industry (e.g. basic commodities such as steel or oil), the value-added content of trade can be large since it would embody only a small fraction of intermediate goods. Now, if international trade begins to occur in both downstream and upstream

[42]These trade data are available in the revision 2 of the SITC classification. I have used various concordance tables between SITC and SIC industries to combine the trade data with input-output measures. Alternatively, I obtain extremely similar results by using more precise concordance tables between SITC and HS product classifications, and then between HS and NAICS classification, to be finally combined with distance indexes contructed from the NAICS-based 2002 input-output table.

industries, the value-added content of trade may well decrease since trade flows in downstream industries would rely more intensively on intermediate goods, part of them being imported.

It is interesting to see that the shift toward downstream activities is not a uniqueness of the US economy and is also reflected in world trade flows. We obtain the same trend, whether US imports and exports are included in the computation above.

4.5 Vertical specialization and trade patterns

This section provides an application of the two measures of fragmentation developed in this paper. These two measures provide novel information on the position of each industry along production chains which is not captured by existing indices of fragmentation.

We have seen in Table 9 that import penetration is not significantly correlated to the number of production stages across industries. However, trade patterns and the source of imports may be related to the degree of fragmentation. A recent paper by Costinot, Vogel and Wang (2011) develops a simple model where stages along production chains are naturally sorted across countries depending on their productivities. They predict that poor countries specialize in early stages while more developed countries specialize in final stages. They also predict that poor countries should be involved in shorter production chains, while developed countries specialize in longer production chains.

In order to test these predictions, I regress US imports in 1992 (by industry i and source country c) on industry dummies, country dummies and two interaction terms: i) between GDP per capita of the source country c and the number of production stages in industry i (measured for the US as above); ii) between GDP per capita and the distance to final demand (fragmentation index D_i measured as above):[43]

$$\log E[M_{ic}] = \beta_N \,.\, N_i \,.\, \log(pcGDP_c) + \beta_D \,.\, D_i \,.\, \log(pcGDP_c) + \alpha_i + \eta_c$$

Such approach using interaction terms has been put forward by Romalis (2004) and Nunn (2007) among others. In line with Costinot et al (2011), we should find a positive coefficient β_N (richer countries specialize in goods involving more stages) and a negative coefficient β_D (richer countries specialize in stages that are closer to final demand). Such a comparative advantage of richer or poorer countries for fragmented industries might well be explained by “traditional” sources of comparative advantage since patterns of fragmentation are related to other industry characteristics (Table 5). Thus, I further control for interactions between capital intensity and capital endowments, skill intensity and skill endowments (as in Romalis, 2004).

[43] I estimate this equation with Negative Binomial PML which allows for zeros and overdispersion in the residual.

Another related question that we can ask is how vertical fragmentation interacts with trade costs. Hillberry and Hummels (2000) and Yi (2010) suggest that multi-stage production magnifies the impact of trade costs.[44] By looking at US imports across source countries and industries, I examine this claim by regressing imports (by industry i and source country c) on industry dummies, country dummies and an interaction term between physical distance from the source country c and the number of production stages in industry i (measured for the US as above), as well as an interaction term between physical distance and the number of stages to final demand:

$$\log E[M_{ic}] = \gamma_N \,.\, N_i \,.\, \log(distance_c) + \gamma_D \,.\, D_i \,.\, \log(distance_c) + \alpha_i + \eta_c$$

In particular, Hillberry and Hummels (2000) suggest that upstream industries are more sensitive to distance, thus predicting $\gamma_D < 0$. In line with Yi (2010), we can also expect a negative interaction term γ_N between physical distance and the number of embodied production stages.

As the fragmentation of production may be correlated with the transportability of goods (weight or other traits rendering a good less tradable), I also control for an interaction term between a proxy for tradability and physical distance. The variable on tradability is constructed using the ratio of the difference between c.i.f. and f.o.b. values over c.i.f. values of imports coming from a few key Asian countries.[45]

Table 11 presents the results for the two types of regressions specified above. Surprisingly, I find that rich countries are more likely to export goods involving fewer production stages, as shown by the negative and significant interaction terms in column (1). [46] Moreover, richer countries seem to specialize in industries that are closer to final demand. The latter is consistent with Costinot et al (2011) while the former is not. In column (2), I further control for endowments in skills and capital and interactions with skill and capital intensities (which are both positive and significant as in Romalis, 2004). With these controls, results are more in line with Costinot et al (2011) with a stronger coefficient for the interaction with the number of stages to final demand and a smaller coefficient for the interaction with the number of production stages. These results hold after including even more controls as in column (4).

As shown in column (3), the interaction term between physical distance and the number of production stages N_i is positive, showing that distance has a smaller effect when more stages are required. Distance however has a stronger effect in industries that are more upstream, as measured by the index of vertical fragmentation D_i. The interaction term between physical distance and distance to final demand is not significant in column (3) but it bcomes significant in

[44] Yi (2010) focuses on the border effect but the same argument applies to transport costs.

[45] This method follows Rauch (1999).

[46] As N_i and D_i are not significantly correlated with each other, the same interaction term is obtained for either of them whether I include the other one or not.

Table 11: Comparative advantage along supply chains

Dependent variable:	Imports	Imports	Imports	Imports
pcGDP$_c$ * production stages N_i	-0.421	-0.209		-0.190
	[0.091]***	[0.096]***		[0.099]***
pcGDP$_c$ * stages to final demand D_i	-0.065	-0.180		-0.192
	[0.033]*	[0.042]***		[0.043]***
Distance$_c$ * production stages N_i			0.623	0.412
			[0.153]***	[0.166]***
Distance$_c$ * stages to final demand D_i			-0.044	-0.114
			[0.055]	[0.058]***
Skill endowment$_c$ * Skill intensity$_i$		6.118***		6.458***
K endowment$_c$ * K intensity$_i$		0.314***		0.315***
Distance$_c$ * transportability$_i$				-7.078***
Industry fixed effects	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Observations	46412	31696	46412	31696
Log pseudolikelihood	-34629	-31032	-34635	-30970

Notes: Negative binomial ML regressions. Robust standard errors into brackets; * significant at 10%; ** significant at 5%; *** significant at 1%.

column (4) after including additional controls such as an internation between physical distance and goods transportability.

These results exhibit interesting patterns of trade and comparative advantage as well as differential effects of distance along value chains. Other applications of the second index D_i ("upstreamness") are explored in Antràs, Chor, Fally and Hillberry (2012) where we examine the role of institutional quality in explaining patterns of specialization along value chains.

5 Robustness

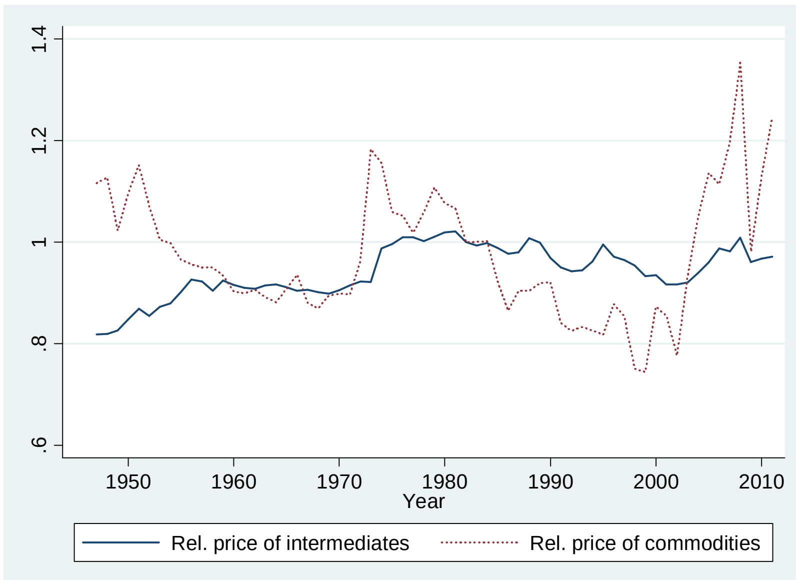
5.1 On the evolution of relative prices

The measure of vertical fragmentation is constructed using values of purchased intermediate goods, not quantities. Hence, one may be concerned that the decline in fragmentation is simply driven by changes in prices.

