On the fragmentation of production in the US

Thibault Fally*
University of Colorado-Boulder

October 2011

Abstract

What is the average number of production stages for the US? Is production more fragmented now than decades ago? 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 on tradable goods. At least half of this decrease corresponds to a shift of value-added towards industries that are closer to final demand, while early stages contribute less to the value of final goods. Also, the production of more complex goods appears to be relatively less vertically fragmented. I show that international trade has provided new opportunities to fragment production and 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 show that goods involving fewer production stages and 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, L16, L23, O14.

*I specially thank Russell Hillberry for very helpful suggestions. I also thank Pol Antras, Jeff Bergstrand, Yongmin Chen, Arnaud Costinot, David Hummels, Wolfgang Keller, Kalina Manova, Jim Markusen, Keith Maskus, Emanuel Ornelas, Scott Savage, Robert Staiger and seminar participants at Stanford University, Bank of France, Graduate Institute, European Trade Study Group, London School of Economics and the University of Colorado-Boulder for helpful comments, and Ben Li for great research assistance.
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 tends to be more fragmented across countries (Hummels, Ishii and Yi, 2001), 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? Is production more fragmented now than decades ago? How many plants are sequentially involved in production chains (henceforth referred to as vertical fragmentation) matters for several key issues in trade and other economic aspects. As trade costs decline, gains from trade are magnified when production is or can be fragmented: not only consumer can 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 to fragment production constitute one of the main sources of gains from agglomeration according to Marshall and recently confirmed by Ellison et al. (forthcoming). Economic development has also put a traditional emphasis on the role of vertical linkages (Hirshman, 1955) and more recently the "O-ring" theory (Kremer 1993, Jones 2010).

In this paper, I provide new quantitative analysis on the average length of production chains, its evolution over time, and its determinants. I develop a simple measure to document the length of production chains using input-output tables. For the aggregate economy, it corresponds to a weighted average of the number of plants sequentially involved in production chains, where the weight is the value that has been added at each stage.[1] In closed economy, it equals the ratio of total gross output to value added. At the industry level, I construct two separate measures to reflect: i) the number of stages required for production; ii) the number of stages between production and final consumption.

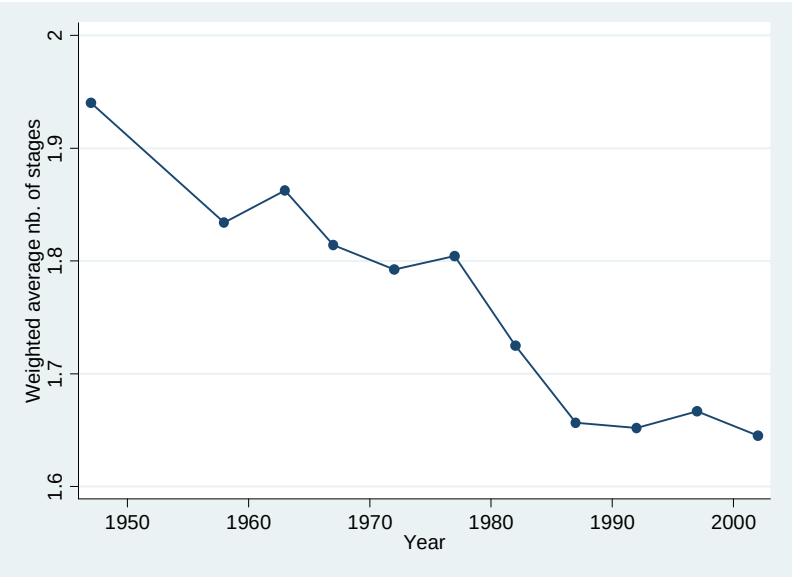
I calculate these measures of vertical fragmentation for the US using benchmark input-output tables from the Bureau of Economic Analysis for a period covering 1947 to 2002. Without the need of plant-level data,[2] I can compute the average number of sequential stages (plants) in production chains weighted by value added at each stage. I find that production chains are short on average or, equivalently, that most of the value added comes from later stages: the weighted number of stages is smaller than 2 on average for the aggregate economy.

More surprisingly, I find that the weighted number of production stages has been decreasing

[1] Here, stages correspond to plants. This definition may differ from a task-level approach where each task is defined as one stage.

[2] Computing an unweighted measure of the length of production chains at the plant level would require matched supplier-buyer transaction data which are not available. However, by just using industry input-output tables it is possible to compute a measure of fragmentation across plants weighted by value added by each plant, thus bypassing the need of plant-level data.

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



by more than 10% over the past 50 years. In part, this decrease can be partly explained by the increasing share of services in total production: services now account for 70% of US GDP and services generally require fewer production stages than manufacturing industries. However I find that the number of production stages has also decreased for primary and manufacturing industries ("tradable" goods). Figure 1 plots the evolution of the weighted number of stages aggregated over all tradable goods excluding petroleum.

Using more disaggregated data between 1967 and 1992, I examine the determinants of the number of production stages across 311 industries over time. In particular, I can decompose the evolution of the aggregate measure of fragmentation into industry compositions effects and within-industry effects. If we exclude services, I do not find that consumption has shifted towards less fragmented industries. However, I find a large and significant shift of production towards industries that are closer to final demand. In other words, early stages contribute less to the final value of production, whereas more value is added at later stages. This shift can explain about one half of the overall decrease in the measure of fragmentation.

By looking at fragmentation across industries, I find that vertical fragmentation is negatively correlated with product specificity, R&D intensity, skill intensity and dependence in external finance. The number of sequential stages however 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). These findings seem surprising if one expects high-tech industries to have more complex sourcing strategies, but these results seem consistent with incomplete

contract theories of the firm applied to the study of the fragmentation of production across plants. Furthermore, R&D-intensive industries have become relatively less fragmented over time. Industry characteristics can also explain a large part of the shift in value added towards final stages. In particular, industries that are more intensive in advertising and less intensive in capital have experienced a larger growth rate, which explains a significant part of this shift.[3]

I specifically investigate the role of trade. The decrease in the overall fragmentation of production remains puzzling regarding the reorganization of supply chains across borders ("slicing up the value chain" using Krugman 1996 terminology). We expect that the large decline in transport costs over the past decades has provided new opportunities. I find indeed that increased import penetration induced an increase in vertical fragmentation, showing that foreign outsourcing is not just a substitute to domestic outsourcing.

I perform various checks to confirm the robustness of this measure of fragmentation. One may be concerned that the input-output classification system is not detailed enough and far from reflecting the transformation of products along production chains (even if US input-output matrices are available at a very disaggregated level since 1967). It is therefore essential to examine how the constructed measure of fragmentation can be biased when using imperfectly disaggregated tables. I do it both from a theoretical and empirical perspective. First, I show that having an aggregated input-output table does not bias the aggregate measure of fragmentation for a closed economy. Second, I examine the conditions that are required such that partial aggregation of the input-output table does not yield any bias at the industry level. Then, I verify empirically that the measure of fragmentation is not significantly biased when it is constructed from an input-output table artificially aggregated over broader industries (e.g. at the 2- or 3-digit level instead of 6-digit level).

I also conduct further robustness checks to verify that changes in this measure of fragmentation are not driven by price effects. Rapid changes in oil prices may explain short-term changes in observed fragmentation by magnifying the weight put on early stages (e.g. oil extraction). Over the long term, however, changes in relative prices of commodities and intermediate goods cannot explain the overall observed decline. Another concern is that input-output matrices are computed in producer prices. After reincorporating trade and retail margins, I find a similar evolution of vertical fragmentation. Finally, in order to better reflect how production chains are "sliced up", I construct an alternative index that is (inversely) related to the concentration of value added along stages of production chains. Using this alternative index leads to the same conclusion: production has become less vertically fragmented.

