

Disentangling Global Value Chains*

Alonso de Gortari

alonso.degortari@princeton.edu

First version: August 2016

This version: September 2018

[\[latest version\]](#)

Abstract

This paper studies the implications of a key fact: That many different global value chain (GVC) networks aggregate up to the same multi-country input-output data. I argue in favor of networks where the use of inputs varies depending on the use of output - in contrast to the current literature based on networks where all output uses the same inputs. I provide evidence for this approach using Mexican customs data: cars exported to the U.S. use a higher share of imported American inputs than those exported elsewhere. This heterogeneity matters since both quantitative counterfactual estimates and measures of globalization such as value-added trade vary across GVC networks. I argue that GVCs are better measured when leveraging additional information: incorporating Mexican customs data implies that 30% of U.S. imported Mexican manufactures is U.S. value returning home - higher than the conventional estimate of 17%.

*This paper is based on my 2017 job market paper of the same title and the 2018 paper “Bounding the Gains from Trade”. I am extremely grateful to my advisors Pol Antràs, Elhanan Helpman, and Marc Melitz for their mentorship and guidance throughout this project, and am also thankful for the comments of numerous participants at seminars and conferences. I gratefully acknowledge the hospitality of Banco de México, where part of this paper was written. All errors are my own.

1 Introduction

While writing this paper, the North American Free Trade Agreement (NAFTA) was renegotiated for the first time since its inception in 1994, the United Kingdom discussed its potential exit from the European Customs Union, and the seeds of a possible full-blown trade war between the U.S. and China were sown. What are the potential costs of these economic shocks? How do they ripple across country borders? Are the shocks amplified in this age of globalization where over two-thirds of world trade is in intermediate inputs? Specifically, would a 25% tariff on imported Mexican vehicles - as considered by the current U.S. federal administration - hurt the American worker more or less depending on the share of American value built into these cars? Further, if Mexican car exports to the U.S. rely heavily on American value-added - as suggested by anecdotal evidence - why is this so? Throughout the next pages, I will argue that, addressing these questions requires first developing a more accurate, systematic, understanding of the nature of the global value chains (GVCs, henceforth) underlying world trade than has so far been achieved.

In a nutshell, this paper is about, first, showing that the conventional framework used to study GVCs may be mismeasuring these objects severely. Second, that this mismeasurement matters because it affects the quantitative exercises carried out by both academics and policymakers to study global trade in a world of highly fragmented production. And, third, that there is a lot of readily available information (data) that can help improve measurement and thus provide more precise answers to these questions.

This paper is based upon one central fact: That any multi-country input-output dataset - the data typically used to study GVCs - is consistent with many different GVC networks. With this idea in hand, the paper makes four main contributions - with one section devoted to each. First, I develop a general theory of GVCs that can be used to compare how different theories of production prescribe the way in which GVCs should be constructed from input-output data. Second, I show that an infinite number of microfounded GVC models can perfectly replicate any given input-output database and that each delivers different counterfactual predictions to any shock such as a NAFTA trade war. Third, I show that any given database is also consistent with a wide range of values for any measure used to study globalization such as the U.S. content of imported Mexican goods or the U.S.-China value-added trade balance. Fourth, and finally, I argue that information beyond that contained in input-output tables, such as transaction-level customs data, can be used to complement the former and improve the precision of GVC flow estimates.

I kick off in section 2 by developing a general theory of GVCs that can accommodate, with further assumptions, how different classes of microfounded models behave in equilibrium and what implications they have on GVC flows. For example, the vast majority of trade models with intermediate inputs assume that technology features *roundabout production* in which all of a country-industry's output is produced using the exact same input mix.¹ While specific microfoundations differ substantially, I show that roundabout production models all have the same implications for how GVC flows are mapped across stages of production in equilibrium and thus on how GVC flows should be constructed from input-output data. This class of models thus represent one set of assumptions that can be imposed in order to further restrict the above general theory of GVCs. However, there are many other ways of doing so.

¹Roundabout models come in many varieties, some examples include [Krugman and Venables \(1995\)](#), [Eaton and Kortum \(2002\)](#), [Balistreri et al. \(2011\)](#), [di Giovanni and Levchenko \(2013\)](#), [Bems \(2014\)](#), [Caliendo and Parro \(2015\)](#), [Ossa \(2015\)](#), [Allen et al. \(2017\)](#).

I argue in favor of the class of models featuring *specialized inputs* - models in which goods sold to different countries or industries are built with different input mixes and value-added shares. While the class of roundabout models prescribe a single way of constructing a GVC network from any given input-output database, imposing the specialized inputs assumptions on the general GVC theory has a different set of implications in that an infinite number of GVC networks can be constructed from the same data. More formally, I show that roundabout models prescribe that GVCs should be constructed recursively using first-order Markov chains while specialized inputs models prescribe higher-order Markov chains.

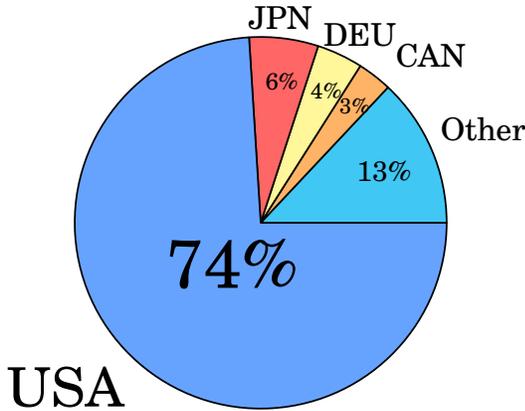
Specialized inputs are consistent with the modern supply chains in which firms make complex decisions when deciding where to source inputs at each stage of production and where intermediate input suppliers customize their goods to be compatible with only specific downstream uses. For example, the lithium battery supplier in Apple's famously long iPod supply chain manufactures it exactly to the size of the metal frame while the screen supplier ensures that the touch, color, and dimming capabilities are in line with Apple's iOS software (Linden et al. 2011). Today, this form of input compatibility is ubiquitous (Rauch 1999, Nunn 2007, Antràs and Staiger 2012, Antràs and Chor 2013) and implies that the type of inputs used to produce exports, at the country-industry level, should vary depending on use of output since firms exporting to different countries and industries have different supply chains.² Indeed, figure 1 takes a first peek at Mexican customs microdata and shows the distribution of foreign inputs used in Mexico to produce car exports to the U.S. and Germany. While roundabout models assume that these input mixes should be exactly the same, in reality the firms exporting cars from Mexico to the U.S. tend to have supply chains using U.S. car parts heavily while the firms exporting to Germany use much fewer U.S. car parts.

Having made the case for specialized inputs, section 3 then shows that quantitative counterfactual predictions depend crucially on the GVC network. For the sake of clarity, and at the cost of generality, I illustrate this point with the simplest structural model as given by an extension of a perfect competition Armington model to specialized inputs - basically, an Armington model where each country-industry produces a specific variety for each market. I show that there is an infinite ways of parameterizing this model so that it perfectly replicates any input-output dataset. Moreover, the GVC network matters quantitatively since the welfare gains from trade depend on the expenditure share on domestic inputs used for the production of domestically sold goods. These two points imply that mapping the model to different GVC networks produces different counterfactual estimates following any economic shock - even though all parameterizations replicate the same input-output data in the benchmark equilibrium. In particular, this model nests the special case of roundabout production in which the gains from trade are given by the formula in Arkolakis et al. (2012) and where any counterfactual exercise delivers a single point estimate.³

²Various recent studies suggest that the use of inputs, within country-industries, depend on the downstream use of output. For example, within-industry exports vary across destinations due to quality (Bastos and Silva 2010), trade regime (Dean et al. 2011), and credit constraints (Manova and Yu 2016). Likewise, the use of imports varies across firm size (Gopinath and Neiman 2014, Blaum et al. 2017a, 2017b, Antràs et al. 2017), multinational activity (Hanson et al. 2005), firm capital intensity (Schott 2004), and the quality of output (Fieler et al. 2017). Further, recent research has made explicit connections between imports and exports through quality linkages (Bastos et al. 2018), trade participation (Manova and Zhang 2012), and rules-of-origin (Conconi et al. 2018). Finally, production processes vary also in terms of the intensity of labor inputs. Processing trade firms export lower-cost labor assembly goods (De La Cruz et al. 2011, Koopman et al. 2012) while firms exporting to richer countries hire higher-skilled workers (Brambilla et al. 2012, Brambilla and Porto 2016). Thus, value-added shares also differ depending on the use of output.

³The Arkolakis et al. (2012) formula does not apply with specialized inputs since the partial elasticity of relative imports from

Foreign Inputs in Exports to the U.S.



Foreign Inputs in Exports to Germany

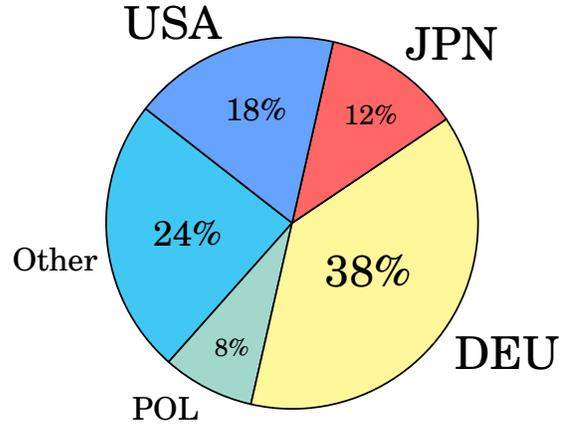


Figure 1: Distribution of Foreign Inputs Used In Mexican Final Good Motor Vehicle Exports to the U.S. and Germany: The shares are constructed using Mexican customs shipment-level data for 2014; details are discussed in section 2.4.1. In contrast to these charts, the conventional approach for constructing GVCs assumes common input distributions across destinations.

Further, I show how to use specialized inputs models to construct bounds on any counterfactual estimate - i.e. finding the upper and lower bounds across all GVC networks that replicate the input-output data - and show, empirically, that the bounds tend to be quite wide. For example, using the World Input-Output Database (WIOD), I find that a NAFTA trade war where both the U.S. and Mexico increase trade barriers on each other by 50% delivers a welfare cost to the U.S. between 0.19-0.29% of real income while the cost from Mexico lies between 3.25-4.66%. In contrast, the special case of roundabout production predicts 0.26% and 3.68%, respectively. The goal of this exercise is not to provide highly credible numbers - since the Armington model is highly simplistic - but rather to illustrate, first, the fact that any dataset is consistent with a range of counterfactual values and, second, that one can construct bounds on them. I conjecture that future work with richer, more credible, microfoundations will yield similar qualitative implications.⁴

Analogous to the case of counterfactuals, section 4 shows that measures of globalization - that is, measures seeking to quantify the current fragmentation of production such as value-added trade (Johnson and Noguera 2012, Koopman et al. 2014) or average downstreamness (Antràs et al. 2012) - also depend crucially on the GVC network. In particular, in current practice, papers tend to define these measures directly in terms of input-output analysis (Leontief 1941). In contrast, I define these measures more broadly using the general theory of GVCs described in section 2 and show that the measures given by input-output analysis are special cases obtained by further imposing the roundabout assumptions on the general theory. This approach is useful because the measures defined with input-output analysis are not consistent with

two sources depends on third country trade costs. Hence, the macro-level restriction “the import demand system is CES” fails.

⁴Examples of richer specialized inputs microfoundations include Yi (2010), Costinot et al. (2012), Antràs and Chor (2013), Fally and Hillberry (2016), Johnson and Moxnes (2016), Antràs and de Gortari (2017), and Oberfield (2018).

a world of specialized inputs, but the correct measures can be derived by instead imposing the specialized inputs assumptions on the general theory. Hence, the general theory provides a common framework for comparing measures of globalization across different equilibrium theories of production.

I then show how to construct approximate bounds on any measure of globalization and use the WIOD for 2014 to illustrate that, in practice, value-added trade might be severely mismeasured. For example, a set of statistics often cited during the NAFTA debate were the share of U.S. value-added returning home through Mexican final good imports.⁵ Higher shares are typically interpreted as proxying a higher cost of supply chain disruption - i.e. when the U.S. hurts Mexico, this will ripple back and hurt the U.S. more when it provides more value to these supply chains - and conventional estimates (based on the roundabout approach) predict that about 17% of the value-added in Mexican manufacturing imports corresponds to U.S. value returning home. In contrast, I show that the same input-output data is consistent with bounds as low as 6% and as high as 47%. In other words, in reality Mexican-American supply chains may be much less or much more integrated depending on how one constructs GVCs. In a second exercise, I focus on the U.S.-China trade deficit. As famously shown by [Johnson and Noguera \(2012\)](#), the value-added deficit based on the roundabout approach is lower than the gross deficit since China re-exports a lot of U.S. value-added back to the U.S. With specialized inputs, however, I find that the upper bound is given by a value-added surplus (i.e. China re-exports a lot more U.S. value-added) but the lower bound is given by a much higher value-added deficit (thus, potentially, reversing the key finding in [Johnson and Noguera 2012](#)).