Intermediate vs. final goods prices

A first concern is that commodity prices and intermediate goods prices might have decreased compared to the price of final goods. Keeping quantities constant, this would explain a downward trend in the fragmentation index. To investigate this issue, I compare producer price index series from the Federal Reserve Economic Database (FRED) for different types of goods. In particular, I consider the following series: i) "Finished Consumer Goods"; ii) "Intermediate Materials: Supplies & Components"; iii) "Crude Materials for Further Processing". Figure 6 plots the ratio of the price index of the second and third category over to the first one (yearly average).

Figure 6: Relative price of commodities and intermediate goods compared to final goods



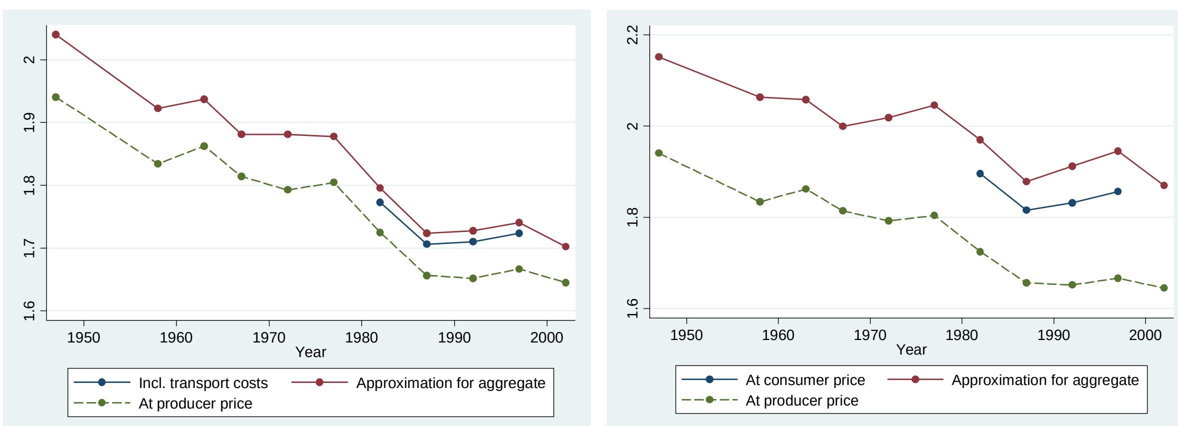
There is no evidence that intermediate goods prices have declined compared to final goods over the 1947-2002 period. As shown in Figure 6, there has been instead an overall increase in the relative price of intermediate goods. Concerning the relative price of commodities, there is no decline over the period 1967-1992 (period corresponding to the results presented in Table 1 to 10) and only a small decline if we compare 1947 to 2002. Given the relatively small share of commodities in total production (10% of value added and gross output), this change is not large enough to explain the decrease of the measure of fragmentation.

Consumer vs. producer prices

A second issue is that the BEA input-output tables are mainly based on producer prices. This might be a concern if the main focus is the decision to outsource by the downstream firm: consumer prices would be more appropriate. From 1982 onward, the BEA input-output tables include coefficients based on consumer price, with details on transport margins, retail and wholesale margins. Such data are not available for previous tables (1947-1977) at the industry level. For the aggregate economy, we can however approximate the index of fragmentation. If μ is the ratio of intermediate goods use to gross output, and τ the total amount of spent on trade costs divided by gross output, the corrected measure of fragmentation equals $\frac{1}{1-\mu-\tau}$ instead of $\frac{1}{1-\mu}$. In order to approximate τ, I use input-output coefficients associated with the use of retail, wholesale and transportation industries as inputs.

Figure 7 (a) plots the measure of fragmentation after incorporating transportation margins only. The corrected index of fragmentation is larger as it puts more weight on intermediate goods. The approximated curve is even above the curve using actual consumer prices, but not by far. As Figure 7 (a) shows, transportation margins have remained fairly constant over the past decades and thus the negative trend in vertical fragmentation is confirmed. Similarly, the negative trend still appears after incorporating retail, wholesale as well as transportation margins (Figure 7b), even if retail and wholesale margins have slightly increased.

Figure 7: Incorporating (a) transportation and (b) retail margins



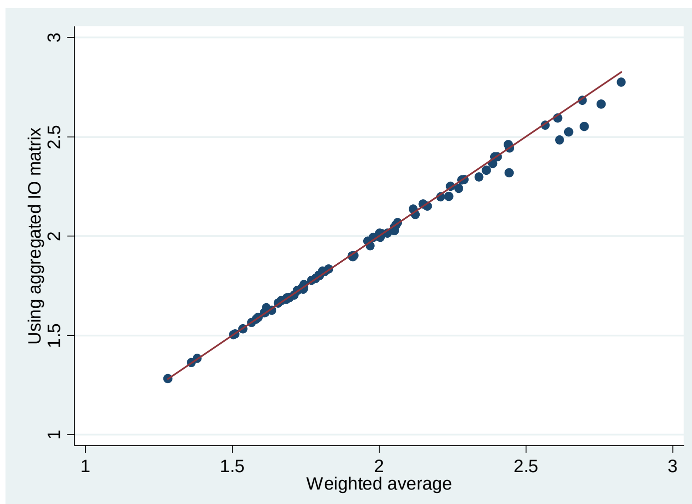
5.2 Aggregation

As shown by Proposition 3, the level of disaggregation possibly matters for an open economy when net imports are correlated with either measure of fragmentation. By aggregating too much, one might underestimate this correlation. However, for the US, the correlation between trade and fragmentation measure is so small (at the 6-digit level) that it's unlikely that we would find large correlations if we had more disaggregated data. Hence, it is quite unlikely that our main result on the aggregate decline is driven by an aggregation bias.

As shown by Proposition 4 and 5, results at the industry-level might be sensitive to the level of disaggregation when characteristics of production across varieties within an industry are systematically related to characteristics of the buying industry. In order to check whether the level of aggregation matters, I artificially construct an aggregated input-output matrix at the 3-digit level (similar results are obtained at the 2-digit level), I reconstruct the index of fragmentation using this aggregate matrix, and I compare with the appropriately-weighted average of the disaggregated measure.

I find that the new index is always very close (less than 1% difference on average) to the average of the disaggregated ones. This is depicted in Figure 8 where I plot the measured index using the aggregated input-output table as a function of the average of the index calculated across sub-industries using the disaggregated input-output table. We can see that the two measures differ only for extreme industries (generally belonging to the food industry).

Figure 8: Aggregation at the 3-digit level



This robustness to aggregation is comforting and promising for future studies as most countries beside the US do not have precise input-output tables. For the US, where more precise but still imperfect input-output tables are available, this suggests that the results of this paper would probably not be very different if even more detailed tables were available.