This paper is related to various trends of the literature. As the fragmentation of production

[3]Industries at later stages of production chains are more intensive in advertising and less intensive in capital.

is closely related to the decision to outsource, it relates to an extensive amount of studies in industrial organization on the determinants of vertical integration (see Lafontaine and Slade, 2007, for a survey of previous empirical works). The fragmentation of production also reflects the division of labor, the development of markets, which in turn may depend on institutions.[4] The field of international trade has traditionally drawn a lot of attention to the fragmentation of production, and even more recently as different production stages may occur in different countries and trade in intermediates now accounts for a large fraction of total world trade (Yeats, 2001, Campa and Goldberg, 1997). Recent macroeconomic models also incorporate intermediate goods and vertical linkages which can magnify business cycles and productivity differences (see Jones, 2010): a better grasp on the extent of the fragmentation of production is essential to understand the role of these mechanisms.

Previous work in the field of industrial organization generally rely on case studies or industry studies to evaluate the extent of the vertical fragmentation,[5] and lack systematic ways to characterize the aggregate economy and compare industries. Various measurements of integration have been used. Firm size (e.g. Brynjolfsson, Malone, Gurbaxani and Kambil, 1994) provides a simple index of fragmentation. However it does not disentangle vertical from horizontal fragmentation. Several studies have used the ratio of value added to gross output (first investigated by Adelman, 1955, applied e.g. by Machiavello, 2009) as an index of vertical integration or fragmentation.[6] The closest to this paper are vertical integration indices combining information on multi-product firms and industry linkages from input-output matrices. Such an index of vertical integration takes 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). Such an approach has been taken by Maddigan[7] (1981), Hitt (1999) Fang and Lan (2000), Acemoglu, Johson and Mitton (2007), Acemoglu, Aghion, Griffith and Zilibotti (forthcoming) among others. This approach has at least three caveats. The first is that it requires detailed firm-level data with sufficient information on the range of products that are produced. This makes it difficult to study the evolution of the economy over an extended period of time and difficult to characterize the economy as a whole. A second caveat is that it is

[4]See for instance Acemoglu, Johnson and Mitton (2007), Carluccio and Fally (2010) on the role of financial institutions and entry regulations in shaping vertical integration.

[5]These include the auto industry, see Helper (1991), Abraham and Taylor (1996); the electronic industry, see Sturgeon (2002); the aerospace industry, see Masten (1984); the aluminum industry, see Stuckey (1993).

[6]My index presents many advantages over this one. By construction, it better accounts for inter-industry linkages. Second, it is more directly related to vertical fragmentation as it provides an estimation of the average number of production stages by industry. Interestingly, I show that these two indices coincide for the aggregate economy: there is equality between my index and the ratio of gross output to value added when I take the average across all products weighted by their contribution to final demand.

[7]"Vertical Industry Connection Index"

sensitive to the product classification employed and the lack of sufficiently disaggregated data.[8] Another issue is that this index is based on ownership structure rather than actual shipments of intermediate goods.[9] These differences in sample coverage and methodologies might explain why previous studies have not identified the decrease in vertical fragmentation documented in this paper.

The trade literature provides various examples of global supply chains and the cross-border fragmentation of production (e.g. Feenstra, 1998). Moreover, previous papers have developed indices to measure the extent of vertical specialization (e.g. Hummels, Ishii and Yi, 2001). In comparison, my paper aims at capturing the fragmentation of production across plants instead of fragmentation across borders. There is of course a connection: the large decrease in transport costs over the past decades has provided more opportunities to fragment production. I show that there isn't just a substitution between domestic outsourcing and foreign outsourcing, but firms also rely less on within-plant production.[10]

I also provide an illustration of an alternative use of these measures of fragmentation. 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 in goods that involve fewer production stages and 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. 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.

The remaining of the paper contains four sections. Section 2 defines the key indices and describes their aggregation properties. Section 3 describes the data. In Section 4, I present descriptive statistics, document the shift of value added towards final stages, discuss potential explanations of the changes in vertical fragmentation and examine the role of trade. Section 5 presents several robustness checks and Section 6 concludes.

[8] For instance, if one input is classified in the same category as one of the firm's product, this firm willl be interpreted as being vertically integrated even if the output is actually distinct from the input that is needed.

[9] A recent paper by Hortacsu and Syverson (2009) show that, remarkably, intra-firm shipments constitute a very small fraction of all shipments across plants, even when these firms own upstream and downstream plants in vertically-linked industries. They conclude that the main purpose of cross-ownership in vertically-related industries is *not* to facilitate input-output relationships.

[10] In ongoing research, Fort (2011) 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.

2 Definition and properties

2.1 Production stages and distance to final demand

In this section, I start by constructing two measures N_i and D_i defined by industry or product[11] (e.g. autos vs. steel) to characterize the position along production chains. For each product, I define:

i) N_i to reflect how many plants (stages) are sequentially involved in the production of this good i;

ii) D_i to measure how many plants 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.

To construct N, 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}. 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.[12]

If a product doesn't 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.

Let us consider a very stylized example. Suppose that a car industry requires 50 cents of auto parts for each dollar worth of car produced by this industry (the other 50 cents being value added by the car industry) and assume that auto parts are made from scratch in another plant. The measure of vertical fragmentation defined above equals 1 for the auto part industry, and $1 + 0.5 = 1.5$ for the car industry. I henceforth name this index the "number of production stages" but in general, this index is not an integer. Instead it could be considered as the average number of stages (plants) involved in the production chain, weighted by the value added at

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

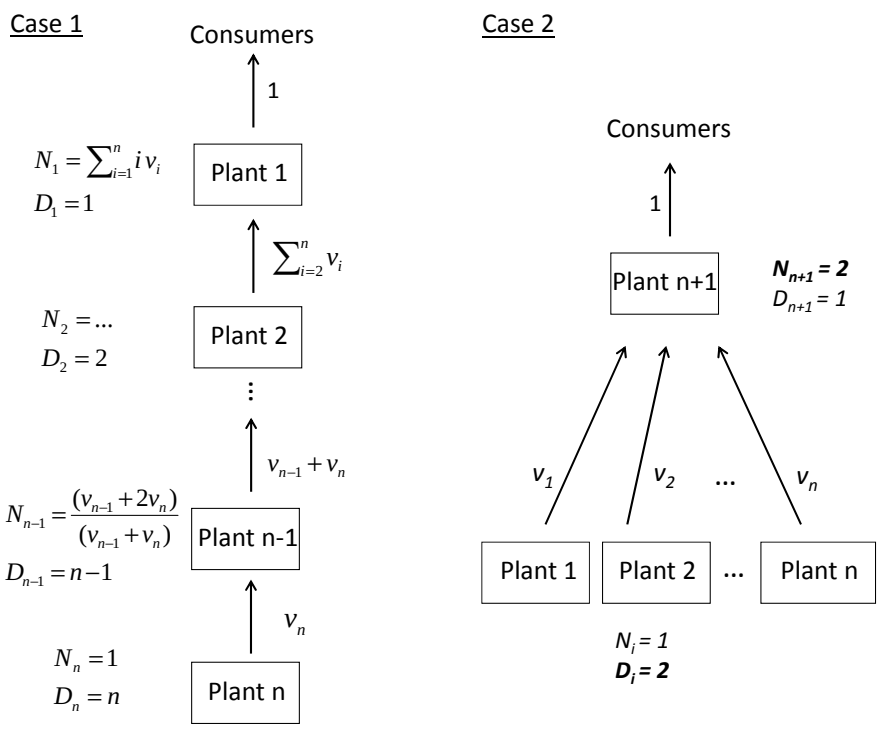
[12] By inverting this system of equations, we obtain the (transposed) matrix of total requirements. This measure of production stages corresponds to the sum of "total requirement" coefficients for a given industry.

each stage. In this simple example, half of the value has been added in the first stages (going through 2 plants), while the other half has been added at the last stage. The weighted average equals 1.5.[13]

While this measure aims at capturing the sequentiality in 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. In particular, this index does not capture the complexity of production when lots of different components are required for a complex assembly (e.g. aerospace industry). This point is illustrated in Figure 2, cases 1 and 2.