In sum, the key message of sections 2, 3, and 4, is that any given input-output dataset is consistent with many GVC networks and that the specialized inputs approach together with the bounds approach is useful for determining the potential mismeasurement that trickles over from the GVC flows to quantitative counterfactual estimates and measures of globalization. In other words, since all GVC networks aggregate up to the same bilateral flows they all exhaust the information contained in the input-output data. That is, there is no further information that can shed light on which specific point estimates - be it the roundabout point estimates, the upper or lower bound, or any point contained therein - are most accurate.

Finally, I devote section 5 to improving measurement, in the sense of obtaining the best informed guess of the true GVC network underlying input-output data, by incorporating information from other sources. Specifically, I propose an optimization problem that searches over all GVCs consistent with specialized inputs and a given input-output dataset and chooses the network closest to a set of targets chosen by the researcher. This framework is useful when a researcher has information about the underlying GVC network that is informative but insufficient for constructing the GVC network directly. For example, in some countries, there exist datasets including the universe of firm-to-firm transactions in which case the microdata is sufficient for measuring GVC flows directly.⁶ More often, however, a researcher has some information such as Mexican customs data that is not sufficient for measuring GVCs directly since it contains no data on domestic transactions. In such a case, the researcher can take a stand on how to use the Mexican customs data - i.e. her best informed guess about Mexican GVC flows - by disciplining the targets

⁵For example, U.S. Secretary of Commerce Wilbur Ross argued in the Washington Post (September 21, 2017) that disrupting Mexican-American supply chains was not worrisome since Mexican imports contained 'only' 16% of U.S. value-added (in 2011).

⁶In reality, no dataset truly covers the universe of domestic and foreign transactions (in the sense of including data on every single buyer/seller transaction). However, a Belgian dataset comes quite close; see [Tintelnot et al. \(2017\)](#) and [Kikkawa et al. \(2017\)](#).

with this information, and then using the optimization problem to fill in the missing pieces by ensuring that the Mexican GVC flows aggregate up to the input-output data.

I illustrate this approach by using the Mexican customs data to construct the input mix used in each Mexican manufacturing industry for exports to each destination and industry. In order to use the data, I take the stand that Mexico only does processing trade - that is, that imported intermediate inputs are used exclusively for producing exports. While a strong assumption, it is not too far-fetched since processing trade is widely prevalent in Mexico (De La Cruz et al. 2011). Then, by combining this microdata with the input-output data or, rather, by disciplining GVC flows with this informed best guess, I am able to find that the share of U.S. value-added in Mexican manufacturing imports returning home is not 17% as given by the roundabout model but rather around 30%. Hence, Mexican-American supply chains are much more integrated than suggested by conventional estimates.

This measurement framework is quite flexible and can be used to incorporate various forms of information into GVC measurement in future work. To begin, my analysis using Mexican customs data does not incorporate information from other countries but can readily be extended to the latter case by exploiting the information in firm-level datasets across countries. Further, the approach can also incorporate more abstract forms of information since it only requires that a researcher take a stand on how to use this information to discipline the optimization targets. For example, rules of origin are widely believed to shape supply chains in the NAFTA region (Conconi et al. 2018) - and so even if a researcher had no customs-level data to support this claim, she could obtain a GVC network with highly integrated NAFTA supply chains by mapping the (abstract) rules-of-origin information into targets of the optimization problem.

From a history of science standpoint, this paper is inspired by Samuelson (1952) which asked how to measure bilateral trade flows in the presence of only aggregate export data. This paper takes the same idea to the next iteration: How to measure GVC flows in the presence of only bilateral input-output data? From a philosophy of science standpoint, this paper is inspired by Popper (1959) and argues for a falsifiable approach to GVC measurement. That is, instead of imposing the theoretically-based roundabout approach outright, I argue in favor of studying GVCs under initially broad sets of plausibly accurate GVCs obtained through specialized inputs and to then refine these estimates as more information becomes available.

2 The Hunt for GVCs: The Challenge

This section provides the GVC framework to be used throughout the rest of the paper to discuss counterfactuals, measures of globalization such as value-added trade, and measurement in a GVC world. I proceed in four steps. First, I describe the data contained in a multi-country input-output table. Second, I develop a general theory that provides notation and a unifying framework for comparing different theories of production. I will argue in the rest of the paper that this framework is also useful for deriving explicitly the connection between the literature on structural models and counterfactuals and the literature on measures of globalization such as value-added trade. Third, I discuss three specific theories of production, two widely used theories given by a world of 'only trade in final goods' and a world of roundabout production and a third more modern theory which I will call specialized inputs, and show how they relate through the use

of the general GVC theory. Fourth, and finally, I argue empirically in favor of the theory of specialized inputs using Mexican customs shipment-level data and U.S. domestic input-output tables.

2.1 Multi-Country Input-Output Data

Let \mathcal{J} denote both the set and number of countries and \mathcal{K} the set and number of industries. I define $\mathcal{S} = \mathcal{J} \times \mathcal{K}$ as the set and number of country-industries, with a generic element $s \in \mathcal{S}$ being a country-industry denoted as $s = \{j, k\}$ with $j \in \mathcal{J}$ and $k \in \mathcal{K}$. Multi-country input-output datasets typically contain data on bilateral intermediate input flows across two country-industry pairs, with $X(s', s)$ the dollar value of intermediate inputs sold from country-industry s' to country-industry s , and final good flows between a country-industry and consumers, with $F(s', j)$ the dollar value of final goods sold from country-industry s' to consumers in country j . These are the basic building blocks from which all other aggregate moments are built. For example, gross output of country-industry s' equals

$$GO(s') = \sum_{s \in \mathcal{S}} X(s', s) + \sum_{j \in \mathcal{J}} F(s', j),$$

while its gross domestic product equals

$$GDP(s') = GO(s') - \sum_{s \in \mathcal{S}} X(s, s').$$

There are currently various sources of multi-country input-output datasets such as those produced by the World Input-Output Database Project (WIOD), the Global Trade Analysis Project (GTAP), the Institute for Developing Economies (IDE-JETRO), the Eora Global Supply Chain Database (Eora MRIO), and the OECD Inter-Country Input-Output Tables (ICIO). Each dataset has its own advantages and limitations and the analysis in this paper can be readily applied to each. I will focus throughout on the WIOD, which is the most widely used dataset by the international trade literature, and which is available in its 2016 release for $\mathcal{J} = 44$ countries, $\mathcal{K} = 56$ industries (17 in manufacturing), and for the years 2000-2014.

2.2 A General GVC Theory

I introduce notation that centers attention on GVCs as the central objects of interest instead of parting directly from a specific theory of production. I argue that this general theory can accommodate almost any specific theory of production if one imposes additional assumptions and that this approach is useful since it provides a unifying framework for comparing the implications of each specific theory.

GVC flows comprise the key building blocks of this theory. Let $\mathcal{G}(\cdot)$ denote the dollar value of production flowing through an initial node across a specific ordered set of country-industries all the way to final consumption. To fix ideas, I begin by describing notation assuming a single industry world, that is $\mathcal{S} = \mathcal{J}$. Take three countries $j, j', j'' \in \mathcal{J}$. Then $\mathcal{G}(j', j)$ denotes the dollar value of final consumption goods that j' sells to j while $\mathcal{G}(j'', j', j)$ is the dollar value of intermediate inputs that j'' sells to j' which j' uses as inputs for goods then sold as final consumption to j .

More generally, intermediate inputs may be traded at a stage of production that is $N \in \mathbb{N}$ stages upstream relative to the production of final consumption goods, and I will write a generic truncated GVC flow as $\mathcal{G}^N(j^N, j^{N-1}, \dots, j^1, j)$. The superscript N on $\mathcal{G}^N(\cdot)$ indicates the dimension of this function, i.e. N is the number of nodes previous to final consumption that are specified. Every node corresponds to a country with $j^n \in \mathcal{J} \forall n$ and the n is only meant to indicate the dimension for which country j^n is relevant. The flow $\mathcal{G}^N(j^N, j^{N-1}, \dots, j^1, j)$ thus indicates the dollar value of inputs from j^N sold to j^{N-1} , that j^{N-1} uses to produce new inputs sold to j^{N-2} , so on and so forth, until the goods arrive at j^1 and are put into final goods shipped and sold to consumers in j . Since using apostrophes is cumbersome with large N , in general I will use the notation $\mathcal{G}^1(j^1, j)$ instead of $\mathcal{G}(j', j)$ and likewise $\mathcal{G}^2(j^2, j^1, j)$ instead of $\mathcal{G}(j'', j', j)$.

The extension to a multi-industry world is immediate. Let \mathcal{K} be the set of sectors and $\mathcal{S} = \mathcal{J} \times \mathcal{K}$ be the set of country-sectors. GVCs can be defined generically as follows.

Definition 2.1. For any length $N \in \mathbb{Z}^+$, $\mathcal{G}^N : \mathcal{S}^N \times \mathcal{J} \rightarrow \mathbb{R}^+$ is the function describing truncated GVC flows leading to final consumption in countries in \mathcal{J} through a sequence of N upstream stages of production given by an element of \mathcal{S}^N .

A generic GVC is $\mathcal{G}^N(s^N, \dots, s^1, j)$ and, as before, I refer to the elements of a country-industry pair as $s^n = \{j^n, k^n\}$ with $j^n \in \mathcal{J}$ the country and $k^n \in \mathcal{K}$ the industry of $s^n \in \mathcal{S}$, where the n is only meant to indicate the dimension of $\mathcal{G}^N(\cdot)$ for which s^n is relevant. For example: a flow of length $N = 1$ could be $\mathcal{G}^1(s^1, j) = \mathcal{G}^1(\{\text{Mexico, cars}\}, \text{U.S.})$, the sales of Mexican cars to U.S. consumers, while a flow of length $N = 2$ could be $\mathcal{G}^2(s^2, s^1, j) = \mathcal{G}^2(\{\text{U.S., steel}\}, \{\text{Mexico, cars}\}, \text{U.S.})$, the sales of U.S. steel in the form of intermediate inputs that are used exclusively by the Mexican car industry to produce final goods sold to U.S. consumers. Analogously for any $N \in \mathbb{Z}^+$ and any sequence of production in \mathcal{S}^N that produces a final good eventually sold to consumers in some country in \mathcal{J} .

The crucial challenge embedded in this theory of GVCs is that the word *truncated* appears in Definition 2.1. Specifically, $\mathcal{G}^N(\cdot)$ is a truncated GVC because it only specifies the flow through N stages of production even though its most upstream stage, s^N , also uses inputs and the full chain of production is characterized by a (potentially) infinite number of stages of production. The challenge is thus to develop a theory of production that links GVC flows across different stages of production. That is, take an arbitrary GVC $\mathcal{G}^N(s^N, s^{N-1}, \dots, s^1, j)$. Since this tells how many inputs are sold from s^N to the sequence $s^{N-1} \rightarrow \dots \rightarrow s^1 \rightarrow j$ then there has to be some relation with the GVC flow $\mathcal{G}^{N-1}(s^{N-1}, \dots, s^1, j)$ denoting the inputs that s^{N-1} itself sells to this sequence of production.

In its most general form, the only restriction I impose is that flows across different stages of production must satisfy

$$\sum_{s^N \in \mathcal{S}} \mathcal{G}^N(s^N, s^{N-1}, \dots, s^1, j) \leq \mathcal{G}^{N-1}(s^{N-1}, \dots, s^1, j). \quad (1)$$

That is, the right-hand side denotes the value of intermediate inputs sold by s^{N-1} to be used through the sequence in $\mathcal{G}^{N-1}(s^{N-1}, \dots, s^1, j)$. The left-hand side denotes the total value of intermediate inputs, across all sources $s^N \in \mathcal{S}$, sold to s^{N-1} and used down this same sequence of production. Imposing equation (1) thus implies that the total value of inputs purchased by s^{N-1} for a specific downstream sequence of production need be less or equal than the value of the output that s^{N-1} itself produces for that sequence.

This theory is general and can encompass most production processes. It relies only on the key restriction that the dollar value of output not fall as goods flow down the value chain. Whenever the value of output increases, which implies that equation (1) holds with strict inequality, I will say that value was added at the $N - 1$ th stage of production to the inputs purchased from stage N by $\mathcal{G}^{N-1}(s^{N-1}, \dots, s^1, j)$. For example, this theory assumes that

$$\sum_{s^2 \in \mathcal{S}} \mathcal{G}^2(s^2, \{\text{Mexico, cars}\}, \text{U.S.}) \leq \mathcal{G}^1(\{\text{Mexico, cars}\}, \text{U.S.}).$$

The right-hand side indicates the dollar value of Mexican cars sold to U.S. consumers and corresponds to a truncated GVC flow because the Mexican car industry uses intermediate inputs produced further upstream to produce these cars. Meanwhile, $\mathcal{G}^2(\{\text{U.S., steel}\}, \{\text{Mexico, cars}\}, \text{U.S.})$ is the dollar value of U.S. steel bought as inputs directly in order to produce these exports, so that the summation across all possible input sources $s^2 \in \mathcal{S}$ yields aggregate input sales to the downstream sequence on the right-hand side. The inequality holds strictly if the Mexican car industry adds domestic value-added directly into the intermediate inputs purchased from the previous stage of production.