5.3 An alternative index of vertical fragmentation

While this measure of fragmentation aims at reflecting the number of plants that production is sequentially going through, it might not well reflect whether production is actually dispersed along the value chain. For instance, if plant A ships one dollar of an intermediate good to plant B, and plant B only add one cent of value added to the product, our measure of fragmentation associated with the final product will be equal to 2 whereas production is mostly concentrated within just one plant.

For this purpose, I construct an alternative measure of fragmentation inspired from the Herfindahl-Hirschman Index (HHI). For each product i, I define H_i by:

$$H_i \;=\; \frac{1}{\sum_n^\infty (v_i^{(n)})^2}$$

where $v_i^{(n)}$ is defined as in Proposition 1 and corresponds to the share of the value added that has gone through n stages. Note that the sum of these shares equal one for each industry: $\sum_n^\infty v_i^{(n)} = 1$, hence $H_i \geq 1$ by construction. This index can be interpreted as a HHI-index of the concentration of value added across production stages. If value added originates from only one stage (i.e. if $v_i^{(n)} = 1$ for a particular stage n), this index equals 1. If the source of value added is rather dispersed across production stages, this index will take larger values.

Table 12: Dispersion of value added along supply chains

Year	1967	1972	1977	1982	1987	1992
H-Index	2.68	2.36	2.38	2.36	2.15	2.18
Weighted	3.01	2.76	2.77	2.56	2.43	2.45

Notes: First row: average across industries of index H_i for each year; second row: average weighted by final consumption.

I calculate this index for all tradable goods (excluding services and petroleum-related industries as in previous tables).[47] I find a very large correlation between this new index H_i and the previous index N_i across industries (taking averages across years): the correlation is above 90% each year. This suggests that both H_i and N_i capture very similar aspects of fragmentation.

Using H_i, I also find that production has become less vertically fragmented. Table 12 shows the average of H_i across industries for each year. The unweighted average of H_i across industries has steadily decreased from 2.68 in 1967 to 2.18 in 1992. The average weighted by final consumption has decreased from 3.01 to 2.45.

6 Conclusion

In this paper, I provide a novel measure of the fragmentation of production reflecting the average number of production stages by industry weighted by the contribution of each stage

[47] In practice, I compute this H-index by summing up to $n = 20$, but this captures more than 99.99% of the value-added.

to value added. A variant of this measure reflects the number of stages between an industry's production and final demand. These indices offer simple structural interpretations. These indices are simple to calculate and only require input-output tables that are generally publicly available. They satisfy interesting aggregation properties: i) the weighted average equals the gross-output-to-value-added ratio in a closed economy; ii) at the industry level, these indices are not likely to be biased by using more aggregated input-output matrices.

The key finding is that US industries have become less vertically fragmented over the past 50 years. The average number of production stages seems to have decreased according to the above fragmentation index computed using the BEA US input-output tables since 1947. This fact is not just limited to a composition effect between services and tradable goods. When I exclude services, I also find a decline in the number of production stages on aggregate. Among manufacturing industries, I find a relatively smaller declines in more specific, R&D- and capital-intensive industries, and larger declines in skill-intensive and financially dependent industries.

Trade and prices do not seem to play an important role in explaining these results. While the commodity-price shock of the mid-70's can explain temporary increases in measured fragmentation, the long-term change in fragmentation does not reflect a systematic change in relative prices of upstream vs. downstream goods along value chains. Also, import penetration is not correlated with an industry's position on the value-added chain across industries, and increases in import penetration in upstream industries are not correlated with changes in the number of production stages for downstream industries.

In order to provide a more intuitive view on the decrease in vertical fragmentation, I examine the evolution of the relative contribution of stages to value added. In particular, I find a large and significant shift of value added towards production stages that are closer to final demand, which generates an overall decrease in weighted-average number of production stages. Half of this shift can be explained by observable industry characteristics such as intensities in the use of capital, skilled labor and advertising services.

While this paper mainly focuses on the vertical fragmentation of production in the US, the measures of fragmentation developed here may have other applications. I illustrate one of those by investigating patterns of US imports depending on the position of industries along value chains and the level of development of the exporting country. In particular, I find that rich countries have a comparative advantage in industries that are closer to final demand and less vertically fragmented.

References

[1] Abowd, J. "Appendix: The NBER Immigration, Trade, and Labor Market Files," in John M. Abowd and Richard B. Freeman, eds. Immigration, Trade, and the Labor Market. Chicago: Univ. of Chicago Press and NBER, 1991, 407-422.

[2] Abraham, Katharine G. and Taylor, Susan K. (1996), "Firm's Use of Outside Contractors: Theory and Evidence," *Journal of Labor Economics*, 14, 394-424.

[3] Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti,"Vertical integration and technology: theory and evidence," *Journal of the European Economic Association*, 85:5 (2010), 1-45.

[4] Acemoglu, Daron, Simon Johnson and Todd Mitton, "Determinants of Vertical Integration: Finance, Contracts and Regulation," *Journal of Finance*, 63:3 (2009), 1251-1290.

[5] M. A. Adelman, 1955. "Concept and Statistical Measurement of Vertical Integration," NBER Chapters, in: Business Concentration and Price Policy, pages 279-328 National Bureau of Economic Research, Inc.

[6] Antràs, Pol "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics*, 118:4 (2003), 1375-1418.

[7] Antràs, Pol "Incomplete Contracts and the Product Cycle," *American Economic Review*, 95:4 (2005), 54-1073.

[8] Antràs, Pol, Davin Chor, Thibault Fally and Russell Hillberry, "Measuring the Upstreamness of Production and Trade Flows" forthcoming in the *American Economic Review* Papers and Proceedings, May 2012.

[9] Arkolakis, C., Costinot, A. and Rodriguez-Clare, A. (2009) "New Trade Models, Same Old Gains?" NBER working paper 15628.

[10] Baldwin, R. E. and A. Venables (2010), "Relocating the Value Chain: Offshoring and Agglomeration in the World Economy", NBER Working paper No. 16111.

[11] Bartelsman, Becker and Gray, 2000, NBER-CES Manufacturing Industry Database.

[12] Bosma, N. S., E. Dietzenbacher, and I. Romero "Using average propagation lengths to identify production chains in the Andalusian economy", *Estudios de EconomÃa Aplicada*, 23 (2005), 405-422.

[13] Brynjolfsson, E., T. Malone, V. Gurbaxani, A. Kambil, "Does Information Technology Lead to Smaller Firms?", *Management Science*, Vol. 40, No. 12, December 1994, pp. 1628-1644.

[14] Campa, Jose M., and Linda Goldberg, "The Evolving External Orientation of Manufacturing Industries: Evidence from Four Countries," *Federal Reserve Bank of New York Economic Policy Review* 4 (1997), 79-99.

[15] Carluccio, Juan and Thibault Fally (2010), "Global Sourcing under Imperfect Capital Markets," CEPR discussion paper no. 7868.

[16] Chenery, Hollis B. and Tsunehiko Watanabe (1958) "International Comparisons of the Structure of Production", *Econometrica*, 26,487-521.

[17] Costinot, A., L. Oldenski, J. Rauch (2011), "Adaptation and the Boundary of Multinational Firms," *The Review of Economics and Statistics*, MIT Press, vol. 93(1), pages 298-308, October.

[18] Costinot, A., Vogel, J. and S. Wang (2011), "An Elementary Theory of Global Supply Chains", manuscript

[19] Dietzenbacher, E. and I. Romero (2007) "Production chains in an interregional framework: identification by means of average propagation lengths". *International Regional Science Review*, 30, 362-383.

[20] Ellison, Glenn, Edward L. Glaeser and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", *American Economic Review*, forthcoming.