Figure 2: Vertical vs. horizontal fragmentation: an illustration



In both cases, each plant i contributes to a fraction v_i of the final value of the product ($\sum_{i=1}^{n} v_i = 1$). Case 1 involves sequential production whereas case 2 involves simultaneous production. In case 1, the measure of fragmentation increases with the number of suppliers

[13]This weighted measure will be of course strictly less than the actual number of plants involved in a production chain. In Section 5.3, I examine an alternative index, inspired from the Herfindahl-Hirschman Index, to measure the dispersion of value added along the chain.

because each of them enters sequentially in production. In case 2, however, they all ship to the same plant, so the degree of verticality does not depend on how many of them ship to this plant. In the second case, the measure of vertical fragmentation does not depend on the number of suppliers ($N = 2$ in the final stage). Baldwin and Venables (2010) classify these two cases as "snakes" and "spiders"; my index only captures snakes and is indifferent to spiders.

- When the input-output table is constructed at the plant level (such as the BEA input-output matrix for the US), this index reflects the fragmentation of production across plants independently from the ownership structure. 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. Note however that, according to Hortacsu and Syverson (2009), shipments across 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.

- This measure does not depend on the share of imported inputs in intermediate goods purchase as long as products of the same classification requires the same number of production stages abroad as domestically.[14] 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.[15] In other words, the 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 fragmentation. There is an effect only if sourcing substitutes to within-plant production.

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 production used as intermediate goods in industry j. We denote this coefficient by φ_{ij}. In other words, φ_{ij} denotes the fraction of production from industry i that is purchased as an intermediate good

[14] 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.

[15] 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.

by industry j. In an open economy, this coefficient φ satisfies:

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

where Y_i stands for the value of production of good i, M_i for imports and X_i for exports.

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 is similar to N. 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.

If we think of a sequential production chain where each plant contributes to value added by the same amount (case 1 in Figure 2), this index D is generally inversely related to N. D takes high values at early stages and decreases for later stages as N increases. 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. A similar result holds for the aggregate economy (see Proposition 2).

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 important measurement errors when the input-output table is partially aggregated.

2.2 Index for the aggregate economy

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 correct appropriate weights to compute averages.

With these two indices in hand, we can compute:

i) The average number of production stages associated with final goods on average. This makes use of index N. For this purpose, a natural weight is the total value of good i used for final consumption.

ii) The average number of stages between production and final consumption (distance to final demand), making use of index D. 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. It satisfies: $V_i = (1 - \sum_j \mu_{ij}) Y_i$. It equals production of good i minus the total use of intermediate goods for the production of good 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 1 *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.[16]

Note that the ratio of gross output to value added would also be the solution for the fragmentation index if the input-output matrix only had one industry (i.e. if it were fully aggregated). If μ is the input-output coefficient for the aggregate economy (i.e. the average value of intermediate goods needed to produce one dollar of output, which is also the share of production used as intermediates), this index would be the solution of the following equation:

$$N = 1 + \mu N$$

It's straightforward to verify that the ratio of gross output to value added is equal to the solution of this equation: $\frac{1}{1-\mu}$.

[16] However at the industry level, please note that both indices N_i and D_i differ from the gross-output-to-value-added ratio (see Table 2).

Open economy

In an open economy, net imports $M_i - X_i$ no longer equal zero. In particular, there is no longer equality between supply and demand for intermediate goods by domestic industries. In an open economy, 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 2 *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:*

$$\frac{\sum_i C_i N_i}{\sum_i C_i} = \bar{N} + \frac{\sum_i (M_i - X_i)(N_i - \bar{N})}{\sum_i C_i} \tag{3}$$

$$\frac{\sum_i V_i D_i}{\sum_i V_i} = \bar{N} + \frac{\sum_i (X_i - M_i)(D_i - 1)}{\sum_i V_i} \tag{4}$$

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

When net trade $(M_i - X_i)$ is not correlated with either fragmentation indices 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 underestimate the weighted average number of production stages as it does not account for the number of production stages embodied in imports. Conversely, the gross output to value added ratio underestimate the average number of stages to final demand when a country tends to export goods that are further from final demand.

2.3 From varieties to industries

Ideally, the unit of observation would be 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 subsection 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.

A few additional notations are 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.[17] 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 i that are sold as final goods. For a given industry i, the sets Ω_{ij} and Ω_{iF} offers 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 3 *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

[17] 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).

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 the aggregate economy (the gross-output-to-value-added ratio) 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 4 *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: 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.[18]

I use the BEA input-output tables for benchmark years, which are available online.[19] Unfortunately, industry classifications are not always homogenous across periods:

- The 1997 and 2002 IO tables are available in the NAICS classification (430 industries);
- The 1967, 72, 77, 82, 87 and 92 IO tables follow the SIC classification (6-digit level);
- The 1963 table follows the SIC classification (4-digit level);
- Previous tables (1947 and 1958) are aggregated across 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 382 homogenous industries.[20] 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.[21]

[18] 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).

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

[20] 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.

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

Note that the industry classification is more precise for manufacturing goods and commodities, with 311 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. 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 311 consolidated input-output industries, 270 of which having information on all variables.

4 Empirical Findings

4.1 Descriptive statistics

Aggregate number of production stages, 1947-2002

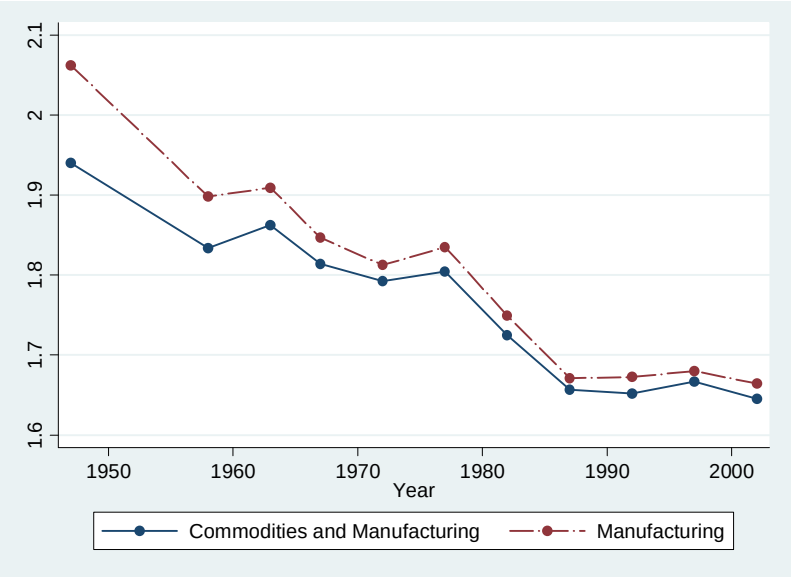
The first striking fact is that the weighted average number of production stages for the US is below 2. This can be seen in Figure 3 showing the ratio of gross output to value added. Production is not as disintegrated as we could expect. In other words, the value added embodied in production goes across less than two plants on average before reaching final demand.