I refer to equation (1) as the *GVC challenge* which can only be solved by imposing a theory of production linking GVC flows across different stages of production. In other words, solving the GVC challenge requires taking a stand on how $\mathcal{G}^N(s^N, s^{N-1}, \dots, s^1, j)$ and $\mathcal{G}^{N-1}(s^{N-1}, \dots, s^1, j)$ relate to each other across all stages and sequences of production. Throughout the rest of the paper I will restrict attention to a static world in which all goods are produced simultaneously in a single period since this is how input-output data is typically interpreted in the GVC literature. However, in future work this framework could potentially incorporate an extension with dynamic production since one can interpret s as a triple of country-industry-time period in which inputs of past periods flow down the value chain to be used as inputs in future periods.

2.2.1 Relation to Multi-Country Input-Output Data

Needless to say, $\mathcal{G}^N(s^N, \dots, s^1, j)$ is not observed directly in input-output tables. If these flows were observed, then the GVC challenge in equation (1) would be solved trivially since it would rely only on measuring the objects of interest directly. This does not mean that input-output data is useless, but rather that it only contains some (non-exhaustive) information about how to solve the GVC challenge in equation (1). I now show how the GVC flows relate to this data.

The first thing to note is that input-output data typically provides perfect information about the very last stage of production, namely that of final good production. This implies that the simplest GVC flows, those with $N = 1$, are observed. Indeed, bilateral final good flows can be defined in terms of GVCs as

$$F(s', j) = \mathcal{G}^1(s', j). \tag{2}$$

This mapping is the basic building block from which all theories of intermediate input trade will build upon since this is the only part of the supply chain that is observed directly in input-output data.

Second, bilateral intermediate input flows are much more complicated since they aggregate the dollar

value of inputs traded across two country-industries across all stages of the supply chain. The relation between these aggregate flows and GVC flows is given by

$$X(s', s) = \sum_{N=2}^{\infty} \sum_{s^{N-2} \in \mathcal{S}} \cdots \sum_{s^1 \in \mathcal{S}} \sum_{j \in \mathcal{J}} \mathcal{G}^N(s', s, s^{N-2}, \dots, s^1, j). \quad (3)$$

The flow $\mathcal{G}^N(s', s, s^{N-2}, \dots, s^1, j)$ is the value of inputs from s' sold to s at the N th stage of production and to be used through a specific downstream sequence of production. Summing up across $s^{N-2} \in \mathcal{S}, \dots, s^1 \in \mathcal{S}, j \in \mathcal{J}$ thus delivers the aggregate value of inputs from s' sold to s at the N th stage of production used across all downstream sequences of production. The first summation across $N \geq 2$ then sums up the value of inputs traded across all stages of the supply chain. This aggregate value is thus what is typically reported in input-output tables.

Specific theories of production provide a guide for disentangling GVC flows across different stages of production while taking into account the observed input-output data. That is, since $\mathcal{G}^1(s', j)$ is observed, any theory of production needs to take a stand on how to disentangle the GVC flows $\mathcal{G}^N(s^N, \dots, s^1, j)$ for all $N \geq 2$ taking into account the restrictions in equation (1) and the fact that a lot of the information is potentially lost in the aggregation into bilateral intermediate input flows in equation (3).

2.3 Two Old Solutions, and One New One

I now discuss three possible solutions to the GVC challenge. The first corresponds to a theory in which only final goods are traded, this was the standard GVC theory a few decades ago. The second corresponds to the benchmark GVC theory used presently, both in structural models and in terms of literature on measures of globalization, which incorporates intermediate inputs albeit in a very simplified manner. The third corresponds to a new solution which takes seriously the fact that, in reality, supply chain networks are very complex and that most of this richness is often lost through the aggregation into input-output data. Each subsequent solution is more general and nests the previous one.

Importantly, the theories of production I now discuss require only specifying how the theory behaves in equilibrium since the input-output data is always interpreted as an equilibrium in the real world. In the rest of the paper, thinking about counterfactuals will require digging deeper into the specific microfoundation underlying these solutions since one needs to account for how the equilibrium changes after a given shock. In contrast, for the purposes of both studying measures of globalization and improving measurement, though, specifying how the theory behaves in equilibrium is sufficient.

2.3.1 The ‘Only Trade in Final Goods’ Solution

The simplest solution to disentangling GVCs in (1) is to assume that GVC linkages are non-existent. In other words, parting from the observed GVCs $\mathcal{G}^1(s^1, j)$ assume that the mapping into previous stages of production at $N \geq 2$ is given by

$$\mathcal{G}^N(s^N, s^{N-1}, \dots, s^1, j) = 0, \quad (4)$$

across any sequence of production. Substituting into the definition of intermediate input flows in equation (3) implies that for all country-industry pairs s' and s

$$\Rightarrow X(s', s) = 0,$$

since intermediate inputs are not used at any stage of production. Hence, this system of GVC flows implies that every dollar of final output is made up entirely of domestic value-added created at the most downstream stage of production

$$\Rightarrow GO(s') = GDP(s') = \sum_{j \in \mathcal{J}} \mathcal{G}^1(s', j) = \sum_{j \in \mathcal{J}} F(s', j).$$

These assumptions are extreme and at odds with today's global economy since over two-thirds of world trade is in intermediate inputs. Indeed, most datasets report $X(s', s) > 0$ for the majority of country-industry pairs so that this GVC characterization cannot be squared with current data. However, these restrictions were still widely imposed even a few decades ago. For example, both the modern version of the Armington model developed by [Anderson \(1979\)](#) and the classical Ricardian model of [Dornbusch et al. \(1977\)](#) assume that intermediate inputs play no role so that both models can be characterized, in equilibrium, by this GVC representation. While this paradigm is less prevalent today, it is an useful starting point for showing how to map more complex theories of international trade into the above GVC framework.

2.3.2 The Roundabout Solution

Disentangling GVCs in the presence of intermediate input trade is much more complex since, in principle, there are many theories of production that can solve the GVC challenge in equation (1). This observation motivates this paper since, so far, the literature has largely focused on a solution in which every single dollar of output within each country-industry is produced using the exact same input mix. Formally, the mapping is solved by assuming the existence of a set of technical coefficients $\alpha(s' | s)$ denoting the share of inputs from s' used by s to produce output at any stage of production and for any sequence of production. Parting from the observed GVCs, $\mathcal{G}^1(s^1, j)$, the mapping into previous stages of production at $N \geq 2$ is given by

$$\mathcal{G}^N(s^N, s^{N-1}, \dots, s^1, j) = \alpha(s^N | s^{N-1}) \mathcal{G}^{N-1}(s^{N-1}, \dots, s^1, j). \quad (5)$$

Rearranging, any GVC flow can thus be characterized entirely by final good flows and the technical coefficients as

$$\Rightarrow \mathcal{G}^N(s^N, s^{N-1}, \dots, s^1, j) = \prod_{n=2}^N \alpha(s^n | s^{n-1}) F(s^1, j). \quad (6)$$

Substituting into the relation between GVC flows and intermediate input flows in (3), it is straightforward to see that the following relation holds

$$\Rightarrow X(s'', s') = \alpha(s'' | s') \left(\sum_{s \in \mathcal{S}} X(s', s) + \sum_{j \in \mathcal{J}} F(s', j) \right).$$

In other words, since $X(s', s)$ and $F(s', j)$ are observed in input-output data, this theory of production can only be squared with the data if the technical coefficients are given by

$$\Rightarrow \alpha(s' | s) = \frac{X(s', s)}{GO(s)}. \quad (7)$$

The expenditure by s on inputs from s' is simply given by aggregate value of inputs purchased from s' relative to gross output. Since gross output is typically larger than aggregate intermediate input purchases, this implies that every dollar of output of s has a share of domestic value-added given by

$$\Rightarrow \beta(s) = 1 - \sum_{s' \in \mathcal{S}} \alpha(s' | s) = \frac{GDP(s)}{GO(s)}. \quad (8)$$

The roundabout solution has two very useful properties. First, it incorporates intermediate inputs, whereas the previous ‘only trade in final goods’ solution did not. Second, it is so highly tractable that the measurement problem regarding how to disentangle GVCs is completely eliminated as long as one has input-output data at hand. Roundabout production implies that any GVC flow in (6) is characterized by final good flows and the technical coefficients, but since the latter are characterized by input-output data as well in (7), then any GVC flow is fully and uniquely characterized by input-output data.

In other words, any microstructure with roundabout production has GVCs that can be characterized, in equilibrium, by the mapping in (6), and this is equivalent, in terms of measurement, to input-output analysis. As is well known, the latter is a measurement framework that leaves no degrees of freedom open to the researcher since it is fully characterized by input-output data. Importantly, though, while input-output analysis is often defined directly as a set of input and value-added shares given by (7) and (8), I derived these input shares from first principles in the sense that I imposed assumptions on the mapping of GVCs across different stages of the supply chain in (5) and then derived the input shares as an implication.⁷ This latter approach is more useful since it parts from a general theory of GVCs, consistent with many different theories of production, and in which different measurement frameworks can be contrasted in terms of the additional assumptions imposed in order to disentangle GVCs from the observable data.

2.3.3 The Specialized Inputs Solution

The specialized inputs solution departs from the roundabout solution and assumes that the use of inputs depends on the destination of output and the use of output, both in terms of whether goods are sold as final goods or intermediate inputs and to which industry they are sold to as inputs in the latter case. The GVC flow of inputs used directly for the production of final goods is thus given by

$$\mathcal{G}^2(s^2, s^1, j) = \alpha_F(s^2 | s^1, j) F(s^1, j), \quad (9)$$

⁷Input-output analysis is typically described using matrix algebra. Imposing the GVC mapping (5) on the definition of bilateral intermediate input flows in (3) and using matrix algebra implies that

$$\Rightarrow \mathbf{X} = \mathbf{A}\mathbf{F} + \mathbf{A}^2\mathbf{F} + \dots = \mathbf{A}[\mathbb{I} - \mathbf{A}]^{-1}\mathbf{F},$$

where $\mathbf{GO} = [\mathbb{I} - \mathbf{A}]^{-1}\mathbf{F}$ is gross output and $[\mathbb{I} - \mathbf{A}]^{-1}$ is known as the Leontief inverse matrix.

where $\alpha_F (s'' | s', j)$ is the share of inputs from country-industry s'' used in the final goods produced by country-industry s' that are sold to consumers in market j . Analogously, the use of intermediate inputs in the production of new intermediate inputs is given by

$$\mathcal{G}^{N+1} (s^{N+1}, s^N, \dots, s^1, j) = \alpha_X (s^{N+1} | s^N, s^{N-1}) \mathcal{G}^N (s^N, \dots, s^1, j), \quad \forall N \geq 2, \quad (10)$$

where $\alpha_X (s'' | s', s)$ is the share of inputs from country-industry s'' used in the production of intermediate inputs by country-industry s' sold to country-industry s . Note that while the intermediate input shares depend on the destination and use of inputs, they are common across all stages of production. That is, the input mix used to produce inputs in s' and sold to s is the same in all production stages $N \geq 2$.

In this context, value-added shares also depend on the destination and use of output and are given by

$$\begin{aligned} \beta_F (s', j) &= 1 - \sum_{s'' \in \mathcal{S}} \alpha_F (s'' | s', j) \geq 0, \\ \beta_X (s', s) &= 1 - \sum_{s'' \in \mathcal{S}} \alpha_X (s'' | s', s) \geq 0. \end{aligned}$$

These shares have to be greater or equal than zero given the assumption in (1) that the dollar value of output never falls as goods flow along the value chain. Further, at least one of these shares has to be strictly positive since GDP in every country-industry s' in the data is positive.

Relative to the roundabout solution, now it is not possible to characterize input shares directly using the observable data. Rather, this GVC solution is richer than the input-output data since there are many different sets of input shares that perfectly fit this data (i.e. there are many supply chain networks that replicate the same bilateral trade, gross output, and gross domestic product flows). This can be seen by substituting in the specialized inputs solution in (9) and (10) into the set of linear constraints relating GVC flows to the observed input-output data in (3), which delivers

$$X (s'', s') = \sum_{N=2}^{\infty} \sum_{s^N \in s''} \sum_{s^{N-1} \in s'} \sum_{s^{N-2} \in \mathcal{S}} \dots \sum_{s^1 \in \mathcal{S}} \sum_{j \in \mathcal{J}} \left[\prod_{n=3}^N \alpha_X (s^n | s^{n-1}, s^{n-2}) \right] \alpha_F (s^2 | s^1, j) F (s^1, j). \quad (11)$$

Equation (11) is tedious but straightforward and simply sums up the use of inputs sold by s'' to s' across all stages and chains of production. Conditional on N , the first two stages of the sequence of production are $s^N = s''$ and $s^{N-1} = s'$, and I have slightly abused notation by indicating two summations over single-valued sets. The subsequent summations sum up the use of inputs across all downstream sequences of production $s^{N-2} \in \mathcal{S}, \dots, s^1 \in \mathcal{S}, j \in \mathcal{J}$, while the summation over $N \geq 2$ sums up the exchange of inputs at all upstream stages of production.