[21] Fan, Joseph, and Larry Lang, 2000, "The measurement of relatedness: An application to corporate diversification", *Journal of Business* 73, 629-660.

[22] Feenstra, Robert (1996) "NBER Trade Database, Disk1: U.S. Imports, 1972-1994: Data and Concordances," NBER Working Paper no. 5515, March 1996.

[23] Feenstra, Robert (1998) "Integration of Trade and Disintegration of Production in the Global Economy", April 1998, *Journal of Economic Perspectives*, Fall 1998, 31-50.

[24] Fort, Teresa (2011) "Breaking up is hard to do: how firms fragment production across locations", University of Maryland mimeo.

[25] Guo, J., A. M. Lawson and M. A. Planting (2002), "From Make-Use to Symmetric I-O Tables: An Assessment of Alternative Technology Assumptions," BEA Papers 0021, Bureau of Economic Analysis.

[26] Hall, Robert and Charles Jones "Why do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics*, 114(1999), 83-116.

[27] Hanson, Gordon H. 1995. "Incomplete Contracts, Risk, and Ownership." International Economic Review, 36(2): 341-63.

[28] Helper, Susan (1991) "How much has really changed between US automakers and their suppliers?", *Sloan Management Review*, 32, 15-28.

[29] Hillberry, R., Hummels, D., (2000). "Explaining home bias in consumption: production location, commodity composition and magnification." Purdue University, manuscript.

[30] Hitt, Lorin M. (1999) "Information Technology and Firm Boundaries: Evidence from Panel Data" *Information Systems Research.*

[31] Horowitz, K. and M. Planting (2009), *Concepts and Methods of the U.S. Input-Output Accounts*, Bureau of Economic Analysis.

[32] Hortacsu, A. and C. Syverson (2011), "Why Do Firms Own Production Chains?", University of Chicago, manuscript.

[33] Houseman, Susan, Christopher Kurz, Paul Lengermann, and Benjamin Mandel (2010), "Offshoring Bias in U.S. Manufacturing", *Journal of Economic Perspectives*, Vol. 25, 2, Spring 2011, pp111-132.

[34] Hummels, Davis, Jun Ishii and Kei-Mu Yi, "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics*, 54:1(2001), 75-96.

[35] Johnson, R. and Noguera, G. (2010) "Accounting for Intermediates: Production Sharing and Trade in Value Added", manuscript.

[36] Johnson, R. and Noguera, G. (2012) "Fragmentation and Trade in Value Added Over Four Decades", manuscript.

[37] Jones, Charles (2010) "Intermediate Goods and Weak Links in the Theory of Economic Development", September 2010, Forthcoming in the *American Economic Journal: Macroeconomics.*

[38] Kaufmann, D. Kraay, A. and Mastruzzi, M. (2003), "Governance Matters III: Governance Indicators for 1996-2002," Working Paper No. 3106, World Bank.

[39] Kremer, M. (1993) "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 108:3: 551-576.

[40] Krugman, Paul R. 1996. "Does Third World Growth Hurt First World Prosperity?" Harvard Business Review 72, pp. 113-121.

[41] Lafontaine, Francine and Margaret Slade,"Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*,45:3 (2007), 631-687.

[42] Leontief, V. (1941) *The structure of American economy, 1919-1939: An empirical application of equilibrium analysis.*

[43] Macchiavelo (2009) "Financial Development and Vertical Integration: Theory and Evidence", forthcoming, *Journal of the European Economic Association*

[44] Maddigan, R. (1981). "The Measurement of Vertical Integration." *Review of Economics and Statistics*, 63(3): 328-335.

[45] McLaren, J. (2000) "Globalization and Vertical Structure," *American Economic Review* 90:5, December, pp. 1239-54.

[46] Masten, Scott (1984) "The Organizational Production: Evidence from the Aerospace Industry", *Journal of Law and Economics*, 27, 403-17.

[47] Nunn, Nathan, "Relationship Specificity, Incomplete Contracts and the Pattern of Trade", *Quarterly Journal of Economics*,122:2 (2007), 569-600.

[48] Pierce, Justin and Peter Schott, concording HS system categories to SIC and NAICS industries, NBER WP 15548

[49] Rajan, Raghuram and Luigi Zingales . "Financial Dependence and Growth." *American Economic Review* 88 (1998),559-86.

[50] Rasmussen, P. N. (1956) *Studies in Intersectoral Relations*. Amsterdam: North-Holland.

[51] Rauch, James E.,"Networks Versus Markets in International Trade" *Journal of International Economics* 48:1 (1999), 7-35.

[52] Romalis, John, "Factor Proportions and the Structure of Commodity Trade", *American Economic Review*, Vol. 94(1), March 2004, pp.67-97.

[53] Woodrow E. (1979) "A Note on the Empirical Measurement of Vertical Integration", *The Journal of Industrial Economics*, Vol. 28, No. 1 (Sep., 1979), pp. 105-107.

[54] Yeats, Alexander (2001), "Just How Big is Global Production Sharing?", in Sven W. Arndt and Henryk Kierzkowski (eds.), *Fragmentation: New Production Patterns in the World Economy*, Oxford University Press.

[55] Yi, Kei-Mu (2003) "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy* 111:52-102.

[56] Yi, Kei-Mu (2010) "Can multi-stage production explain the home bias in trade?" *American Economic Review* 100 (1): 364-393.

Mathematical Appendix

Proposition 1: If N_i is defined recursively as in equation (1) and $v_i^{(n)}$ is defined as above, then:

$$N_i = \sum_{n=1}^{\infty} n\, v_i^{(n)}$$

In other words, N_i is the average number of stages to produce good i weighted by the share $v_i^{(n)}$ of value added at each stage n.

Proof: Suppose that N_i is defined by: $N_i = \sum_{n=1}^{\infty} n\, v_i^{(n)}$ where the fraction $v_i^{(n)}$ is defined recursively by $v_i^{(n+1)} = \sum_j \mu_{ij} v_j^{(n)}$, with $v_i^{(1)} = \frac{V_i}{Y_i}$. We need to show that N_i verifies the recursive definition of equation (1).

First, note that $\sum_{n=1}^{\infty} v_i^{(n)} = 1$. To see this point, note that

$$\begin{aligned} 1 - \sum_{n=1}^{\infty} v_i^{(n)} &= 1 - v_j^{(1)} - \sum_{n=2}^{\infty} \sum_j \mu_{ij} v_j^{(n)} \\ &= 1 - \frac{V_i}{Y_i} - \sum_{n=1}^{\infty} \sum_j \mu_{ij} v_j^{(n)} \\ &= \sum_j \mu_{ij} - \sum_j \mu_{ij} \left(\sum_{n=1}^{\infty} v_j^{(n)} \right) \\ &= \sum_j \mu_{ij} \left(1 - \sum_{n=1}^{\infty} v_j^{(n)} \right) \end{aligned}$$

Assuming that the identity matrix minus the input-output matrix is invertible (see foonote 12), $1 - \sum_{n=1}^{\infty} v_j^{(n)} = 0$ is the only solution of the system of equation $x_i = \sum_j \mu_{ij} x_j$.