Figure 3: Weighted average number of production stages

Moreover, the fragmentation of production 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. 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.[22]

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.[23] In comparison, data on manufacturing sectors are mode detailed.

Figure 4: Weighted average number of production stages (tradables excl. petroleum)



In Figure 4, I compute the aggregate index of fragmentation by only considering tradable goods (manufacturing goods and commodities, excluding services and petroleum-related industries). We now focus on industries that require a larger number of production stages compared to services. However, the number of production stages drops if we do not account for part of the inputs (services). As a result, the average number of production stages after excluding services is roughly the same as before (around 2). Even if we exclude services, the downward trend is confirmed. The average number of production stages of tradable goods declined from 2

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

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

to 1.6 over the past 50 years. We can further restrict our attention to manufacturing industries but the picture remains similar if we account for all tradable inputs.

The results so far 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 2. In particular, the aggregate number production stages (weighted by final consumption) can differ from the aggregate number of stages to final demand (weighted by value added). Using industry-level trade data from 1967 to 1992, I compute the deviations from the closed economy case as described in Proposition 2 (differences between aggregated indices 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 than 0.02, i.e. less than a 1% error. Figures 3 and 4 would thus remain the same 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 to be somehow related to fragmentation. This issue is further discussed in Section 4.4.[24]

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	1.895	0.002	-0.013
1972	0.064	1.800	0.011	-0.009
1977	0.073	1.806	0.011	-0.011
1982	0.094	1.723	0.015	0.000
1987	0.140	1.662	0.013	0.028
1992	0.157	1.658	0.012	0.020

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 deviations from GO/VA.

[24]In Section 4.4 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.

Fragmentation of production across industries in 1992

I begin by providing examples of industries with the largest measures of production stages. Food industries typically involve long production chains (see Table 2a). Among the top-5 industries with the largest number of production stages, 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 average number of stages is 2.8, and it is 2.4 for auto parts.

Table 2: Industries with the largest index values

Table 2a: **Measure of production stages**

Production stages	All inputs	Tradables	GO/VA
Top-5 industries:			
Meat packing plants	3.50	2.67	8.74
Sausages and other prepared meat products	3.40	2.65	4.88
Leather tanning and finishing	3.17	2.43	3.93
Natural, processed, and imitation cheese	3.16	2.35	3.94
Creamery butter	3.15	2.36	5.12
Motor vehicle industries:			
Motor vehicles and passenger car bodies	2.81	2.04	6.09
Motor vehicle parts and accessories	2.42	1.78	3.15
Truck and bus bodies	2.42	1.82	2.83
Truck trailers	2.61	1.92	3.75

Table 2b: **Measure of stages between production and final demand**

Stages to final demand	All inputs	Tradables
Electrometallurgical products, except steel	8.06	6.40
Iron and ferroalloy ores	6.68	5.33
Primary and secondary nonferrous metals, n.e.c.	5.46	4.38
Copper ore	5.11	4.33
Primary smelting and refining of copper	5.04	4.12

Note that the fragmentation index differs from the gross-output-to-value-added ratio for

the aggregate economy 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 services 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".

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
Greenhouse and nursery products	1.33	Greenhouse and nursery products	1.08
U.S. Postal Service	1.37	Manufactured ice	1.14
Other Federal Government	1.44	Forestry & fishery products	1.16
Real estate	1.45	Brick and structural clay tile	1.16

An overall comparison between commodities, manufacturing goods and services confirms the previous picture (Table 4). Manufacturing industries involve 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 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 depending on various industry characteristics. The choice of these industry characteristics is primarily motivated by the literature on firm boundaries (see Lafontaine and Slade, 2007, for a survey). 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 (2009) 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.

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	2.21	2.12	1.61	1.53
Commodities	2.07	3.02	1.38	2.46
Services	1.78	1.81	/	/
Petroleum	2.36	3.55	/	/

The literature on the boundaries of the firm has identified various factors. 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).[25] Following Antras (2003) model based on the property-right approach, the internalization 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 Antras, 2005). Other factors affecting integration include competition and market thickness (McLaren, 2000) and financial constraints (Acemoglu, Johnson and Mitton, 2007). We proxy competition by the fraction of output produced by the 4 largest companies in the industry[26] 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 diffcult 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. We can expect a negative correlation between specificity and vertical fragmentation.

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

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

Pairwise correlations between the fragmentation 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 are generally less fragmented, which is surprising if we expect hi-tech industries to be more complex and combine multiple inputs.[27] Nevertheless, these results are in line with the literature on vertical integration. In particular, there is a negative and significant correlation with product specificity, R&D intensity, skill intensity and dependence in external finance. We find however no significant correlation with capital intensity, productivity and industry concentration. Turning to the second column, 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.[28]

Table 5: Pairwise correlations with industry characteristics

Variable:	Production Stages	Distance to final demand
Specificity	-0.196*	-0.496*
R&D	-0.197*	-0.042
Capital intensity	0.065	0.465*
Skill intensity	-0.268*	-0.170*
Advertising intensity	-0.075	-0.266*
Productivity	-0.124	-0.116
Financial Dep	-0.191*	0.233*
Share of top 4 firms	-0.008	0.066

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

We should also note that production stages and distance to final demand are weakly correlated across all commodities and manufacturing industries. The correlation is negative until 1982: -7% in 1967, -4% in 1972, -2% in 1977. Then it is smaller than 1% (in absolute value) after 1982. This small correlation shows that these two indices capture different dimensions of the fragmentation of production and can be both informative to characterize the position of an industry along supply chains.

[27] As mentioned before, the measure of fragmentation depends on the share of the cost of inputs in production but does not depend on how many different inputs are assembled. It depends however on whether the production of these intermediates goods has required other intermediate goods etc.

[28] Very similar results are obtained with multivariate OLS regressions.

4.2 The shift of value added towards final stages

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. According to Proposition 1, both shifts can contribute to the aggregate decrease in fragmentation.[29]

To answer these questions, 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 added (Decomposition 2):

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

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.

I first decompose the change in the index calculated for all industries, including all inputs (Table 6, Panel A). Panel A shows very similar results for both decompositions. Whereas the within effects can be quite large in magnitude for some periods, especially in 1977, the contribution of the between effect is consistently negative across periods. Overall, the between effect dominates. The negative trend in the between effect for both indices can be explained by

[29] 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 equal for the US.

a shift of demand and production towards services. Services require fewer stages and are closer to final demand. Also note that the positive within effect for 1977 and 1982 in Decomposition 1 can be related to the oil-price shock as we saw previously. Petroleum is essentially consumed as intermediate demand and an increase in its price shifts the fragmentation index (its consumption is inelastic). Similarly, a positive oil-price shock can explain a positive between effect in 1977 in Decomposition 2.

Table 6: Within and between decompositions

Panel A: **All industries**

	Aggregate measure of fragmentation:		Number of production stages: (decomposition 1)		Distance to final demand: (decomposition 2)	
Year	Level	Change	Between	Within	Between	Within
1967	1.932	/	/	/	/	/
1972	1.865	-0.067	-0.021	-0.046	-0.046	-0.021
1977	1.939	0.074	-0.004	0.078	0.038	0.036
1982	1.944	0.005	-0.037	0.042	0.007	-0.001
1987	1.855	-0.089	-0.019	-0.07	-0.058	-0.031
1992	1.822	-0.033	-0.023	-0.010	-0.015	-0.018
All		-0.110	-0.124	0.015	-0.089	-0.021

Panel B: **Tradables only**

	Aggregate measure of fragmentation:		Number of production stages: (decomposition 1)		Distance to final demand: (decomposition 2)	
Year	Level	Change	Between	Within	Between	Within
1967	1.895	/	/	/	/	/
1972	1.800	-0.095	0.017	-0.111	-0.042	-0.052
1977	1.806	0.006	-0.029	0.035	0.034	-0.028
1982	1.723	-0.082	-0.032	-0.051	-0.064	-0.019
1987	1.662	-0.062	0.003	-0.065	-0.043	-0.018
1992	1.658	-0.004	-0.004	0.001	0.010	-0.013
All		-0.237	-0.035	-0.202	-0.101	-0.136

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.