Fortunately, the recursive structure of the specialized inputs solution assumed in (10) implies that the mapping between input shares and input-output data in (11) can be rewritten much more succinctly as

$$X (s'', s') = \sum_{s \in \mathcal{S}} \alpha_X (s'' | s', s) X (s', s) + \sum_{j \in \mathcal{J}} \alpha_F (s'' | s', j) F (s', j). \quad (12)$$

In words, the right-hand side sums up all the intermediate inputs from s'' used by s' to produce further downstream inputs sold to all $s \in \mathcal{S}$ and final goods sold to all $j \in \mathcal{J}$. Since this is the total value of inputs sold from s'' to s' this has to equal the observed flow $X(s'', s')$.

All of the information in input-output data is contained in $X(s', s)$ and $F(s', j)$. Thus, any set of input shares $\alpha_X(s'' | s', s)$ and $\alpha_F(s'' | s', j)$ satisfying (12) for all bilateral pairs characterize a system of GVC flows that perfectly fit the observable data. Crucially, fitting the data requires imposing $\mathcal{S} \times \mathcal{S}$ restrictions but the specialized inputs GVC network depends on $\mathcal{S} \times \mathcal{S} \times (\mathcal{S} + \mathcal{J})$ input shares. These degrees of freedom imply that there are many different GVC networks that replicate the same observable data.

The roundabout solution is the knife-edge case of specialized inputs in which the input shares for both final and intermediate inputs are assumed to be common and independent of the use or destination of output. That is, in this knife-edge case the share of inputs from s'' used by s' at any stage of production equals

$$\alpha_X(s'' | s', s) = \alpha_F(s'' | s', j) = \alpha(s'' | s'), \quad \forall s \in \mathcal{S} \text{ and } \forall j \in \mathcal{J}.$$

And the restrictions in (12) imply that the input shares fit the observable data if and only if

$$\alpha(s'' | s') = \frac{X(s'', s')}{GO(s')},$$

which are exactly the input shares assumed outright in the roundabout solution or input-output analysis.

2.3.4 Taking Stock

I have discussed three solutions showing how theory can be used to disentangle the GVC challenge described in (1). Each subsequent theory is more general than the previous one and all three are useful for understanding the aggregation issues present in input-output data.

First, a few decades ago, bilateral trade data did not distinguish between intermediate input and final good trade and so, in practice, the data was silent regarding whether the ‘only trade in final goods’ solution was potentially accurate or not. Current input-output datasets, however, show that the majority of world trade is in intermediate inputs and so are now disaggregate enough to be able to reject this GVC theory.

Second, the roundabout solution incorporates intermediate input flows but assumes that the use of inputs is common regardless of the destination or use of output and across all stages of production. This theory is the knife-edge case that fits the data perfectly and in a unique way. However, it also implicitly implies assuming that further disaggregating the data would yield no additional insights or information.

Third, and finally, the specialized inputs solution can fit the data perfectly in many ways and thus implicitly assumes that there is important information that is hidden by the aggregation present in input-output datasets. The rest of the paper is concerned with using the specialized inputs solution to understand the implications of such aggregation in currently available input-output datasets.

As a final comment, note that there are many other potential ways of disentangling the GVC mapping in (1). For example, a richer form of input specialization could be characterized by input shares $\alpha_X(s''' | s'', s', s)$ where the input mix used in s'' for exports to s' is tailored according to the further downstream production stage at s . More formally, this corresponds to building GVCs recursively using

third-order Markov chains while the above specialized inputs and roundabout solutions correspond to the special cases of second-order and first-order Markov chains.⁸ Alternatively, one could assume intermediate input trade but that GVCs cannot be characterized recursively and are instead finite with output at some stage $N > 1$ consisting entirely of domestic value-added produced at that stage. I focus throughout the rest of the paper on the specialized inputs solution since it is, in my view, the most natural and tractable generalization of the roundabout solution. But the reader should keep in mind that the GVC framework in (1) can be used to study many other solutions in future research.

2.4 Evidence for Specialized Inputs

2.4.1 Evidence from Firm-Level Data

I now use customs shipment-level data to study the presence of specialized inputs in the sales of a given country-industry to different export markets. Specifically, I use the universe of Mexican customs data for 2014 to impute the type of inputs used in exports to different markets. I proceed in three steps. First, for each firm I construct its aggregate intermediate input purchases from each country and its aggregate exports to each country. Second, I assume the roundabout solution at the firm-level so that all of a firm's exports have the same input mix.⁹ This lets me obtain a measure of the dollar value of imports from each country used in the exports to each country at the firm-level. Third, I take all of the firms within a manufacturing industry and compute the aggregate value of imports from a given source used in the exports to a given destination within a manufacturing industry. This lets me construct the distribution of foreign inputs used in exports to each destination market - which should be common across markets if the roundabout solution were accurate at the industry-level.¹⁰

Figure 2 confirms the prevalence of specialized inputs in Mexican manufacturing final good exports at the level of aggregation consistent with typical multi-country datasets.¹¹ Specifically, each chart plots the distribution of foreign inputs from the four main suppliers and a rest of world remainder (rows) used in the exports to each of the five main export destinations (columns) in the top nine Mexican manufacturing industries. In other words, the cells across a column represent the distribution of foreign inputs in a specific type of exports and add up to 100%. For example, motor vehicles is Mexico's main export industry and the corresponding chart shows that the use of inputs in exports to the U.S. and Germany, Mexico's main North American and European trade partners, differ substantially (i.e these are the distributions in figure 1).

⁸A previous version of this paper, [de Gortari \(2017\)](#), shows how to disentangle GVCs using Markov chains of any order.

⁹Most firms are multi-product firms and so different inputs are probably used for different exported products within each firm. However, the data has no information on what happens within the firm so this issue cannot be addressed. Having said this, assuming the roundabout solution at the firm-level is a much weaker assumption than assuming it at the industry-level.

¹⁰Customs data does not contain domestic purchases so value-added shares cannot be measured at the firm-level and this analysis also rests on assuming common value-added shares across firms within an industry. Imposing the roundabout solution at the industry-level also assumes this and so, in this respect, this analysis is just as far-fetched as the standard approach.

¹¹The latter is an important point since one could define manufacturing industries at the firm-level and then the distribution of inputs used in exports to different destinations would be common by construction since I have assumed the roundabout solution at the firm-level. However, the charts in figure 2 are presented at the relevant level of aggregation since, for example, manufacturing flows in the WIOD are available for only 17 aggregate manufacturing industries. Going forward, while multi-country datasets are likely to become more disaggregate over time it is unlikely that these datasets become available at a disaggregate enough level to be consistent with the roundabout solution at the industry-level anytime soon.

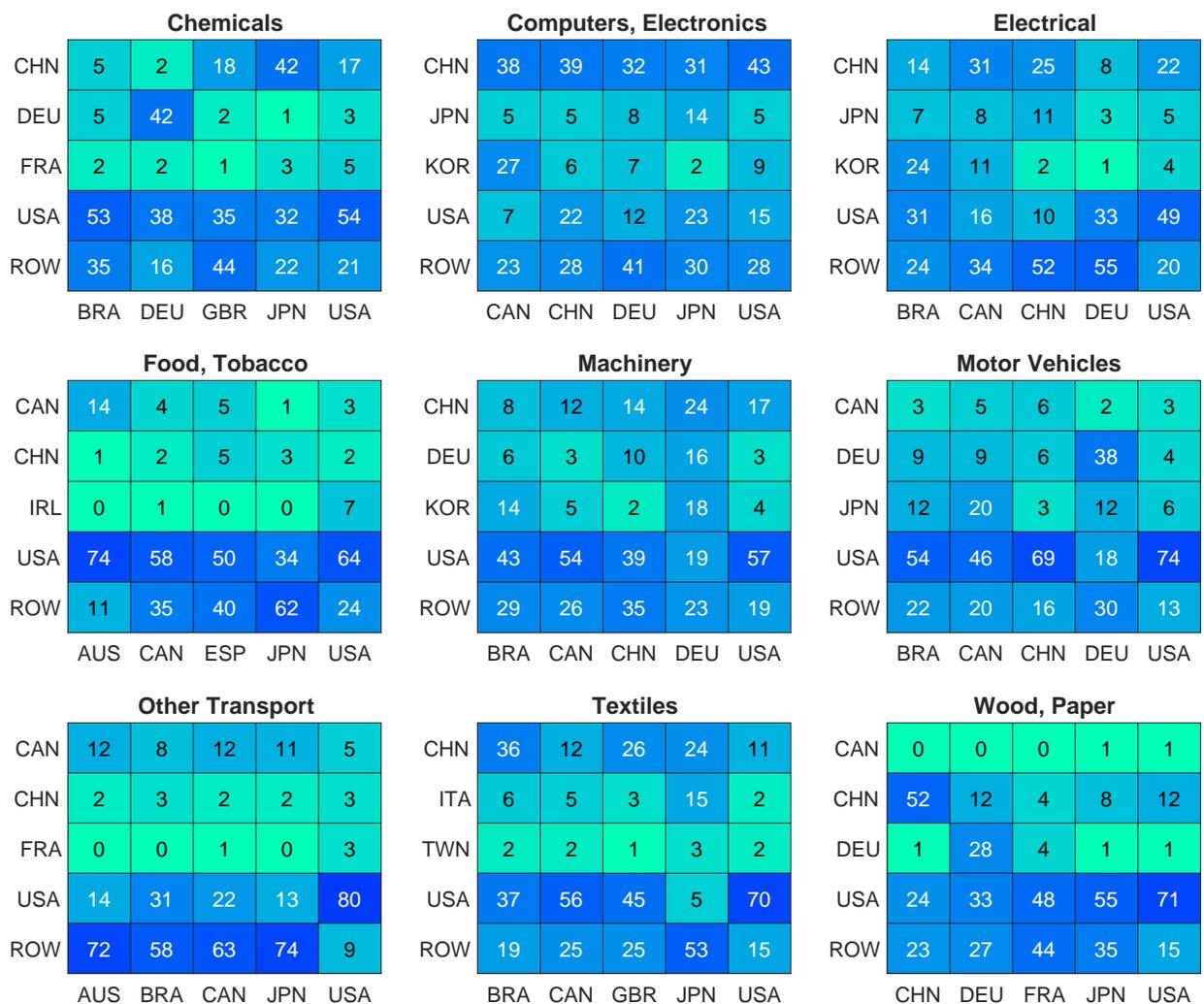


Figure 2: Foreign Input Shares in Mexican Manufacturing Exports Across Destinations: Each chart presents the distribution of foreign inputs from the four main input suppliers and a rest of world remainder (rows) used in the exports shipped to the five main export destinations (columns) for each manufacturing industry (i.e. cells across rows within each column sum up to 100%). These nine manufacturing industries account for 95% of Mexico's final good manufacturing exports. Shares are constructed using Mexican customs shipment-level data assuming that the roundabout solution holds at the firm-level. In contrast to these charts, assuming the roundabout solution at the industry-level implies common input distributions across export destinations.

Overall, figure 2 shows substantial heterogeneity in input shares in sales to different destinations and confirms the fact that NAFTA supply chains are highly integrated. The U.S. is always one of the five main export markets and always one of the top four input suppliers. Furthermore, the U.S. tends to have an outsized role as input supplier in the exports that return to its own market - thus confirming the widely-available anecdotal evidence that Mexico-U.S. trade is based heavily on goods that cross the border back and forth. For example, U.S. inputs account for over 70% of foreign inputs in the exports to the U.S. in four of these nine manufacturing industries and the share is around or above 50% in all but one.

2.4.2 Evidence from Disaggregate Domestic Input-Output Tables

I now use domestic input-output tables to study the presence of specialized inputs in the sales of a given industry to other industries. For example, the computer and electronics industry produces diverse products such as semiconductors, often sold to downstream electronics producers, and navigation instruments, often sold to ship-building companies. However, since computer and electronics is often reported as a whole, GVC estimates based on the roundabout solution assume that the input mix in goods sold to both the downstream electronics and the ship-building industries is the same. Domestic tables can be used to proxy the importance of this type of bias since they are often available at different levels of disaggregation.

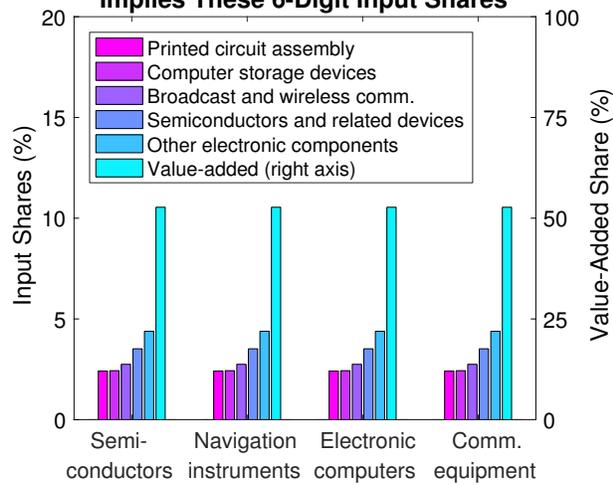
The U.S. Bureau of Economic Analysis reports domestic input-output tables for the year 2007 at a level of disaggregation of both 389 and 71 industrial categories (corresponding to the 6- and 3-digit NAICS classification). I use this data to study the industry aggregation bias through the following thought experiment: I assume the roundabout solution is accurate at the 389 industry-level, so that all output within a given industry is built with the same input mix regardless of what industry it is sold to, and then compare these input shares to the ones implied by the more aggregate data with only 71 industrial categories. If there were no industry aggregation bias, then the input mix used in the production of each of the 6-digit industries bundled into a single 3-digit industry should be the exactly same and equal to the aggregate 3-digit input mix. If not, then there is an aggregation bias because industries with different input mixes are being bundled together — thus breaking the assumption that all output within a given industry, at the 3-digit level, is built with the same input mix, even if it is true at the 6-digit level.