Using the above definition of N_i, we obtain successively:

$$\begin{aligned} N_i &= \sum_{n=1}^{\infty} n\, v_i^{(n)} \\ &= \sum_{n=0}^{\infty} (1+n)\, v_i^{(n+1)} \\ &= \sum_{n=0}^{\infty} v_i^{(n+1)} + \sum_{n=0}^{\infty} n\, v_i^{(n+1)} \\ &= \sum_{n=1}^{\infty} v_i^{(n)} + \sum_{n=1}^{\infty} n \sum_j \mu_{ij} v_j^{(n)} \end{aligned}$$

Then, using the recursive definition of $v_i^{(n)}$ and using the fact that $\sum_{n=1}^{\infty} v_i^{(n)} = 1$, this becomes:

$$\begin{aligned} N_i &= 1 + \sum_{n=1}^{\infty} n \sum_j \mu_{ij} v_j^{(n)} \\ &= 1 + \sum_j \mu_{ij} \sum_{n=1}^{\infty} n\, v_j^{(n)} \\ &= 1 + \sum_j \mu_{ij} N_j \end{aligned}$$

which corresponds to equation (1).

Section 2.3: Structural interpretations of N: details on examples ii, iii)

Example ii): If the production function for product i has constant returns to scale in all inputs j plus labor, the unit cost is a homogenous function of degree one in prices of each input and labor, and is inversely proportional to productivity. Keeping wages constant, we obtain that the relative change in prices satisfies:

$$\hat{P}_i = -\hat{Z} + \sum_j \mu_{ij} \hat{P}_j$$

where hats denote relative changes and μ_{ij} is the share of input j in total cost of production for i. We can see that $\frac{-\hat{P}_i}{\hat{Z}}$ satisfies the same reccursive definition as N_i. Hence: $\hat{P}_i = -\hat{Z} N_i$.

Concerning welfare, the result shown in equation (3) is obtained by considering the expenditure function and the envelop theorem. In equilibrium, quantities of goods for final consumption maximize utility given the set of prices. Hence the change in expenditures generated by a change in prices is given by:

$$\hat{e} = \sum_i \alpha_i \hat{P}_i$$

where $\alpha_i = \frac{C_i}{\sum_j C_j}$ is the share of good i in final consumption. Using the previous results on price changes, we obtain the formula in the text.

Example iii): In a first case, let us consider an economy with J industries, in perfect competition, characterized by the following equations:

$$\begin{array}{rcl} Q_i & = & Q_i^F + \sum_j Q_{ji}^M \\ Q_i & = & Z A_i \,.\, \prod_{j=1}^J (Q_{ij}^M)^{\mu_{ij}} \,.\, L_i^{1-\sum_j \mu_{ij}} \\ U & = & \prod_{i=1}^J (Q_i^F)^{\alpha_i} \\ \bar{L} & = & \sum_i L_i \end{array}$$

where U defines preferences in terms of consumption of goods i, with the sum $\sum_i \alpha_i$ normalized to unity; Q_i^F referes to the quantity of final goods i whereas Q_{ij}^M refers to the quantity of goods j used an inputs for the production of good i. In addition, we normalize wages (and nominal income) to unity. Nominal GDP is therefore equal to population $\bar{L}$.

In this framework, final consumption (in value) is a constant fraction of total income: $C_i \equiv P_i Q_i^F = \alpha_i \bar{L}$. Intermediate demand (in value) is also a constant fraction of downstream production Y_i (in value): $Y_{ij} \equiv P_j Q_{ij}^M = \mu_{ij} Y_i$. Hence, the value of production in sector j satisfies: $Y_j = \alpha_j \bar{L} + \sum_i \mu_{ij} Y_i$. Taking all sectors, this system of equation determines sectoral production as a function of total income and parameters α_i and μ_{ij}. In particular, the value of production does not depend on Z.

This framework is a special case of example ii). The result on prices applies: $\frac{\partial P_i}{\partial Z} = -N_i$. Since the value of production does not depend on Z, quantities should satisfy:

$$\frac{\partial Q_i}{\partial Z} = N_i$$

Now, suppose instead that we have the following Leontief production function:

$$Q_i = Z. \min_j \left\{ \frac{Q_{ij}}{\alpha_{ij}}, \frac{L_i}{\alpha_{iL}} \right\}$$

with Q_i denoting the production (in quantity) of good i, Q_{ij} is the quantity of input j used for the production of i, Z reflects productivity, L_i is the amount of labor for the production of good i, α_{ij} and α_{iL} are parameters.

In the spirit of the O-ring theory (Kremer, 1993) and Costinot Vogel and Wang (2012), we can interprete Z as being determined by the probability that no mistake arise, assuming that mistakes potentially arise at each stage of production (i.e. for the production of each good i, whether it is a final or intermediate good).

Suppose also that utility is a Leontief function of final consumption Q_i^F:

$$U = \min_i \left\{ \frac{Q_i^F}{\alpha_{iF}} \right\}$$

In this framework, we obtain that $Q_i^F = \alpha_{iF} U$ where U is the level of utility attained at equilibrium. Total production quantities of good i satisfies:

$$Q_i = \alpha_{iF} U + \sum_j \alpha j i Q_j / Z$$

Given a change in productivity $\hat{Z}$ (generating a change in utility $\hat{U}$), the effect on production is:

$$\hat{Q}_i = (1 - \sum_j \varphi_{ij}) \hat{U} + \sum_j \varphi_{ji} (\hat{Q}_j - \hat{Z})$$

where $\varphi_{ij} = \frac{\alpha j i Q_j / Z}{Q_i}$ denotes the share of production of good i absorbed as intermediate goods for industry j.

From the previous equation, we obtain that:

$$\hat{Q}_i - \hat{Z} - \hat{U} = -\hat{Z} + \sum_j \varphi_{ji} (\hat{Q}_j - \hat{Z} - \hat{U})$$

We can see that $\frac{\hat{Q}_i - \hat{Z} - \hat{U}}{-\hat{Z}}$ satisfies the same recursive equation defining D_i, the index of "distance to final demand", and thus should be equal to D_i. Therefore:

$$\hat{Q}_i - \hat{Z} - \hat{U} = -\hat{Z} D_i$$

Taking the diifference between any two industries, we obtain:

$$\hat{Q}_i - \hat{Q}_j = -\hat{Z}(D_i - D_j)$$

which corresponds to the result shown in the main text.

Proposition 2: In a closed economy, the aggregate measure of fragmentation equals the gross output to value added ratio: $\frac{\sum_i C_i N_i}{\sum_i C_i} = \frac{\sum_i Y_i}{\sum_i V_i}$ (part 1) and $\frac{\sum_i V_i D_i}{\sum_i V_i} = \frac{\sum_i Y_i}{\sum_i V_i}$ (part 2).

Proof: We use two equalities: the definition of measure of fragmentation $N_i = 1 + \sum_j \mu_{ij} N_j$, and the link between final consumption, intermediate demand and production (in a closed economy): $C_i = Y_i - \sum_j \mu_{ji} Y_j$. We obtain:

$$\begin{aligned}
\sum_i C_i N_i &= \sum_i \Big(Y_i - \sum_j \mu_{ji} Y_j\Big) N_i \\
&= \sum_i Y_i N_i - \sum_{i,j} \mu_{ji} Y_j N_i \\
&= \sum_i Y_i N_i - \sum_{i,j} \mu_{ij} Y_i N_j \\
&= \sum_i Y_i N_i - \sum_i Y_i \Big(\sum_j \mu_{ij} N_j\Big) \\
&= \sum_i Y_i N_i - \sum_i Y_i (N_i - 1) \\
&= \sum_i Y_i
\end{aligned}$$

Similarly, for the other measure D_i (part 2), we obtain: $\sum_i V_i D_i = \sum_i Y_i$ by using the definition $D_i = 1 + \sum_j \varphi_{ij} D_j$ and the equality $V_i = Y_i - \sum_j \mu_{ij} Y_i = Y_i - \sum_j \varphi_{ji} Y_j$.