Now, I decompose the change in fragmentation by considering tradable goods only (manufacturing and commodities excluding petroleum). Panel B shows that the between effect is much smaller for tradable goods, and a large part of the evolution across years is explained by the within effect. This confirms that the results from Panel A are mostly driven by the shift towards services.

The between effect in Decomposition 2 remains quite large: there is a shift in value added towards final stages of production. The magnitude of the between effect is comparable to the within-effect. Except for 1977 (oil price shock), the between effect is large and negative. The within effect is smaller in magnitude but consistently negative across periods.

Table 7 examines the shift in value added in more details. In columns (1) to (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). The coefficient is negative and significant; the beta coefficient equals -0.221.

This result confirms the negative between effect found in Table 6 (Panel B, Decomposition 2) for the shift in value added. In column (2), we control for the number of production stages. The coefficient is not significant, which reflects the small between effect found in decomposition 1.[30] 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 in column (5).

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

Table 7: Shift of value-added towards final stages

Dependent variable:	VA Growth	VA Growth	VA Growth	Increase in VA/GO	Increase in VA/GO
Stages to final demand	-4.471	-4.39	-2.940	-0.671	-0.381
	[1.114]***	[1.105]***	[1.259]***	[0.149]***	[0.210]**
Number of stages		-2.409	-3.296		
		[2.952]	[2.970]		
Specificity			-7.566		-0.410
			[2.647]***		[0.361]
R&D intensity			0.358		-0.163
			[0.625]		[0.085]*
Capital intensity			-6.937		-1.015
			[2.048]***		[0.309]***
Skill intensity			4.830		2.271
			[9.276]		[1.634]
Advertising intensity			0.666		0.017
			[0.257]***		[0.071]
Productivity growth			23.748		-0.260
			[10.823]**		[1.166]
Financial Dependence			1.274		0.262
			[0.469]***		[0.074]***
Top 4 share			-0.076		0.000
			[0.039]*		[0.007]
Import penetration			-45.145		-4.176
			[8.474]***		[1.007]***
Number of industries	311	311	270	311	270
R-squared	0.05	0.05	0.30	0.06	0.24

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. 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%.

4.3 What explains the decrease in fragmentation?

Behind the shift in value added

After describing the decline in fragmentation and the shift of value added towards final stages, a important question is why does it occur. This section attempts at providing partial answers, but the main reasons behind these changes remain unknown.

Is international trade part of the answer? As shown in column (3) of Table 7, value-added has particularly decreased in industries facing higher import penetration. However, import penetration is not significantly correlated with the distance to final demand.[31] The same holds

[31] This result is described in Table 1: the correlation term between net imports and distance to final demand

for changes in import penetration which are also not correlated with distance to final demand across industries. This suggests that imports are not the main reason behind this shift in value added towards final stages. This issue will be further discussed in the next section (Section 3.4): results show instead that trade induced an *increase* in fragmentation.

An alternative potential explanation 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 2) which can explain why value-added growth is negatively correlated with distance to final demand. To examine this explanation quantitatively, I perform the following exercise:

i) First, I regress value-added growth on industry characteristics (all control variables from column 3 of Table 7) 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 7 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 -2.766 (significant at 1%). It equals 60% of the coefficients from Table 7, column 1. This result suggests that these industry characteristics can explain nearly two thirds of the negative correlation between value-added growth and distance to final demand, which itself explains half of the aggregate decrease in vertical fragmentation.

Behind the negative within effect

Another way to investigate this decline in vertical fragmentation is to examine the changes in the measure of production stages (index N_i) by industry. That and the shift of value added are two faces of the same coin. As documented in Table 6 (Panel B, Decomposition 1), the overall decline essentially corresponds to "within" changes, i.e. a decrease in the measure of fragmentation rather than a change in the composition of final demand with respect to the number of production stages ("between" effect).

Table 8 explores the determinants of the change in fragmentation by industry. The dependent variable is increase in the index of fragmentation: $\Delta N_i = N_{i,1992} - N_{i,1967}$. Results in

is very small (last column). This is also confirmed in Table 9: distance to final demand is not significantly correlated with import penetration (column 2).

column (1) show that the change in fragmentation is positively related to product specificity, R&D intensity and capital intensity, and negatively related to skill intensity and financial dependence. All-in-all, these industry characteristics can account for 14% 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. Antras 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.

Table 8: Within-industry changes

Dependent variable:	ΔN	ΔN	ΔN
Specificity	1.449	1.416	1.001
	[0.604]**	[0.605]**	[1.875]
R&D intensity	0.482	0.455	1.728
	[0.143]***	[0.172]***	[0.435]***
Capital intensity	1.828	1.828	2.614
	[0.477]***	[0.467]***	[1.335]*
Skill intensity	-5.500	-5.701	-16.249
	[2.878]*	[2.923]*	[5.328]***
Advertising intensity	-0.097	-0.094	-0.092
	[0.073]	[0.074]	[0.071]
Productivity growth	-2.521	-2.047	-11.709
	[2.067]	[2.294]	[6.540]*
Financial Dependence	-0.392	-0.376	-1.024
	[0.143]***	[0.159]**	[0.335]***
Top 4 share	-0.008	-0.012	-0.009
	[0.012]	[0.014]	[0.030]
Δ R&D int.		0.044	1.136
		[0.513]	[1.278]
Δ Capital int.		0.032	-0.132
		[0.028]	[0.050]***
Δ Skill int.		-0.048	-0.461
		[0.174]	[0.399]
Characteristics	Same industry	Same industry	**Upstream** industry
Number of industries	270	270	270
R-squared	0.14	0.14	0.13

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%.

Could the decline in fragmentation be explained by technological change? As Table 6 shows, R&D and skill-intensive industries are less vertically fragmented. Hence one might suspect that changes in R&D intensity and skill intensity 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.[32]

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 Nunn (2007) methodology and construct a set of variables $x_{it}^{k,up}$ such that:[33]

$$x_{it}^{k,up} = \frac{\sum_j \mu_{ijt} x_{it}^k}{\sum_j \mu_{ijt}} \tag{5}$$

As shown in column (3), the results based on upstream industry characteristics are however very similar to column (2).

In Section 4.5, I investigate alternative explanations of the decline of fragmentation. I examine the measure of fragmentation by taking consumer prices instead of producer prices, hence reincorporating transport and retail margins in intermediate goods consumption. I also show that the relative price of intermediate goods compared to final goods has remained stable over the past 50 years. Therefore, changes in prices are not likely to explain the observed decrease in fragmentation.

4.4 Trade and vertical fragmentation

This section focuses on the effect of international trade on vertical fragmentation in the US. 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

[32]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.

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

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 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. Remind that the measure is constructed using the value of intermediate goods usage in production. If trade reduces the price of intermediate goods, there is a possibility that it also reduces the amount spent on these goods. It can occur if there is a very low substitution between outsourced intermediate goods (domestically or internationally) and intermediate goods produced within the plant: a reduction in the price of outsourced inputs would lead to a reduction in their share of total production costs.[34]

A first question is whether fragmentation is correlated with import penetration across industries (in cross section). In Table 1 on the difference between the GO/VA ratio and aggregate measures of fragmentation corrected for trade, results show that there is only a very 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 correlation (OLS regression with robust standard errors).