Figure 3 illustrates the industry aggregation bias in the 3-digit computers and electronics industry. Specifically, the latter is composed of 20 6-digit industries, of which the four largest are semiconductors, navigation instruments, electronic computers, and communication equipment. Meanwhile, its five largest input suppliers are printed circuit assembly, computer storage devices, broadcast and wireless communication equipment, semiconductors and related devices, and other electronic components. Figure 3’s left panel shows that imposing the roundabout solution on the 3-digit computers and electronics implies that every product within each 6-digit subindustry is produced with the same input and value-added mix. For example, computer storage devices accounts for 2.7% of the aggregate output value of the 3-digit computers and electronics and is thus also the input share used in all 6-digit subindustries. Figure 3’s right panel, however, uses the more disaggregate 6-digit data to show that input shares vary substantially within each subindustry. For example, computer storage devices are used intensively as inputs in electronic computers (15% of output value) but only marginally in the other three plotted 6-digit industries.

Figure 4 summarizes the industry aggregation bias across all U.S. manufacturing as proxied by the coefficient of variation—standard deviation relative to mean—of input shares from each source within each 3-digit code.¹² In the absence of aggregation bias, there is no heterogeneity in input shares at the 6-digit level and the coefficient of variation is zero. Alternatively, when the aggregation is done across industries with substantial heterogeneity the coefficient of variation is large. Each column in figure 4 corresponds to a

¹²Specifically, for each 3-digit industry $k^{3\text{dig}} \in \mathcal{K}^{3\text{dig}}$ I compute the coefficient of variation of the input shares $\alpha(t|k)$ from a given source $t \in \mathcal{K}^{6\text{dig}}$ and across all the 6-digit subindustries k bundled in $k^{3\text{dig}}$. For example, for the 3-digit industry computers and electronics, figure 4 plots one circle for the coefficient of variation of $\alpha(\text{printed circuit assembly}|k)$ across all 6-digit subindustries indexed by k , and another circle for $\alpha(\text{computer storage devices}|k)$. Analogously, across all 6-digit suppliers $t \in \mathcal{K}^{6\text{dig}}$.

Imposing Roundabout Production on 3-Digit Data Implies These 6-Digit Input Shares



True 6-Digit Input Shares

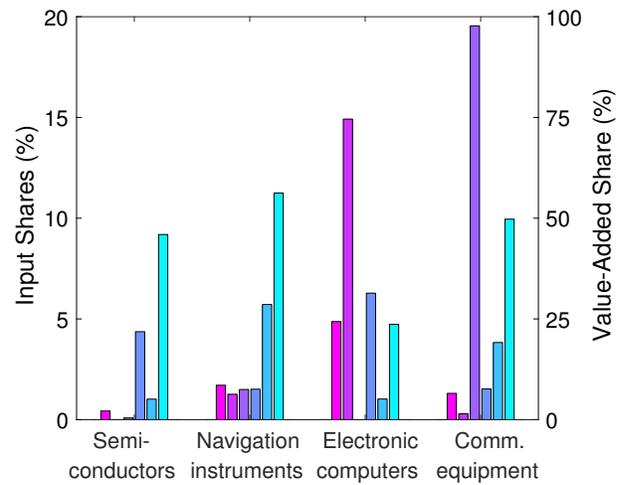


Figure 3: Implied and True Input and Value-Added Shares Within the Computer and Electronics Industry: The left panel plots the input shares, from the top five input suppliers, in the production of the top four subindustries implied by imposing the roundabout solution on the aggregate 3-digit industry computer and electronics while the right panel plots the true shares using the disaggregate 6-digit data. Data is from 2007 U.S. input-output tables from the Bureau of Economic Analysis.

Variation in 6-Digit Input Shares, Within Each 3-Digit Industry

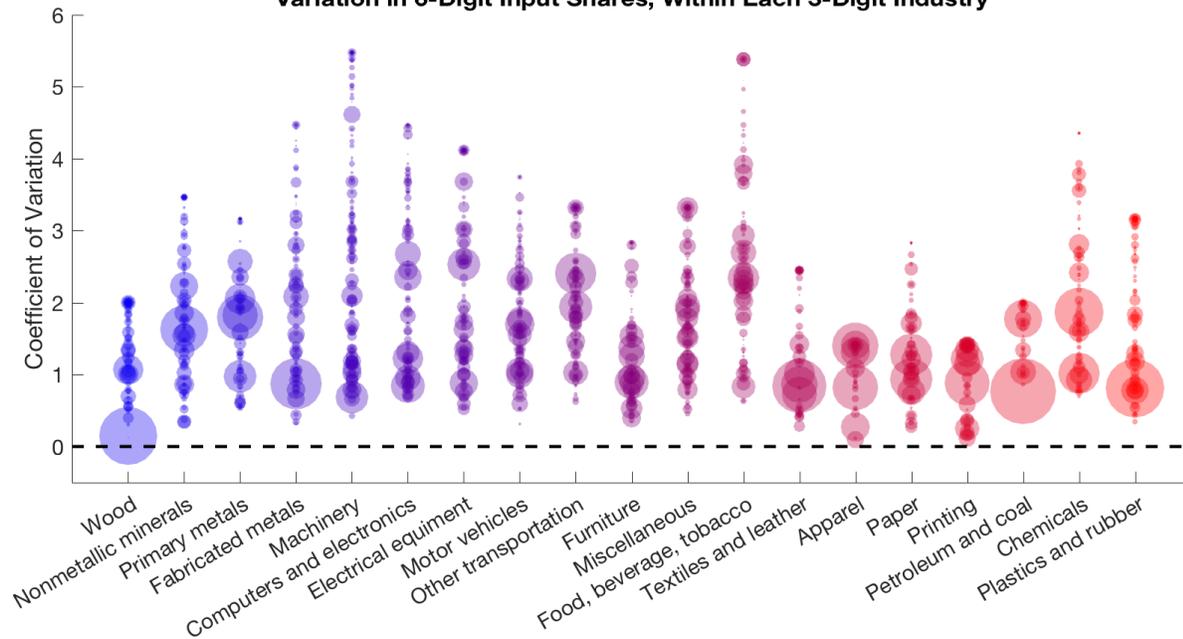


Figure 4: Variation in Domestic Industry Input Shares in U.S. Manufacturing Sales Across Domestic Industries: Each circle corresponds to the coefficient of variation—standard deviation relative to mean—of the input shares from a specific 6-digit input supplier across all 6-digit subindustries within each 3-digit industry. Circle size is proportional to the share of aggregate input purchases by the 3-digit industry from each source. In contrast to this chart, assuming the roundabout solution at the 3-digit industry-level implies zero variation across all 6-digit subindustries. Data is from 2007 U.S. input-output tables from the Bureau of Economic Analysis.

given 3-digit manufacturing industry, with each circle corresponding to the coefficient of variation of input shares from some 6-digit input supplier across the 6-digit subindustries of the 3-digit industry; the size of each circle is proportional to the importance of each input supplier. Figure 4 reveals one key takeaway: There is substantial variation in input shares within each 3-digit sector. For example, for computers and electronics the five biggest circles are those corresponding to input shares from the sources in figure 3. The largest circle corresponds to other electronic components (the most important supplier) and, as figure 3 shows, there is relatively little variation in input shares so the coefficient of variation is 0.8. In contrast, the high variation in computer storage devices visible in figure 3 yields a coefficient of variation of 2.7.

Overall, figure 4 reveals substantial heterogeneity in input shares across sales to different industries and is informative about specialized inputs in multi-country tables since the latter are typically available at an industrial classification level similar to the 3-digit NAICS. Hence, while this exercise cannot be done with multi-country tables, it is likely that the industry aggregation bias is as prevalent as implied by the U.S. domestic tables. In particular, since multi-country tables are based on domestic tables, this implies that assuming common input shares for the U.S. in the latter leads to distorted supply chain networks.¹³

3 GVCs and Counterfactuals

A first strand of the GVC literature is concerned with understanding the implications of economic shocks, such as changes in trade barriers, on international trade. In particular, in a seminal contribution, [Arkolakis et al. \(2012\)](#) (ACR henceforth) argued that, with some assumptions in hand, the welfare gains from trade can be studied across a variety of different microfoundations with a simple formula depending on the change in the domestic expenditure share. Though their benchmark analysis is carried out in a world of ‘only trade in final goods’, they show that their results extend to the world of roundabout production.

This section shows that in richer theories of production, and specifically in models with specialized inputs, the gains from trade vary drastically depending on the GVC network even conditional on a given input-output dataset. In particular, this section builds on the solutions to the GVC challenge described in section 2 but goes deeper in that it provides specific microfoundations for each theory. The goal is not to develop the most general model, but rather to deliver crisp qualitative results and so I focus on the simplest possible microfoundation. Indeed, I will show how the quantitative implications of any counterfactual experiment may differ radically depending on the assumptions made in order to solve the GVC challenge in (1) and recover the GVC flows underlying input-output tables. Though these models are simple, they provide powerful empirical implications and, while outside the scope of this paper, I conjecture that richer and, perhaps, more credible microfoundations yield similar qualitative implications.

I proceed in four steps. First, I begin by describing an Armington model with specialized inputs. Second, I show that the gains from trade can be represented in terms of a set of domestic expenditure shares - though, in contrast to ACR, the relevant expenditure shares are the expenditures on domestic inputs used for the production of domestically-sold goods instead of the aggregate domestic expenditure share.

¹³The issue of aggregation in input-output tables motivated an important literature in the 1950’s with several papers developing conditions under which aggregation is innocuous. The outlook on whether they might hold in practice was grim, though. In the words of [Hatanaka \(1952\)](#) and [McManus \(1956\)](#), “There is very little chance that they will be fulfilled by any model”.

Since any input-output dataset is consistent with many networks delivering different domestic expenditure shares in domestically-sold goods, this implies that any counterfactual exercise may be consistent with a range of quantitative values. Third, I show that the reason why the domestic expenditure share is not the relevant sufficient statistic in a world of specialized inputs is because it does not capture how changes in trade barriers ripple through supply chain linkages. Fourth, and finally, I show how to construct bounds on any counterfactual exercise based on the class of models consistent with the above gains from trade formulas. I illustrate this empirically with the 2014 WIOD and construct the bounds on the autarky gains from trade and on the welfare losses across Mexico and the U.S. following a NAFTA trade war.

3.1 Armington Meets Specialized Inputs

I extend the standard Armington model with roundabout production (for example as in [Costinot and Rodríguez-Clare 2014](#)) to specialized inputs. There are \mathcal{J} countries and \mathcal{K} industries, with each country-industry $s \in \mathcal{J} \times \mathcal{K}$ producing each \mathcal{J} differentiated varieties — each tailored to a specific market. The standard roundabout model is the special case in which each country-industry produces the same differentiated variety for all markets.

The model is based on five main assumptions: (i) both intermediate inputs and final goods are produced with the same technology, (ii) production is specialized in terms of destination country but not destination industry, (iii) production features constant returns to scale with an upper-tier Cobb-Douglas production function across labor and intermediate inputs from different sectors and a lower-tier constant elasticity of substitution (CES) composite of inputs across source countries, (iv) market structure is perfect competition, (v) the only source of value-added in country j is equipped labor $L(j)$ and commands a wage $w(j)$. While these restrictions are highly restrictive, the model will prove powerful empirically when going to the data.

3.1.1 Production

Formally, assumptions (iii) and (iv) imply that the model can be described directly in terms of unit prices, the dual, with the price of a unit of goods from s' sold to j given by

$$p(s', j) = w(j)^{\beta(s', j)} \prod_{k'' \in \mathcal{K}} \left(\sum_{s'' \in \mathcal{J} \times k''} \alpha(s'' | s', j) (p(s'', j') \tau(s'', j'))^{1-\sigma(k'')} \right)^{\frac{\gamma(k'' | s', j)}{1-\sigma(k'')}} , \quad (13)$$

where notation is such that country-industry pairs are summarized by $s'' = \{j'', k''\}$ and $s' = \{j', k'\}$. The upper-tier Cobb-Douglas is characterized by $\beta(s', j)$, the value-added share, and $\gamma(k'' | s', j)$, the expenditure share on sector k'' inputs, with $\beta(s', s) + \sum_{k'' \in \mathcal{K}} \gamma(k'' | s', s) = 1$. The lower-tier CES composite is characterized by two parameters. First, an elasticity $\sigma(k'') \geq 1$ governing the degree of substitutability of the inputs of k'' purchased across sources $j'' \in \mathcal{J}$. That is, the CES composite of inputs from an arbitrary sector $k'' \in \mathcal{K}$ is a combination of the inputs from different sources indexed by $s'' \in \mathcal{J} \times k''$. Second, a set of exogenous input shifters $\alpha(s'' | s', j)$ governing the relative expenditure on industry k'' inputs from each source $j'' \in \mathcal{J}$, note $s'' \in \mathcal{J} \times k''$, satisfying $\sum_{s'' \in \mathcal{J} \times k''} \alpha(s'' | s', j) = 1$. The price of

varieties s' sold to each country j thus depends on the wage paid in s' , $w(j')$, and the prices that j' itself pays for inputs supplied to it from each source s'' , $p(s'', j')$, times an exogenous trade cost $\tau(s'', j') \geq 1$ governing how many units melt when shipped from s'' to j' .