Finally, notice that the sum of final demand $\sum_i C_i$ equals the sum of value added $\sum_i V_i$.

Proposition 3: In an open economy:

$$\begin{aligned}
\frac{\sum_i C_i N_i}{\sum_i C_i} &= \bar{N} + \frac{\sum_i (M_i - X_i)(N_i - \bar{N})}{\sum_i C_i} \\
\frac{\sum_i V_i D_i}{\sum_i V_i} &= \bar{N} - \frac{\sum_i (M_i - X_i)(D_i - 1)}{\sum_i V_i}
\end{aligned}$$

Where $\bar{N}$ denotes the ratio of gross output to value added $\frac{\sum_i Y_i}{\sum_i V_i}$.

Proof: In an open economy, final consumption satisfies $C_i = Y_i - \sum_j \mu_{ji} Y_j + M_i - X_i$. Let's define $F_i \equiv Y_i - \sum_j \mu_{ji} Y_j$. We deduce that $C_i = F_i + (M_i - X_i)$. Following the same path as in the proof of Proposition 2, we can show that $\sum_i F_i N_i = \sum_i Y_i$. Moreover, we can verify that $\sum_i F_i$ equals total value added $\sum_i V_i$ and thus: $\bar{N} \sum_i F_i = \sum_i Y_i$.

Using these three equalities above, we obtain:

$$\begin{aligned}
\sum_i C_i (N_i - \bar{N}) &= \sum_i F_i (N_i - \bar{N}) + \sum_i (M_i - X_i)(N_i - \bar{N}) \\
&= \sum_i Y_i - \bar{N} \sum_i F_i + \sum_i (M_i - X_i)(N_i - \bar{N}) \\
&= \sum_i (M_i - X_i)(N_i - \bar{N})
\end{aligned}$$

After dividing by total consumption, this provides the first equality of Proposition 3.

Turning to the second equality, we use the following relationship between φ_{ij} and input-output coefficients in open economy: $\varphi_{ij} = \frac{Y_j}{Y_i+M_i-X_i}.\mu_{ji}$. We obtain:

$$\begin{aligned}
\sum_i V_i D_i &= \sum_i \Big(Y_i - \sum_i \mu_{ij} Y_i\Big) D_i \\
&= \sum_i Y_i D_i - \sum_{i,j} \mu_{ij} Y_i D_i \\
&= \sum_i Y_i D_i - \sum_{i,j} \mu_{ji} Y_j D_j \\
&= \sum_i Y_i D_i - \sum_{i,j} (Y_i + M_i - X_i)\, \varphi_{ij} D_j \\
&= \sum_i Y_i D_i - \sum_i (Y_i + M_i - X_i)\, (D_i - 1) \\
&= \sum_i Y_i - \sum_i (M_i - X_i)(D_i - 1)
\end{aligned}$$

After dividing by total value added $\sum_i V_i$ and using the definition of $\bar{N} = \sum_i Y_i / \sum_i V_i$, we get the second equality of Proposition 3.

Correspondance with the VAX ratio:

Johnson and Noguera (2010) construct a global input-output matrix A relating the use of input by destination and source country. They use this global IO matrix to derive output as a function of absorption in each country. Using their notation (with i and j being country subscripts):

$$y_j = (I - A)^{-1} f_j$$

where f_j is the vector of final goods to be purchased by final consumers in country j. Gross output y_j is the sum of both domestic sales and gross exports.

Lemma 1: Gross trade x_j of goods absorbed in final destination j can be expressed as:

$$x_j = (I - \widetilde{A})^{-1} \tilde{f}_j$$

where $\widetilde{A}$ is an input-output matrix for trade flows, i.e. describing import requirement for each dollar of gross exports, and $\tilde{f}$ is a vector of export to their final destination.

Proof: Let us define A^D the domestic component of the global IO matrix (i.e. the block-diagonal matrix with blocks A_{ii} describing the use of inputs from country i by industries in i) and let us define A^M the IO import matrix for the use of inputs from other countries: $A^M = A - A^D$.

Similarly, let us denote by f^D the vector of final goods consumption from domestic sources and by f^M the vector of imported final goods: $f = f_D + f_M$. Let us also denote x the vector of gross exports and h the vector of gross domestic shipments. We can obtain the following accounting equality:

$$\begin{aligned}
x &= A^M (x + h) + f^M \\
h &= A^D (x + h) + f^D
\end{aligned}$$

The first term in the first equation corresponds to imported intermediate goods and the second term reflects imported final goods, while the second equation reflects the purchase of intermediate and final goods from domestic sources. Solving for h, we obtain that:

$$h = (I - A^D)^{-1} A^D x + (I - A^D)^{-1} f^D$$

Plugging in h back into the expression for x, we obtain successively:

$$\begin{aligned}
x &= A^M x + A^M h + f^M \\
&= A^M x + A^M (I - A^D)^{-1} A^D x + A^M (I - A^D)^{-1} f^D + f^M \\
&= A^M \left(I + (I - A^D)^{-1} A^D \right) x + A^M (I - A^D)^{-1} f^D + f^M \\
&= A^M (I - A^D)^{-1} x + A^M (I - A^D)^{-1} f^D + f^M \\
&= \widetilde{A} x + \tilde{f}
\end{aligned}$$

where $\widetilde{A} \equiv A^M (I - A^D)^{-1}$ and $\tilde{f} \equiv f^M + \widetilde{A} f^D$. In words, $\widetilde{A}$ is the matrix of import directly required for each dollar of export x and indirectly for domestic output generated by this export through domestic requirements. We can then solve directly for trade:

$$x = (I - \widetilde{A})^{-1} \tilde{f}$$

As in Johnson and Noguera (2010), we can also split trade and output depending on the final destination country j as:

$$x_j = (I - \widetilde{A})^{-1} \tilde{f}_j$$

Lemma 2: Denoting by $m_j \equiv \widetilde{A} x_j$ the vector indirected imports generated by exports x_j, and by $\mathbf{1}$ the column vector (by country and sector), the total value-added content of trade from i to j (summed across all sectors s) can be obtained as:

$$\sum_s va_{ij}(s) = \sum_s x_{ij}(s) - \sum_s m_{ij}(s)$$

where $x_{ij}(s)$ is the value of trade from i to final destination j in sector s minus the sum of import requirements $m_{ij}(s)$ associated with these exports (summing across inputs).

Proof: Direct inputs required for output y_{ij} (output in country i for final absorption in country j) are given by $(I - A_i) y_{ij}$ where A_i is the global IO table component for country i. Output y_{ij} is the sum of exports x_{ij} and domestic output h_{ij} destined to final consumption in country j.

Note that if $i \neq j$, then $h_{ij} = (I - A^D)^{-1} A^D x_{ij}$ and does not depend on final goods purchased from domestic sources in i.

Combining these results, we can obtain the vector of output y_{ij} (production in country i destined to be absorbed in country j) minus the vector of intermediate goods as a difference between the vector of export from i (for final absorption in j) and the vector of imported intermediate goods:

$$(I - A_i) y_{ij} = (I - A_i^D - A_i^M)(x_{ij} + h_{ij})$$

$$
\begin{aligned}
&= (I - A_i^D - A_i^M)\left[x_{ij} + (I - A_i^D)^{-1} A_i^D . x_{ij}\right] \\
&= (I - A_i^D - A_i^M)\left[I + (I - A_i^D)^{-1} A_i^D\right] . x_{ij} \\
&= (I - A_i^D - A_i^M)(I - A_i^D)^{-1} x_{ij} \\
&= \left[I - A_i^M (I - A_i^D)^{-1}\right] x_{ij} \\
&= \left[I - \widetilde{A}_i\right] x_{ij}
\end{aligned}
$$

Then, by taking the column-sum of these vectors, the left-hand side gives the value-added content of trade from i to j as defined by Johnson and Noguera (2010): total output by country i to be absorbed in j minus total intermediate use by country i for the production of goods to be absorbed in j. Taking the column-sum of the right-hand side, we obtain total gross trade from country i to be absorbed in j minus the total use of imported intermediate goods related to these exports.