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. A better check is to test whether increases 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 suggesting that trade indeed creates new opportunities to fragment production. The beta coefficient equals 0.185. Controlling for other industry characteristics does not affect the main coefficient.

One may be worried that the increase in import penetration be endogenous to the change in fragmentation by industry. To mitigate endogeneity biases, I regress the change in fragmentation on import penetration in 1967. As imports have grown faster in industries with higher initial import penetration, I find similar results in column (5): the increase in the number of production stages is positively correlated with initial import penetration. Alternatively, I can regress the change in fragmentation on the increase in import penetration by instrumenting the latter by the initial level of import penetration. The results (not shown) are very similar

[34]Note that Table 7 already suggests that this effect does not dominate. In column (5), I find that the share of value added in production has decreased in industries facing larger import penetration.

Table 9: Import penetration and the measure of production stages

Dependent variable:	N	D	ΔN	ΔN	ΔN	ΔN
Imports	-0.199	-0.002				
	[0.129]	[0.440]				
Increase in imports			0.180	0.170		
			[0.052]***	[0.068]**		
Initial imports					12.158	18.220
					[3.241]***	[6.112]***
Specificity				1.309	1.684	1.730
				[0.604]**	[0.587]***	[1.812]
R&D intensity				0.433	0.447	1.329
				[0.144]***	[0.140]***	[0.384]***
Capital intensity				1.958	1.852	2.247
				[0.469]***	[0.471]***	[1.179]*
Skill intensity				-5.574	-4.643	-14.903
				[2.859]*	[2.832]	[5.468]***
Productivity growth				-2.442	-2.535	-5.512
				[1.991]	[2.017]	[5.602]
Financial Dep.				-0.348	-0.340	-0.460
				[0.148]**	[0.142]**	[0.242]*
Top 4 share				-0.009	-0.010	-0.027
				[0.011]	[0.011]	[0.028]
Characteristics	Same industry	Same industry	Same industry	Same industry	Same industry	**Upstream** industry
Nb of industries	311	311	311	270	270	270
R-squared	0.01	0.01	0.03	0.16	0.19	0.12

Notes: OLS regressions. Dependent variables: Measure of production stages by industry (col. 1: average between 1967 and 1992; col. 3 to 6: increase between 1967 and 1992); measure of stages to final demand (col. 2: average). Independent variables: average import penetration (col. 1 and 2); increase in import penetration (col. 3 and 4); initial import penetration in 1967 (col. 5 and 6). Data on industry characteristics are described in the appendix. Robust standard errors into brackets; * significant at 10%; ** significant at 5%; *** significant at 1%.

to column (4). In the first stage, I find indeed that the increase in import penetration is significantly larger in sectors with higher initial import penetration. In the second stage, the new coefficient is even larger than in the OLS regressions.

I perform a similar exercise using import penetration in upstream industries (taking an average weighted by input-output coefficients as in equation 5) and control variables for the corresponding industries. Note that downstream and upstream variables are highly correlated. It is therefore not surprising to find very similar results: vertical fragmentation has increased more in industries where related upstream industries were facing larger import penetration in 1967 (an IV approach also leads to the same conclusion as in the previous case).

While imports are not correlated with fragmentation in cross-section, our results show that opening to trade is associated with an increase in the fragmentation of production. This finding is in line with common expectations. In light of these results, the overall decrease in fragmentation is even more puzzling.

Vertical specialization

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 (e.g. Hummels, Ishii and Yi, 2001, Johnson and Noguera, 2010).

We have seen in Table 7 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):

$$\log M_{ic} = \beta_N \,.\, N_i \,.\, \log(PCGDP_c) + \beta_D \,.\, D_i \,.\, \log(PCGDP_c) + \alpha_i + \eta_c + \varepsilon_{ic}$$

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. In particular, I have shown in Table 2 that the fragmentation of production is correlated with product specificity, skill intensity, R&D intensity. Thus, a country with better contractual institutions or skill endowments may specialize in less fragmented industries. To account for these explanations, I further control for interactions between capital intensity and capital endowments, skill intensity

and skill endowments (as in Romalis, 2004), product specificity and judicial quality (as in Nunn, 2007).

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.[35] 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:

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

Hillberry and Hummels (2000) and Yi (2010) would predict a negative interaction term between physical distance and the number of production stages ($\gamma_N < 0$) as multi-stage production should magnify the negative effect of distance (note that the direct effect of physical distance is already captured by country dummies).

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.[36]

Table 10 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). [37] Moreover, richer countries specialize in industries that are closer to final demand. The latter is consistent with Costinot et al (2011) while the former is not.

As shown in column (2), physical distance seems to have a stronger negative impact in industries that are further from final demand, as measured by the index of vertical fragmentation D_i. The interaction term between physical distance and the number of production stages N_i is however not significant. These results hold after combining all interaction terms with production fragmentation indices (column 3) which means that the results are not driven by spurious correlations between GDP per capita and physical distance to the US.

In column (4), I include other controls: interaction terms between capital intensity and

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

[36] This method follows Rauch (1999).

[37] As the number of production stages N_i is not significantly correlated with the number of stages to final demand D_i, the same interaction term is obtained for either of them whether I include the other one or not.

Table 10: Comparative advantage along supply chains

Dependent variable:	Imports	Imports	Imports	Imports
PCGDP * production stages	-0.489		-0.480	-0.038
	[0.122]***		[0.125]***	[0.129]
PCGDP * stages to final demand	-0.110		-0.125	-0.021
	[0.040]***		[0.041]***	[0.045]
Distance * production stages		0.201	0.105	0.018
		[0.168]	[0.172]	[0.178]
Distance * stages to final demand		-0.148	-0.171	-0.154
		[0.058]**	[0.057]***	[0.057]***
Distance * transportability				-5.924
				[0.845]***
K endowment * K intensity				0.124
				[0.039]***
Skill endowment * Skill intensity				8.091
				[0.692]***
Judicial Quality * Specificity				2.501
				[0.338]***
Industry FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Observations	11821	11821	11821	11794
R-squared	0.54	0.54	0.54	0.55

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

capital endowments, skill intensity and skilled labor endowments (both significant and positive as in Romalis, 2004), judicial quality and product specificity (positive and significant as in Nunn, 2007). Moreover, the interaction term between physical distance and the proxy for product tradability is negative and significant, as expected: physical distance has a stronger impact on goods are more difficult to trade. However, the two interaction terms between GDP per capita and the stage indices become insignificant when I control for capital, skills and judicial quality. This result suggests that the sorting of stages across countries could simply be explained by traditional sources of comparative advantage.