Production is specialized in that s' puts in specific shares of domestic value-added and inputs from each source s'' into its exports to each market j . That is, of every dollar sold from s' to j a share $\beta(s', j)$ is attributed to the value-added embedded directly by s' while the expenditure on inputs from s'' is determined endogenously and given by

$$\alpha(s'' | s', j) = \frac{\alpha(s'' | s', j) (p(s'', j') \tau(s'', j'))^{1-\sigma(k'')}}{\sum_{t'' \in \mathcal{J} \times \mathcal{K}''} \alpha(t'' | s', j) (p(t'', s') \tau(t'', j'))^{1-\sigma(k'')}} \times \gamma(k'' | s', j). \quad (14)$$

The input expenditure shares are disciplined by the parameters $\alpha(s'' | s', s)$ and I interpret this heterogeneity as a simple way of (exogenously) capturing the interdependencies across different stages and countries of the supply chain. Note, however, that input specificity is *eroded* as goods flow down the supply chain. That is, every country industry s' has access to specific inputs from each source s'' , available at unit cost $p(s'', j') \tau(s'', j')$, but can use them to produce new goods for any downstream market j . This micro-foundation is a bit more restrictive than the specialized inputs framework described in (9) and (10) in that the input shares are common for the production of intermediate inputs and final goods and in that they vary only across destinations. However, I focus on this more restrictive set of networks for the sake of parsimony - the model can be easily generalized to take full advantage of the flexibility in (9) and (10).

3.1.2 Consumers

I model the final demand side as is standard in the literature and assume that consumers aggregate goods across sectors using an upper tier Cobb-Douglas aggregator with $\zeta(k' | j)$ denoting the expenditure share on industry k' final goods by consumers in country j . Further, I assume that within each industry consumers aggregate varieties across source countries into a CES composite with the same elasticity of substitution $\sigma(k') \geq 1$ as above and with the free parameters $\varphi(s' | j)$ disciplining the share of final goods from s' purchased by consumers in each j . With these assumptions the ideal price index in country j equals

$$P(j) = \prod_{k' \in \mathcal{K}} \left(\sum_{s' \in \mathcal{J} \times \mathcal{K}'} \varphi(s' | j) (p(s', j) \tau(s', j))^{1-\sigma(k')} \right)^{\frac{\zeta(k' | j)}{1-\sigma(k')}} , \quad (15)$$

and the expenditure share on final goods from each source country-industry s' equals

$$\pi_F(s' | j) = \frac{\varphi(s' | j) (p(s', j) \tau(s', j))^{1-\sigma(k')}}{\sum_{t' \in \mathcal{J} \times \mathcal{K}'} \varphi(t' | j) (p(t', j) \tau(t', j))^{1-\sigma(k')}} \times \zeta(k' | j). \quad (16)$$

3.1.3 Mapping to Input-Output Data

To map the model to the data I build the model's analogs of the elements of the input-output table. From the final consumption side note that final good purchases by consumers in j from source s' equals the final good share in (16) times aggregate income

$$F(s', j) = \pi_F(s' | j) \times w(j) L(j).$$

The intermediate input side is constructed by noting that a share of the exports to a given market is attributed to the inputs embedded in them. Aggregate intermediate input exports from s'' to s' are obtained by noting that the intermediate input share in (14) delivers the share of inputs from each source in total exports to each country. Thus, aggregate bilateral intermediate input flows must implicitly satisfy

$$X(s'', s') = \sum_{j \in \mathcal{J}} \alpha(s'' | s', j) \left(\sum_{s \in \mathcal{J} \times \mathcal{K}} X(s', s) + F(s', j) \right), \quad (17)$$

where the right-hand side traces the value of inputs from s'' in the exports of both intermediate inputs and final goods from s' to j . The summation then adds up all inputs from s'' used by s' in exports to all markets. In practice, input flows can be computed, conditional on a set of input shares, using this equation as a fixed point.¹⁴

This model has enough degrees of freedom to fit the data perfectly. Conditional on any vector of iceberg trade costs $\tau(s', j) \geq 1$ and any elasticity of substitution $\sigma(k) \geq 1$, the parameters $\varphi(s' | j)$ adjust to match final good flows, the input mix parameters $\alpha(s'' | s', j)$ adjust to match intermediate input flows, and the Cobb-Douglas and value-added shares $\beta(s', j)$, $\gamma(k'' | s', j)$, and $\zeta(k' | j)$ adjust to match GDP. More importantly, there are too many degrees of freedom and so there is a continuum of parameterizations that replicate the same input-output data.

The roundabout model corresponds to the knife-edge case of no specialization in which exports to all markets use the same input mix.¹⁵ With these restrictions, (17) implies the well-known property of roundabout models, and of input-output analysis more generally, that input shares are proportional to bilateral trade shares. That is if value-added and input expenditure shares do not vary across markets then there is a single parameterization of the model that fits the data and delivers the exact same supply chain

¹⁴Alternatively, they can be computed directly with linear algebra through $\mathbf{X} = \alpha [\mathbb{I} - \alpha]^{-1} \mathbf{F}$. This approach is reminiscent of the Leontief inverse matrix but requires a matrix of size $\mathcal{S}^2 \times \mathcal{S}^2$ instead of size $\mathcal{S} \times \mathcal{S}$.

¹⁵More precisely, I use the term roundabout when referring to production processes in which all output uses the same input mix and in which the model is implemented literally in that the sectors in the theory are mapped one-to-one to the sectors in the data (for example [Costinot and Rodríguez-Clare 2014](#), [Caliendo and Parro 2015](#), and [Caliendo et al. 2017](#)). More generally, the above specialized inputs model can also be interpreted as a more disaggregate multi-industry roundabout model in which country j has $\mathcal{K} \times \mathcal{J}$ industries in which the goods produced by industry k for country j are only sold to country j . The mapping to the data is not one-to-one, however, since the theory has $\mathcal{K} \times \mathcal{J}$ industries per country whereas the data has \mathcal{K} . This paper is by no means the first to take issue with the aggregation in input-output data. Rather, I show how to use the same data in new ways by constructing bounds that take the (potential) aggregation concerns into account. In the future, the advent of firm-to-firm data will make the cutting-edge approaches of [Bernard et al. 2018](#), [Lim 2017](#), and [Tintelnot et al. 2017](#) more widely applicable.

network in equilibrium as the one given measured by input-output analysis

$$\begin{aligned} \text{if } \beta(s', j) = \beta(s'), \text{ and } \gamma(k'' | s', j) = \gamma(k'' | s'), \text{ and } \alpha(s'' | s', j) = \alpha(s'' | s'), \forall j \in \mathcal{J}, \\ \Rightarrow \alpha(s'' | s', j) = \alpha(s'' | s') = \frac{X(s'', s')}{\sum_{t'' \in \mathcal{S}} X(t'', s')}. \end{aligned} \quad (18)$$

Hence, while roundabout models may fit the data perfectly, this *cannot* be interpreted as evidence for the roundabout approach since many other specialized inputs models also fit it perfectly. Moreover, input-output data contains no information identifying which networks are most accurate.

3.2 The Gains from Trade

This model delivers a sufficient statistics formula that can be used to calculate the change in welfare following any change in trade barriers. I derive this formula using the exact hat-algebra approach in four steps. Specifically, let a hat variable denote the ratio of a given variable x across two equilibria, i.e. $\hat{x} = x_1/x_0$, and let $\hat{\tau}(s', j)$ denote the (exogenous) change in trade costs of goods shipped from s' to j . As is standard, to make notation cleaner I assume that domestic trade costs do not change, i.e. $\hat{\tau}(s', j) = 1$ for all $s' \in j \times \mathcal{K}$.

First, I derive the change in expenditure shares. From equation (14), the change in input expenditures from source s'' used by s' for goods sold to j as a share of overall expenditure on sector k'' inputs equals

$$\frac{\hat{a}(s'' | s', j)}{\gamma(k'' | s', j)} = \frac{(\hat{p}(s'', j') \hat{\tau}(s'', j'))^{1-\sigma(k'')}}{\sum_{t'' \in \mathcal{J} \times k''} \alpha(t'' | s', j) \times (\hat{p}(t'', s') \hat{\tau}(t'', j'))^{1-\sigma(k'')}}. \quad (19)$$

Analogously, from equation (16), the change in final good expenditures consumed from source s' by consumers in j relative to overall expenditure on sector k' final goods equals

$$\frac{\hat{\pi}_F(s' | j)}{\zeta(k' | j)} = \frac{(\hat{p}(s', j) \hat{\tau}(s', j))^{1-\sigma(k')}}{\sum_{t' \in \mathcal{J} \times k'} \pi_F(t' | j) (\hat{p}(t', j) \hat{\tau}(t', j))^{1-\sigma(k')}}. \quad (20)$$

Both expenditure changes $\hat{a}(s'' | s', j)$ and $\hat{\pi}_F(s' | j)$ depend on exogenous deep parameters given by the Cobb-Douglas industry-level expenditure shares and the elasticities of substitutions, the exogenous change in trade costs, the initial supply chain network as proxied by the original equilibrium's intermediate input and final good expenditure shares, and the endogenous change in unit prices.

Second, I derive the endogenous change in unit prices in terms of domestic expenditure shares. From equation (13), the change in unit prices of goods from s' sold to j can be written in terms of the change in input expenditure shares from some source s'' using the definition in equation (19) as

$$\hat{p}(s', j) = (\hat{w}(j))^{\beta(s', j)} \prod_{k'' \in \mathcal{K}} \left(\hat{a}(s'' | s', j)^{-\frac{1}{1-\sigma(k'')}} \times \hat{p}(s'', j') \hat{\tau}(s'', j') \right)^{\gamma(k'' | s', j)}, \quad (21)$$

where each s'' may in principle be a source located in any country, that is $s'' \in \mathcal{J} \times k''$. As is standard in the sufficient statistics literature, it proves useful to characterize domestic unit prices in terms of domestic

expenditure shares. That is, take (21) and focus on the domestic industries $s' \in j \times \mathcal{K}$ of country j . Domestic unit prices can be written exclusively in terms of the changes in domestic expenditure shares as follows

$$\hat{p}(s', j) = \prod_{s'' \in j \times \mathcal{K}} \left(\hat{w}(j)^{\beta(s'', j)} \prod_{s''' \in j \times \mathcal{K}} \hat{a}(s''' | s'', j)^{-\frac{\gamma(k''' | s'', j)}{1-\sigma(k''')}} \right)^{\delta(k'' | s', j)} \quad (22)$$

where s' , s'' , and s''' all correspond to domestic industries of j . The change in domestic unit prices thus depends on the endogenous change in domestic wages and the change in expenditures on domestic intermediate inputs used for the production of domestically-sold goods. Further, the change in domestic prices depends on the change in these expenditures across all stages of the supply chain as captured by the value-added and Cobb-Douglas expenditure shares and by the auxiliary variable $\delta(k'' | s', j)$ defined as

$$\delta(k'' | s', j) = 1_{[k''=k']} + \gamma(k'' | s', j) + \sum_{s''' \in j \times \mathcal{K}} \gamma(k'' | s''', j) \gamma(k''' | s', j) + \dots$$

In other words, $\delta(k'' | s', j)$ is an expenditure share capturing the aggregate (gross) use of inputs from k'' used in all upstream stages of production of a purely domestic supply chain for inputs that are eventually embedded in goods sold by s' domestically.¹⁶

Third, I derive the endogenous change in the final consumption price index in terms of domestic final good expenditure shares. From equation (15), the change the price index of country j can be written in terms of the change in final good expenditure shares from some source s' using the definition in equation (20) as

$$\hat{P}(j) = \prod_{k' \in \mathcal{K}} \left(\hat{\pi}_F(s' | j)^{-\frac{1}{1-\sigma(k')}} \times \hat{p}(s', j) \hat{\tau}(s', j) \right)^{\zeta(k' | j)} \quad (23)$$

Fourth, and finally, substituting the change in unit prices in (22) into the change in the price index in (23) defined in terms of domestic industries $s' \in j \times \mathcal{K}$ delivers the following formula for the change in welfare in terms of the change in domestic expenditure shares

$$\frac{\hat{w}(j)}{\hat{P}(j)} = \prod_{s' \in j \times \mathcal{K}} \left(\hat{\pi}_F(s' | j)^{\frac{1}{1-\sigma(k')}} \times \prod_{s'' \in j \times \mathcal{K}} \prod_{s''' \in j \times \mathcal{K}} \hat{a}(s''' | s'', j)^{\frac{\gamma(k''' | s'', j) \delta(k'' | s', j)}{1-\sigma(k''')}} \right)^{\zeta(k' | j)} \quad (24)$$

This formula incorporates many of the elements found previously in the literature such as the supply chain elements from [Antràs and de Gortari \(2017\)](#), the domestic expenditure shares from ACR, and the multi-industry input-output linkages of [Caliendo and Parro \(2015\)](#). Specifically, first, in a single-industry world this formula becomes

$$\frac{\hat{w}(j)}{\hat{P}(j)} = \left[\hat{\pi}_F(j | j) \hat{a}(j | j, j)^{\frac{1-\beta(j, j)}{\beta(j, j)}} \right]^{\frac{1}{1-\sigma}} \quad (25)$$

¹⁶Note this expenditure contains value-added counted multiple times. Since the focus is on domestic shares, the Cobb-Douglas expenditure shares $\gamma(k'' | s', j)$ for country j can be written as a $\mathcal{K} \times \mathcal{K}$ square matrix γ with the corresponding $\delta(k'' | s', j)$ shares given by $\delta = [\mathbb{I} - \gamma]^{-1}$.

and the change in welfare depends on the change in the share of final goods purchased domestically and the change in the share of domestic inputs used in the production of domestically-sold inputs. Both shares capture the importance of domestic goods in the production of a purely domestic supply chain with the relative weights given by the power 1 in the case of domestic final goods and the power $(1 - \beta(j, j)) / \beta(j, j)$ in the case of domestic intermediate inputs.¹⁷ This formula is similar to that found previously by [Antràs and de Gortari \(2017\)](#) in a single-industry multi-stage Ricardian model. They find the gains from trade depend on the change in the share of expenditures on goods produced through purely domestic supply chains - an element which is very similar to the specialized inputs expenditure shares $\hat{a}(j|j)$.