Hence, it is equivalent to measure the value-added content of trade by just looking at exports x_{ij} and the related use of imported goods using the IO matrix $\widetilde{A} = A_i^M (I - A_i^D)^{-1}$.

Interpretation of the VAX ratio: Using these two lemmas we can deduce that:

- Exports can be derived from a purely international IO matrix $\widetilde{A} \equiv A^M (I - A^D)^{-1}$ and the vector of trade to be absorbed within the destination country $\tilde{f} \equiv f^M + \widetilde{A} f^D$
- The value-added content of trade (summed across sectors) can be simply obtained from the export flows and the international IO matrix $\widetilde{A}$.

Hence to draw a parallel with Proposition 2, we can treat the world as one closed economy where only international shipments are observed, where both the value-added content of trade and the index of fragmentation can be constructed from the matrix $\widetilde{A}$ relating observed trade flows. The equivalent of an economy's gross output would be the total gross trade in this case, while total value added (GDP) would now correspond to the total value-added content of trade. Using Proposition 2, we htus obtain that the weighted number of border crossings embodied in trade flows forthe world economy (weighted by value added at each "stage" i.e. each country) equals the VAX ratio.

Proposition 4: If $(\int_{\Omega_{ij}} y(\omega)N(\omega)d\omega)/(\int_{\Omega_{ij}} y(\omega)d\omega)$ does not depend on the downstream industry j, for all $j \neq i$ or $j = F$, then:

$$N_i = \frac{\int_{\Omega_{iF}} y(\omega)N(\omega)d\omega}{\int_{\Omega_{iF}} y(\omega)d\omega}$$

is the solution to equation (1) which characterizes index N_i at the industry level.

Proof: If $N(\omega)$ denotes the average number of stages required to produce variety ω (same definition as for the industry-level index but at the variety- or plant-level), then $N(\omega)$ equals 1 plus the weighted average of the index for inputs required to produce variety ω. Aggregating over all varieties $\omega \in \Omega_i$ in industry i, we obtain:

$$\int_{\Omega_i} y(\omega)N(\omega)d\omega = \int_{\Omega_i} y(\omega)d\omega + \sum_j \int_{\Omega_{ji}} y(\omega')N(\omega')d\omega'$$

where ω' refers to varieties of inputs, and where Ω_{ji} refers to the set of input varieties ω' in industry j that enter the production of varieties in industry i. Note that the first term of the righ-hand side corresponds to output in industry i:

$$\int_{\Omega_i} y(\omega)N(\omega)d\omega = Y_i + \sum_j \int_{\Omega_{ji}} y(\omega')N(\omega')d\omega'$$

If we exclude varieties in Ω_i that are used as inputs for industry i (i.e. only consider varieties $\omega \in \Omega_i \backslash \Omega_{ii}$), we have then:

$$\int_{\Omega_i \backslash \Omega_{ii}} y(\omega)N(\omega)d\omega = Y_i + \sum_{j \neq i} \int_{\Omega_{ji}} y(\omega')N(\omega')d\omega'$$

Let us denote by $\tilde{N}_i = \frac{\int_{\Omega_{iF}} y(\omega)N(\omega)d\omega}{\int_{\Omega_{iF}} y(\omega)d\omega}$ the "true" average index across varieties in industry i weighted by final demand. If the conditions enounced in Proposition 4 are satisfied, then the set Ω_{iF} in the previous definition can be replaced by the set $\Omega_i \backslash \Omega_{ii}$ that includes all varieties not sold as input for industry i. By using again the conditions enounced in Proposition 4 (between lines 3 and 4 in the following equalities), we obtain successively:

$$\begin{aligned}
\tilde{N}_i &= \frac{\int_{\Omega_i \backslash \Omega_{ii}} y(\omega)N(\omega)d\omega}{\int_{\Omega_i \backslash \Omega_{ii}} y(\omega)d\omega} \\
&= \frac{\int_{\Omega_i \backslash \Omega_{ii}} y(\omega)N(\omega)d\omega}{Y_i - \mu_{ii} Y_i} \\
&= \frac{Y_i + \sum_{j \neq i} \int_{\Omega_{ji}} y(\omega)N(\omega)d\omega}{(1-\mu_{ii})Y_i} \\
&= \frac{Y_i + \sum_{j \neq i} \tilde{N}_j \int_{\Omega_{ji}} y(\omega)d\omega}{(1-\mu_{ii})Y_i} \\
&= \frac{Y_i + \sum_{j \neq i} \tilde{N}_j \mu_{ij} Y_i}{(1-\mu_{ii})Y_i} \\
&= \frac{1 + \sum_{j \neq i} \mu_{ij} \tilde{N}_j}{1-\mu_{ii}}
\end{aligned}$$

After rearranging, we find:

$$\tilde{N}_i = 1 + \sum_j \mu_{ij} \tilde{N}_j$$

This shows that $\tilde{N}_i = N_i$ if that conditions in Proposition 4 are satisfied.

Proof of Proposition 5: The proof follows the same logic as for Proposition 4.

Partial-aggregation properties

Propositions 4 and 5 can also be used to examine partial aggregation properties: what happens when two industries are merged together in the industry classification?

Let us define N_i as in equation (1) for each industry i. Now suppose that industries "1" and "2" are aggregated into industry "a". The aggregated input-output coefficients satisfy:

$$\begin{aligned} \mu_{aa} &= [Y_1\mu_{11} + Y_1\mu_{12} + Y_2\mu_{21} + Y_2\mu_{22}]/[Y_1 + Y_2] \\ \mu_{aj} &= [Y_1\mu_{1j} + Y_2\mu_{2j}]/[Y_1 + Y_2] \\ \mu_{ia} &= \mu_{i1} + \mu_{i2} \end{aligned}$$

Coefficients μ_{ij} remain the same for any $i, j \notin \{1, 2, a\}$.

Using these aggregated input output coefficients, we can define an alternative staging index $\tilde{N}_i$ where $i = a$ or $i \notin \{1, 2, a\}$. Applying Propositions 4 and 5, we obtain:

Corollary 6 *$\tilde{N}_i = N_i$ and $\tilde{N}_a = [N_1C_1 + N_2C_2]/[C_1+C_2]$ if one of these conditions is satisfied:*
i) $N_1 = N_2$
ii) $\mu_{i1}/\mu_{i2} = C_1/C_2$ across all other industries $i \notin \{1, 2, a\}$.
where C_i is defined by the final use of good i.

Intuitively, this result states that aggregation generates an unbiased measure of fragmentation either if there is no heterogeneity within an industry (all industries have the same number of production stages) or if other industries use the different sub-industries (within the aggregated industry) in the same proportions.[48] For instance, if industry "a" is composed of two sub-industries 1 (downstream) and 2 (upstream) but other industries only use products from 1, then the property ii) above is satisfied and the index constructed with aggregated IO matrix corresponds to the weighted average of the true index. This could apply to airplanes and airplane components: aggregating these two industries doesn't generate a bias if the airplane industry is the only one using airplane components. Condition ii) could also apply to the aggregation of tires and car industries if other industries (and final consumers) always use tires and cars in the same proportions.