5 Robustness

5.1 On the evolution of relative prices

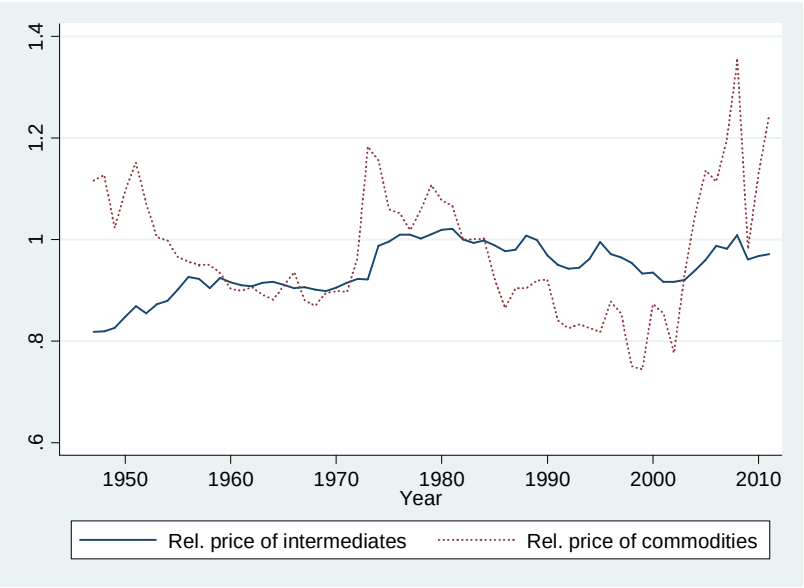
The measure of vertical fragmentation is constructed using values of purchased intermediate goods, not quantities. Hence, one may be concerned that the decline in fragmentation is simply

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 5 plots the ratio of the price index of the second and third category over to the first one (yearly average).

Figure 5: 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 5, 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 9) 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 6 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 6 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 7), even if retail and wholesale margins have slightly increased.

Figure 6: Incorporating transportation margins

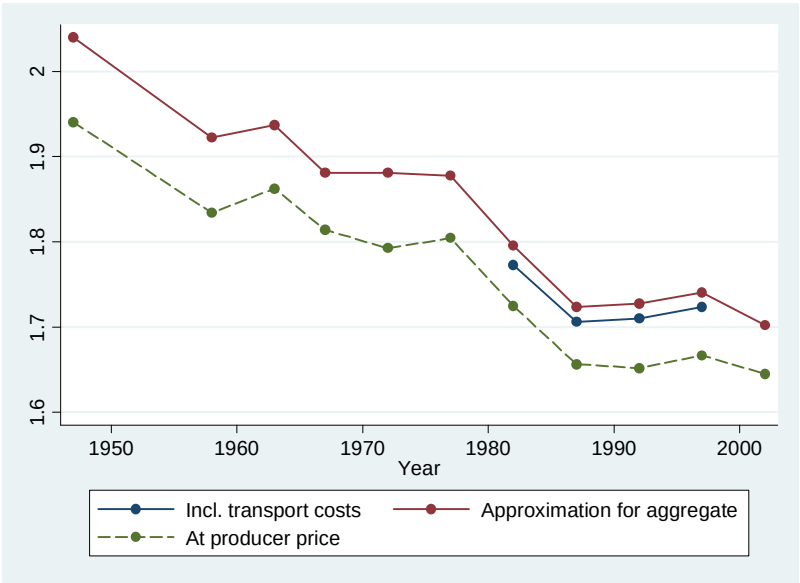
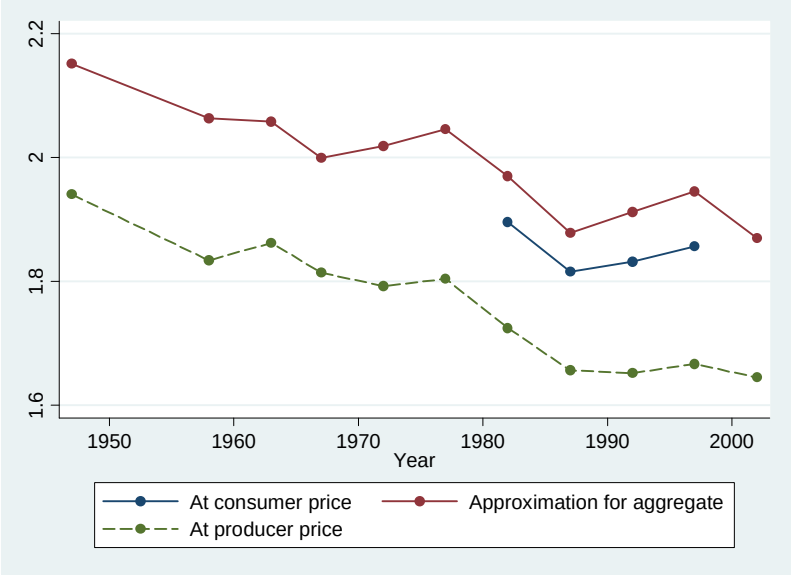


Figure 7: Measure of fragmentation at consumer prices



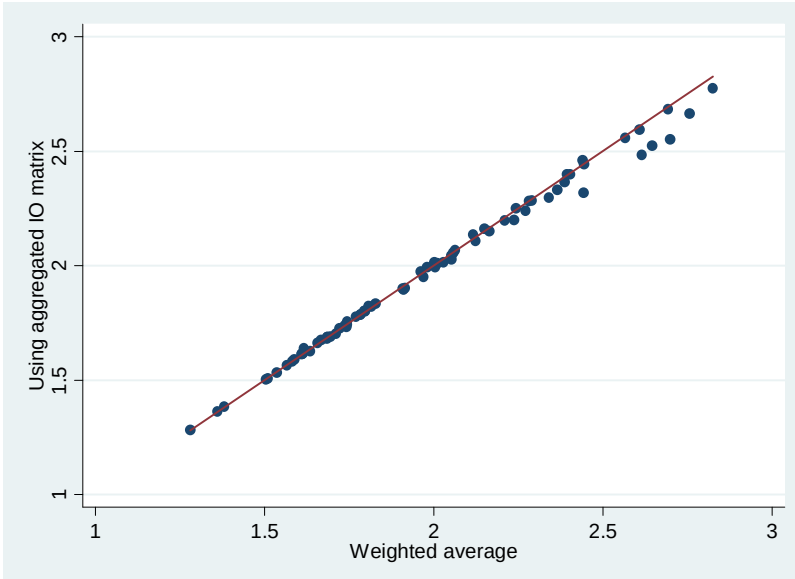
5.2 Aggregation

As shown by Proposition 2, 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 3 and 4, 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 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 industry i, I implicitely define H_i by:

$$H_i \;=\; \Big(1-\sum_j \mu_{ij}\Big)^2 \;+\; \Big(\sum_j \mu_{ij}\Big)\Big(\sum_j \mu_{ij} H_j\Big)$$

This characterizes one equation for each industry. This linear system of equation generally has a unique solution which defines H_i.

Note that H_i is inversely related to the fragmentation of production. The lower it is, the more dispersed is production across stages. There is a close link with HHI. For instance, when the economy is composed of n plants involved sequentially in production (such as case 1 in Figure 2), the measure H_n for the final product corresponds to the Herfindahl-Hirschman Index of the concentration of value added across these n plants.

I calculate this index for all tradable industries (excluding services and petroleum-related industries as in previous tables). I find a very large negative correlation between this new index H_i and the previous index N_i across industries (taking averages across years): the correlation equals -90.5% and is highly significant. 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. The average of H_i across industries (weighted by final consumption) has steadily increased from 0.42 in 1967 to 0.49 in 1992.

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 to value added. A variant of this measure reflects the number of stages between an industry's production and final demand. These two indices are simple to calculate and only require input-output tables that are generally publicly available. Moreover, I show that these two indices have good aggregation properties. In particular, by calculating these indices using more aggregated input-output matrices, we generate only small aggregation biases.

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. In particular, I show that a large part of this decline corresponds to a shift of value added towards final stages of production.

The reasons behind the decrease in fragmentation remain largely unknown. Half of it can be explained by the shift of value added towards industries that are closer to final demand, which in turn can be partially explained by a larger growth rate in advertising-intensive industries. This leaves however more than half of the overall decline in fragmentation unexplained. Increases in demand for skills and innovations do not seem to be related to changes in vertical fragmentation across industries. Moreover, this decrease in fragmentation cannot be simply explained by

decreases in relative commodity or intermediate goods prices, or increasing trade margins. On the contrary, I find that the large increases in import penetration over the past decades have led to increasing vertical fragmentation.