Second, in their seminal contribution, ACR carry out their benchmark analysis without intermediate inputs, $\beta(j, j) = 1$, and obtain

$$\frac{\hat{w}(j)}{\hat{P}(j)} = \hat{\pi}_F(j|j)^{\frac{1}{1-\sigma}}.$$

ACR then incorporate intermediate inputs but under the roundabout solution, corresponding to the assumptions in (18), and assuming symmetry in input shares, $\hat{\pi}_F(j|j) = \hat{a}(j|j)$, which delivers

$$\frac{\hat{w}(j)}{\hat{P}(j)} = \hat{\pi}_F(j|j)^{\frac{1}{\beta(j)(1-\sigma)}}.$$

ACR thus developed the insight that the gains from trade, across a variety of models, depend on some form of domestic expenditure shares - which is also true in the world of specialized inputs. Third, and finally, imposing the roundabout assumptions in (18) directly on (24) delivers the formula of [Caliendo and Parro \(2015\)](#) extending the roundabout single-industry results to multiple-industries with input-output linkages.

3.3 The Import Demand System is not CES

Before delving into empirical analysis, it is helpful to pause and analyze why specialized inputs imply that aggregate expenditure shares are insufficient for tracing the implications of changes in trade barriers. In a nutshell, this occurs because supply chains play a role in propagating trade shocks and specialized inputs determine the structure of these trade linkages. In words, if Ford Mexico exports cars to the U.S. built with domestic inputs while Volkswagen Mexico exports to Germany using Chinese imports then changes in Mexican trade costs with different partners will have asymmetric effects on input suppliers depending on the structure of Mexican supply chains.

Formally, this can be stated in terms of the restriction imposed by ACR concerning how trade shocks pass through into relative imports with third countries. In order to make intuition as crisp as possible I restrict attention to a single-industry world - the extension to multiple-industries is analogous and immediate. To begin, it is useful to define the auxiliary variable denoting the dollar value of inputs from source

¹⁷The gross value of domestic input purchases required for producing a dollar of final goods across all stages of a purely domestic supply chain in a single-industry world is

$$(1 - \beta(j, j)) + (1 - \beta(j, j))^2 + (1 - \beta(j, j))^3 + \dots = \frac{1 - \beta(j, j)}{\beta(j, j)}.$$

j'' that country j' uses to produce exports for market j as

$$\mathcal{X}(j'' | j', j) = \alpha(j'' | j', j) (X(j', j) + F(j', j)).$$

From (12), the use of inputs across markets equals aggregate input purchases $X(j'', j') = \sum_{j \in \mathcal{J}} \mathcal{X}(j'' | j', j)$. With these variables in hand, the partial elasticity of imports in j' from source $j'' \neq j'$ relative to domestic input purchases with respect to changes in trade costs with a third country $i'' \neq j'$ equals

$$\frac{\partial \ln X(j'', j') / X(j', j')}{\partial \ln \tau(i'', j')} = (1 - \sigma) 1_{[j''=i'']} + \sum_{j \in \mathcal{J}} \left(\frac{\mathcal{X}(j'' | j', j)}{X(j'', j')} - \frac{\mathcal{X}(j' | j', j)}{X(j', j')} \right) \frac{\partial \ln \mathcal{X}(j' | j', j)}{\partial \ln \tau(i'', j')}. \quad (26)$$

The first term captures the direct effect on relative imports present when $j'' = i''$. In roundabout models this is the only effect. More generally, however, supply chains play a role. The partial elasticity $\partial \ln \mathcal{X}(j' | j', j) / \partial \ln \tau(i'', j')$ captures the change in domestic input purchases due to both a substitution effect capturing a shift in imports from i'' to domestic inputs in exports to j and a supply chain effect derived from the change in downstream production as proxied by the changing level of exports to j . That is

$$\frac{\partial \ln \mathcal{X}(j' | j', j)}{\partial \ln \tau(i'', j')} = -(1 - \sigma) \alpha(i'' | j', j) + \frac{\partial \ln (X(j', j) + F(j', j))}{\partial \ln \tau(i'', j')}.$$

Further, the term in parenthesis in equation (26) amplifies/dampens the effect on relative imports depending on the differential importance of each export market j for inputs from j'' relative to j' .

In words, if Mexican exports to Germany use mostly Chinese inputs, then a reduction in Mexico-Germany shipping costs reduces both imports from China and domestic input sales following the substitution towards more German inputs. However, imports from China fall relatively more since exports to Germany used Chinese inputs intensively. On net, the supply chain effect exerts an opposing force and increases Chinese imports relatively more than domestic sales following the rise in exports to Germany.

Hence, the supply chain effect illustrates how changes in trade barriers with third countries affect imports asymmetrically depending on the depth of supply chain integration. In contrast, the roundabout model is the knife-edge case in which supply chain linkages are symmetric. In other words, when Mexico uses the exact same input mix to produce all exports this model satisfies what ACR define as the import demand system being CES in which case the asymmetric effect of trade costs on relative imports operating through supply chain linkages disappears. Formally

$$\text{The conditions in (18)} \Rightarrow \frac{\partial \ln X(j'', j') / X(j', j')}{\partial \ln \tau(i'', j')} = (1 - \sigma) 1_{[j''=i'']}.$$

The empirical evidence on supply chain linkages suggests that specialized inputs play a crucial role in propagating trade shocks. For example, [Barrot and Sauvagnat \(2016\)](#), [Carvalho et al. \(2016\)](#), and [Boehm et al. \(2018\)](#) show that supply chain disruptions due to natural disasters are propagated by input specificity through trade networks. Increases in suppliers' marginal costs mostly affect tightly-linked firms, rather than entire industries symmetrically as in roundabout models.

Finally, note that the gravity equation's empirical success is not grounds in favor of the roundabout

model. I show in appendix section X that gravity regressions fare well across simulations of the specialized inputs model even though structural gravity does not hold. While third country trade costs do shift bilateral trade flows, on aggregate the bilateral terms dominate. The overall effect is to attenuate the trade elasticity since this model misspecification, i.e. incorrectly assuming structural gravity, is similar to introducing classical measurement error. This suggests that gravity-based trade elasticity estimates are biased downwards when deep supply chain linkages are pervasive.

3.4 Bounding Counterfactuals

3.4.1 Autarky Gains from Trade - Single Industry Bounds

I begin by showcasing the bounds approach to counterfactuals in a highly simplistic setting. For now, I ignore the data's industrial dimension and assume that there is a single industry per country, i.e. $\mathcal{S} = \mathcal{J}$, and compute the bounds on the gains relative to autarky. In this setting, the change in expenditure shares equal the observed equilibrium's expenditure shares, i.e. $\hat{\pi}_F(j' | j') = \pi_F(j' | j') / 1$ and $\hat{a}(j' | j', j') = a(j' | j', j') / (1 - \beta(j', j')) = a(j' | j', j') / \sum_{j'' \in \mathcal{J}} a(j'' | j', j')$, and since the final good shares are observed in the data the only endogenous variables are the input shares $a(j'' | j', j)$.

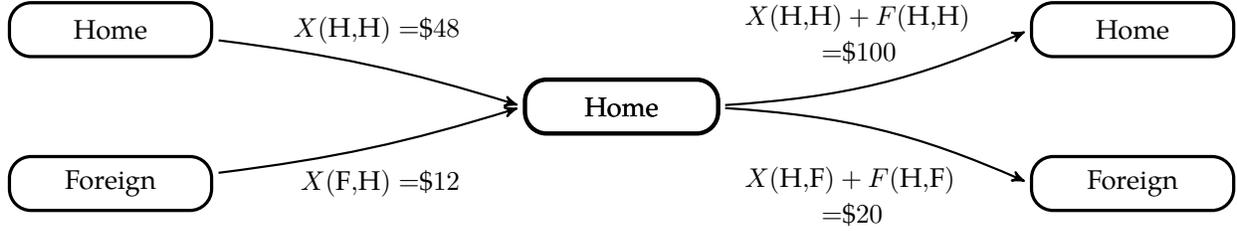
The autarky bounds for country j' in any model delivering a welfare formula as in (25) are given by

$$\begin{aligned} \min/\max \quad & \frac{\sum_{j'' \in \mathcal{J}} a(j'' | j', j')}{1 - \sum_{j'' \in \mathcal{J}} a(j'' | j', j')} \times \ln \frac{a(j' | j', j')}{\sum_{j'' \in \mathcal{J}} a(j'' | j', j')}, \\ \text{subject to} \quad & X(j'', j') = \sum_{j \in \mathcal{J}} a(j'' | j', j) (X(j', j) + F(j', j)), \quad \forall j'' \in \mathcal{J}, \\ & \sum_{j'' \in \mathcal{J}} a(j'' | j', j) \leq 1, \quad \forall j \in \mathcal{J}, \\ & a(j'' | j', j) \geq 0, \quad \forall j'', j \in \mathcal{J}. \end{aligned} \tag{27}$$

The objective function is a concave transformation of equation (25), while the constraints restrict the search to expenditure shares that replicate the input-output data. This optimization problem is relatively easy to solve since the objective function is well-behaved and the constraints are linear. In the special case with constant value-added shares, that is $\beta(j', j) = \beta(j')$ for all $j \in \mathcal{J}$, this reduces to a simpler linear programming problem that bounds $a(j' | j', j')$ directly.

Crucially, computing bounds requires only zooming in on all possible supply chain networks *within* country j' . That is, while the world economy depends on $\mathcal{J} \times \mathcal{J} \times \mathcal{J}$ shares, the optimization solves only for $\mathcal{J} \times \mathcal{J}$ endogenous variables. This occurs because this model features little specialization in that country j' buys specific inputs from j'' , but can then use them to produce exports to any market. Hence, the linkages through specialized inputs of country j' extend at most from its immediate import suppliers to its direct export markets. In other words, the observed bilateral trade flows to and from country j' curtail its specialized network and computing the bounds requires only searching for extremal domestic supply chain linkages.¹⁸ Figure 5 illustrates this in a simple two-country network with constant value-added shares.

¹⁸This breaks down with a higher degree of specialization. For example, suppose that instead $a(j''' | j'', j', j)$ disciplines the



	Roundabout	Lower Bound	Upper Bound
$\alpha(H H, H)$	40%	48%	38%
$\alpha(H H, F)$	40%	0%	50%
$\mathcal{X}(H H, H)$	\$40	\$48	\$38
$\mathcal{X}(H H, F)$	\$8	\$0	\$10
$\hat{w}(H) / \hat{P}(H)$	7.6%	3.7%	8.7%

Figure 5: Domestic Networks in a Simple Home vs Foreign Example: For simplicity, let Home's value-added share be common across destinations and given by $\beta(H) = \$60/\$120 = 50\%$, while the home domestic final good share is $\pi_F(H|H) = \$52/\$60 = 87\%$. The gains from trade are relative to autarky with $1 - \sigma = -5$. Note that home is a relatively closed economy and so the upper bound is mechanically close to the roundabout point estimates. That is, the roundabout shares assign a lot of domestic inputs into all output and so many domestic inputs can be shifted out of exports into domestically sold goods (\$8) but few domestic inputs can be shifted into exports from domestically sold goods (\$2).