Similar aggregation properties are found for the number of stages between production and final demand (distance to final demand). Under certain conditions, this index is stable by partial aggregation within the same aggregated industry.

Corollary 7 *$\tilde{D}_i = \tilde{D}_i$ and $\tilde{D}_a = [D_1V_1 + D_2V_2]/[V_1+V_2]$ if one of these conditions is satisfied:*
i) $D_1 = D_2$
ii) $\varphi_{i,1}/\varphi_{i,2} = V_1/V_2$ across all other industries $i \notin \{1, 2, a\}$.
where V_i is the value-added of industry i.

[48] Notice that property ii) is implicitly satisfied when we aggregate across all industries.

Data Appendix

Treatement of "make" and "use" tables

"Make" and "use" industry-by-commodity tables are available from 1972 onward. I combine information from these two tables to construct a commodity-by-commodity table and estimate the amount of commodity j (input) used to produce commodity i (output).[49]

"Use" tables describe the value of purchases u_{kj} of input j by industry k, while "make" tables describe the value of production m_{ki} of output i for each industry k. I construct commodity-by-commodity input-output ceofficients μ_{ij} by taking the average share of input j in production of industry k weighted by the contribution of industry k to the production of output i:

$$\mu_{ij} = \sum_k \left[\frac{m_{ki}}{\sum_{k'} m_{k'i}} \frac{u_{kj}}{\sum_{j'} m_{kj'}} \right]$$

where $\sum_{k'} m_{k'i} = Y_i$ corresponds to total production of output i and $\sum_{i'} m_{ki'}$ corresponds to total production of industry k – this method is based on the "industry-technology assumption" (see Guo *et al.*, 2002).

Note that this way of constructing intput-output coefficients μ_{ij} is consistent with the construction of coefficients φ_{ij} measuring the fraction of output i used for production of output j if they are defined as:

$$\varphi_{ij} = \sum_k \left[\frac{u_{ki}}{(\sum_{k'} u_{k'i} + u_{Fi})} \frac{m_{kj}}{\sum_{j'} m_{kj'}} \right]$$

where $\sum_{k'} u_{k'i} + u_{Fi}$ includes the use of product i by all industries plus final demand. In an open economy, this corresponds to total absorption $Y_i + M_i - X_i$ i.e. domestic production plus net imports, as discussed in section 2.2. We can verify that:

$$\varphi_{ij} = \frac{Y_j \mu_{ji}}{Y_i + M_i - X_i}$$

Note also that this way to construct input-output coefficients is consistent with aggregation properties discussed in the text. In particular, we find that total value added $\sum_i V_i$, where value-added is defined by $V_i = (1 - \sum_j \varphi_{ij})Y_i$ as in the text, equals total production $\sum_{k,i} m_{ki}$ minus total use of inputs $\sum_{k,j} u_{k,i}$.

Treatment of "non-comparable" and "transferred" imports

In the 1972 table and after, the sum of each column of the use table provides production for each industry (sum of value-added and intermediate purchases). Intermediate goods imports are reported as part of input usage u_{kj} as described above. Total imports and exports by product are also reported in two of the last columns.

A small share of imports, however, are reported as "non-comparable" and correspond to a distinct row in the list of inputs. These non-comparable imports correspond to products that are different from any product produced in the US such as coffee and cocoa beans. Since I need

[49] The 1967 input-output table is treated as a commodity-by-commodity table. I obtain very similar results by extrapolating a "make" table from other years to adjust input-output coefficients.

to have an estimate of the number of production stages necessary to produce all inputs (even if thoses goods are imported), I make changes in the data for two industries: I assume that all non-comparable imports by the coffee-roasting industry (industry 142800) and the chocolate industry (industry 142002) correspond to imports of coffee and cocoa beans respectively and are comparable to "tree nuts" (commodity 020401). These two changes reduce the amount of non-comparable imports of intermediate goods by more than half and the remaining non-comparable account for less than half a percent of total production value (and are thus dropped).[50]

The 1967 input-output table has a different treatment for imports and a few other corrections are needed. Imports are classified in two categories: "non-comparable" imports as described above and "transferred" imports. "Transferred" imports are recorded in two places and would be double-counted if not carefully taken into account. In particular, the column-sum of the 1967 I-O table gives the sum of domestic production plus "transferred" imports classified in the same product category. Hence we need to substract "transferred" imports to obtain domestic output. Note however that "transferred" imports of intermediate goods also appear in input-output coefficient for each input category.[51] In terms of final consumption, some imports destined for final consumption are classified as "non-comparable" imports (while being actually quite comparable) and may account for a large share of absorption in these industries: for instance, most imports of cars are missing in the 1967 consumption data. I thus use import data from the NBER trade database (Feenstra, 1996) to impute the amount of imports for consumption.

Other data sources

Industry characteristics are obtained from various sources. I use the NBER-CES database (Bartelsman, Becker and Gray, 2000) to construct an index of capital intensity (value of capital stock over wages), skill intensity (share of non-production-worker wages in total wages) and productivity. The NBER-CES database is available for manufacturing industries in the SIC 1987 classification and includes all benchmark years between 1967 and 1992. Data on R&D intensity are obtained from the National Science Foundation and is available from 1982. An index of product specificity has been developed by Rauch (1999). Rauch (1999) classifies goods into three categories: goods traded on integrated markets, goods with reference prices and other goods classified as specific. I simply use a dummy being equal to one when goods are specific.[52] I also use an index of dependence in external finance following Rajan and Zingales (1998) methodology. Concentration indices are obtained from the Census, which provides the Herfindahl index and the share of production by the 4 largest companies for each 1987 SIC manufacturing industry. An index of advertising intensity for manufacturing industries is constructed using the input-output coefficient for advertising-related services in 1992. Note

[50]Note that the 1992 table significantly reduced the "non-comparable imports" category by associating these imports will other classified commodities. In particular, the coffee-roasting and chocolate industries in 1992 exhibit large uses of inputs classified as "tree nuts" instead of non-comparable imports, which is consistent with the changes made on earlier tables.

[51]For instance, imports of crude petroleum to be used by the petroleum refinenement industry appear twice: in the row for transferred imports in the column of crude pretroleum, and also in the row for crude petroleum in the column for petroleum refinement.

[52]Rauch classification follows SITC revision 2. My final index is then the fraction of goods within each 1987 being categorized as specific in the SITC classification.

finally that the main results presented throughout the paper are robust to dropping extreme observations for each variable (extreme percentiles).

Table 13: Mean and standard deviation of industry variables

Variable	Mean	Std. Dev.
Number of stages	1.684	0.251
Stages to final demand	1.574	0.672
Specificity	0.744	0.386
R&D intensity	1.944	1.942
Capital intensity	1.124	0.615
Skill intensity	0.357	0.112
Advertising intensity	1.479	2.119
Productivity	0.978	0.113
Productivity growth	0.024	0.081
Financial Dependence	0.166	1.490
Top 4 share	40.36	19.84
Import penetration	0.096	0.110

Notes: Mean and standard deviation of the main variables across industries.

US trade data are available in the 1972 SIC classification (after 1958) and 1987 SIC classification (after 1972) for manufacturing industries from Feenstra (1996). For section 4.5 (on imports across source countries) I complement the trade data by source country with Penn World Table data on GDP per capita (average between 1990 and 1994), physical distance (CEPII) and data on endowments in capital and skilled labor from Hall and Jones (1999).