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

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

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

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

[10] 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.

[11] 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.

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

[13] 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.

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

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

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

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

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

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

[20] 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.

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

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

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

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

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

[26] Hortacsu, A. and C. Syverson (2009), "Why Do Firms Own Production Chains?", University of Chicago, manuscript.

[27] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[43] Stuckey, James (1983) *Vertical Integration and Joint Ventures in the Aluminium Industry*, Harvard University Press.

[44] Sturgeon, T. (2002), "Modular production networks: a new American model of industrial organization," *Industrial and Corporate Change*, 11(3), 451-496.

[45] 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.

[46] 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.

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

[48] 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: 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 2: 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 1, 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 2.

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 2.

Proposition 3: If $(\int_{\Omega_{ij}} y(\omega) N(\omega) d\omega) / (\int_{\Omega_{ij}} y(\omega) d\omega)$ does not depend on the downstream in-

dustry 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 3 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 3 (between lines 3 and 4 in the following equalities), we obtain successively:

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

$$= \frac{Y_i + \sum_{j \neq i} \tilde{N}_j \mu_{ij} Y_i}{(1-\mu_{ii})Y_i}$$
$$= \frac{1 + \sum_{j \neq i} \mu_{ij} \tilde{N}_j}{1-\mu_{ii}}$$

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 3 are satisfied.

Proof of Proposition 4: The proof follows the same logic as for Proposition 3.

Appendix on partial-aggregation properties

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 $i \notin \{1, 2, a\}$. This index is generally imperfect, but it should be very close to the "true" index defined with the disaggregated input-output matrix, taking the average between industries 1 and 2 weighted by the contribution of industries 1 and 2 to final demand. To be more precise, I obtain the following proposition:

Proposition 5 *$\tilde{N}_i = N_i$ and $\tilde{N}_a = [N_1F_1 + N_2F_2]/[F_1 + F_2]$ if one of these conditions is satisfied:*
i) $N_1 = N_2$
ii) $\mu_{i1}/\mu_{i2} = F_1/F_2$ across all other industries $i \notin \{1, 2, a\}$.
where F_i is defined by the total use of good i by other industries and final consumers:

$$\begin{aligned}
F_1 &= Y_1 - \mu_{11}Y_1 - \mu_{21}Y_2 \\
F_2 &= Y_2 - \mu_{12}Y_1 - \mu_{22}Y_2
\end{aligned}$$

Proof: Let's define $N'_i = N_i$ for $i \notin \{1, 2, a\}$ and $N'_a = \frac{F_1N_1 + F_2N_2}{F_1 + F_2}$. In order to prove that $\tilde{N}_i = N'_i$, we show that N'_i satisfies the same equations, i.e.

$$N'_a = 1 + \mu_{aa}N'_a + \sum_{j \notin \{1,2,a\}} \mu_{aj}N'_j \tag{6}$$

and:

$$N'_i = 1 + \mu_{ia}N'_a + \sum_{j \notin \{1,2,a\}} \mu ij N'_j \tag{7}$$

for $i \notin \{1, 2, a\}$ (under either condition specified in Proposition 5).

We begin by showing that equation (6) is always satisfied. Using successively the definition of F_i and N_i, we have:

$$\begin{aligned} N_1F_1 + N_2F_2 &= \Big(Y_1 - Y_1\mu_{11} - Y_2\mu_{21}\Big)N_1 + \Big(Y_2 - Y_1\mu_{12} - Y_2\mu_{22}\Big)N_2 \\ &= \Big(N_1 - \mu_{11}N_1 - \mu_{12}N_2\Big)Y_1 + \Big(N_2 - \mu_{21}N_1 - \mu_{22}N_2\Big)Y_2 \\ &= \Big(1 + \sum_{j\notin\{1,2,a\}} \mu_{1j}N_j\Big)Y_1 + \Big(1 + \sum_{j\notin\{1,2,a\}} \mu_{2j}N_j\Big)Y_2 \\ &= \Big(Y_1 + Y_2\Big)\Big(1 + \sum_{j\notin\{1,2,a\}} \mu_{aj}N_j\Big) \end{aligned}$$

Note that $(Y_1 + Y_2)(1 - \mu_{aa}) = F_1 + F_2$, therefore we obtain that:

$$(1 - \mu_{aa})\Big(\frac{F_1N_1 + F_2N_2}{F_1 + F_2}\Big) = 1 + \sum_{j\notin\{1,2,a\}} \mu_{aj}N_j$$

Hence:

$$\frac{F_1N_1 + F_2N_2}{F_1 + F_2} = 1 + \mu_{aa}\Big(\frac{F_1N_1 + F_2N_2}{F_1 + F_2}\Big) + \sum_{j\notin\{1,2,a\}} \mu_{aj}N_j$$

This proves that equation (6) is satisfied.

Now we need to prove that $N_i = 1 + \mu_{ia}\Big(\frac{F_1N_1+F_2N_2}{F_1+F_2}\Big) + \sum_{j\notin\{1,2,a\}} \mu_{ij}N_j$ for $i \notin \{1, 2, a\}$ (equation 7). Given that $\mu_{ia} = \mu_{i1}+\mu_{i2}$, and given that $N_i = 1+\mu_{i1}N_1+\mu_{i2}N_2+\sum_{j\notin\{1,2,a\}} \mu_{ij}N_j$ (by definition), we obtain that this equality holds if and only if:

$$\mu_{i1}N_1 + \mu_{i2}N_2 = (\mu_{i1} + \mu_{i2})\Big(\frac{F_1N_1 + F_2N_2}{F_1 + F_2}\Big)$$

When $N_1 = N_2$, this equality is obviously satisfied (condition i) of Proposition 5). When $N_1 \neq N_2$, we find that this equality is satisfied if and only if:

$$\mu_{j1}F_2 = \mu_{j2}F_1$$

This equality corresponds to condition ii) of Proposition 5.

Intuitively, this proposition 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.[38] 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

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

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). This index is stable by partial aggregation if it is weighted by the value of production minus the use of inputs within the same aggregated industry.

Proposition 6 *$\tilde{D}_i = \tilde{D}_i$ and $\tilde{D}_a = [D_1\tilde{V}_1 + D_2\tilde{V}_2]/[\tilde{V}_1 + \tilde{V}_2]$ 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 $\tilde{V}_i$ is the value of production in industry i net of the use of inputs by industries 1 and 2:

$$\tilde{V}_1 = Y_1[1 - \mu_{11} - \mu_{12}]$$
$$\tilde{V}_1 = Y_2[1 - \mu_{12} - \mu_{22}]$$

Proof: The proof of Proposition 6 follows the same logic as Proposition 5, by taking $\tilde{V}_i$ instead of F_i, and φ_{ij} instead of μ_{ij}.

Data Appendix: 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.[39] 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 finally that the main results presented throughout the paper are robust to dropping extreme observations for each variable (extreme percentiles).

US trade data are available in the 1972 SIC classification (after 1958) and 1987 SIC classification (after 1972) for manufacturing industries from Feenstra (1996), Abowd (1991) and Pierce and Schott (2009). For the last subsection of this paper (on imports across source countries) I complement the trade data by source country with Penn World Table data on GDP per capita, data on endowments in capital and skilled labor from Hall and Jones (1999) and data on judicial quality from Kaufmann, Kraay, and Mastruzzi (2003).

[39] 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.

Table 11: 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.