Figure 6 plots the gains from trade relative to autarky in the roundabout model (ACR) and in specialized inputs models with both common and destination-specific value-added shares (note the log scale) using the WIOD for 2014. Since the latter class of models nest the former the bounds are wider. Any value within the bounds is feasible since the optimization constraints are linear and any convex combination of the lower and upper bounds is a possible initial trade equilibrium. While the optimization problems do not depend on the trade elasticity, the latter is necessary for transforming the solutions into bounds. Measuring elasticities in specialized inputs models is a fascinating future research topic but beyond this paper's scope and so I simply set a roundabout trade elasticity of $1 - \sigma = -5$, in line with mainstream estimates (Anderson and van Wincoop 2003, Costinot and Rodríguez-Clare 2014, Head and Mayer 2014); the reader can transform any of these numbers χ to her preferred elasticity $1 - \sigma$ through $(1 + \chi)^{(1-6)/(1-\sigma)} - 1$.

The bounds on the gains from trade are wide and increasing in trade openness. For example, the ACR gains for the U.S., a relatively closed economy with only 10% of its total inputs imported from abroad, are low at 2.9% while the range with destination-specific value-added shares lies between 1.2–3.1% indicating that the gains might actually be 60% lower or 10% higher. The range is relatively small, however, with a ratio between the upper and lower bounds of 2.6. In contrast, very open economies are consistent with a wide range of domestic supply chain networks since one can find both trade equilibria in which goods sold domestically use either mostly domestic inputs or almost no domestic inputs. For example, Taiwan

share of inputs from j''' used by country j'' to produce specialized inputs for j' that are further tailored for the downstream market j . This would deliver wider bounds, but comes at a cost in dimensionality (see the previous version of this paper, de Gortari 2017).

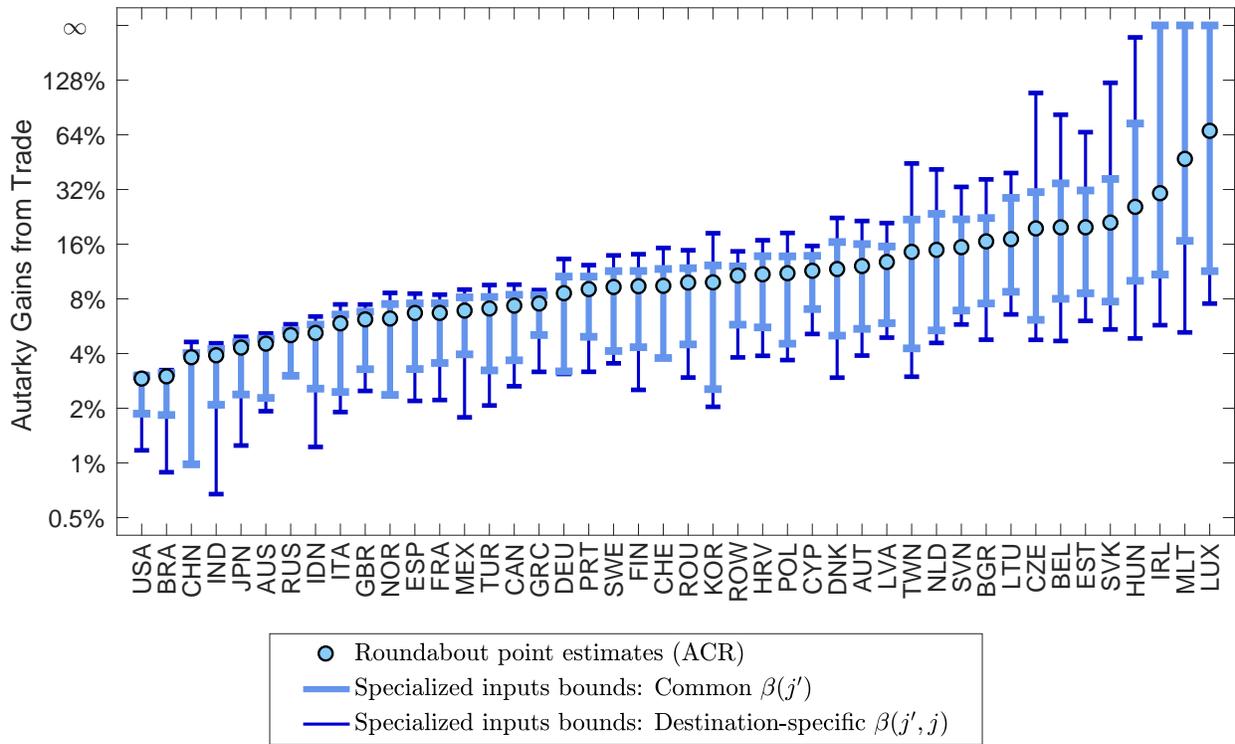


Figure 6: Welfare Gains from Trade Relative to Autarky: All counterfactuals use a roundabout trade elasticity of $1 - \sigma = -5$. Bounds correspond to specialized inputs models as in equation (??). Point estimates correspond to the roundabout model in equation (??). Note the log scale. Data is from the WIOD for 2014 and is aggregated to the country level.

imports about 40% of its inputs from abroad and has a bounds ratio of $45\%/3\% = 15$. The full results are reported in the appendix.

The single-industry bounds feature a mechanical correlation where distance between the ACR gains and the upper bound increases with trade openness. This occurs because trade equilibria in which domestically sold goods use arbitrarily few domestic goods (a high upper bound) can only be found in countries that trade a lot (figure 5 provides further intuition). In practice, extremely open economies like the small European markets on the right of figure 6 feature upper bounds that are quite literally off the charts.

3.4.2 Autarky Gains from Trade - Multiple Industry Bounds

Computing the autarky bounds when incorporating the industrial dimension is analogous to the above but more complex numerically. In particular, incorporating destination-specific value-added and industry-level (Cobb-Douglas) shares is challenging since the gains from trade formula in (25) is highly nonlinear in these terms. Specifically, as discussed in footnote 16, the direct and indirect linkages are captured by the terms $\delta(k''|s', j)$, defined as $\delta = [\mathbb{I} - \gamma]^{-1}$, and are thus endogenous. Optimizing over these terms in high dimensions is very challenging. Further, when these parameters are destination-specific the optimization has to be done globally across all of a country's country-sectors s' since these terms capture cross-sector linkages that cannot be studied in isolation. Hence, to avoid these issues, I focus on computing bounds in the simpler case where only the intermediate input shares within each sector k' vary. That is, I assume

that $\beta(s', j) = \beta(s')$ and $\gamma(k'' | s', j) = \gamma(k'' | s')$ are common across all destinations $j \in \mathcal{J}$ and given by the data. The more general framework is left open for future work.

The autarky bounds for country j' in any multi-industry model delivering a welfare formula as in equation (24) can be found by solving for the extremal supply chain networks within each pair of sectors k'' and k' separately and then incorporating all the solutions into equation (24). Specifically, the extremal domestic shares in country j' for the pair of sectors k'' and k' are found through

$$\begin{aligned} & \min/\max \quad \alpha(s'' | s', j'), \\ & \text{subject to} \quad X(t'', s') = \sum_{j \in \mathcal{J}} \alpha(t'' | s', j) \left(\sum_{s \in \mathcal{J} \times \mathcal{K}} X(s', s) + F(s', j) \right), \forall t'' \in \mathcal{J} \times k'', \\ & \quad \sum_{t'' \in \mathcal{J} \times k''} \alpha(t'' | s', j) = \gamma(k'' | s'), \forall j \in \mathcal{J}, \\ & \quad \alpha(t'' | s', j) \geq 0, \forall t'' \in \mathcal{J} \times k'', j \in \mathcal{J}, \end{aligned} \quad (28)$$

where $s'' = \{j', k''\}$ and $s' = \{j', k'\}$ are domestic country-industries. Optimization problem (28) is a linear program with $\mathcal{J} \times \mathcal{J}$ endogenous variables and easy to solve numerically. The bounds are then given by solving this problem for the $\mathcal{K} \times \mathcal{K}$ industry pairs and inserting the solutions into the terms $\hat{\alpha}(s'' | s', j') = \alpha(s'' | s', j') / \gamma(k'' | s')$ and $\hat{\pi}_F(s' | j') = \pi_F(s' | j') / \zeta(k' | j')$ in equation (24).

Figure 7 confirms that the well known fact that multi-industry models deliver larger gains from trade (Costinot and Rodríguez-Clare 2014) also holds true for the bounds in the 2014 WIOD. Note, however, that this does not hold by construction. Rather, the multi-industry bounds are much larger and overlap little with the single-industry bounds because heterogeneity in openness across sectors leads to disproportionate effects on the gains from trade. To understand this better, imagine a world in which the input-output data across all industry pairs were identical (in shares) to the aggregate single industry data. Following Ossa (2015) we know that even in this very special world the gains computed when using the multi-industry data may be much higher as long as there is heterogeneity in the elasticities of substitution across industries. However, if there were no heterogeneity then the gains should be the same regardless of the level of disaggregation. In Figure 7 I have imposed common elasticities across industries and yet the multi-industry gains are much higher. This occurs because, in reality, even countries that are relatively closed on aggregate often have some very open industries consistent with many supply chain networks.

Hence, in a world of specialized inputs heterogeneity in openness across industries implies that the gains from trade may differ widely depending on the supply chain network even if trade elasticities are common across industries. In future work, mixing both approaches and incorporating heterogeneity in elasticities measured in the context of specialized inputs may produce even wider bounds.

3.4.3 Arbitrary Changes in Trade Costs

Ultimately, perhaps the most relevant empirical exercise is computing bounds on counterfactuals based on real world and policy-motivated events. Relative to the autarky case, however, computing bounds following arbitrary changes to trade barriers is much more complex. While describing how to do it theoretically

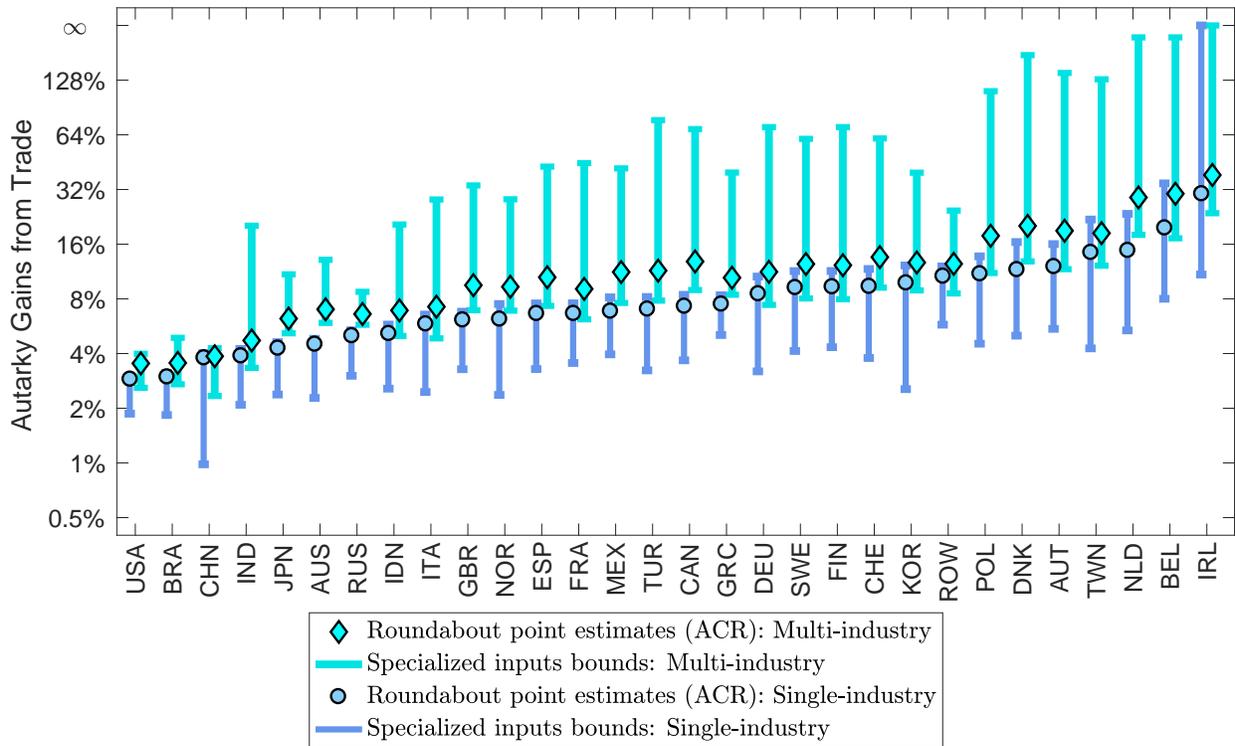


Figure 7: Welfare Gains from Trade Relative to Autarky, Multiple Sectors: All counterfactuals use common roundabout trade elasticities $1 - \sigma(k) = -5$ in all sectors. Point estimates correspond to the roundabout model as in eq X and eq X. Bounds correspond to multi-sector specialized inputs models as in equation (24). Note the log scale. Data is from the WIOD for 2014 and is aggregated the $\mathcal{J} = 30$ largest economies and $\mathcal{K} = 25$ sectors per country (see appendix).

is straightforward, in practice it may be hard to solve numerically with current computing power. Fortunately, though, computing approximate bounds is feasible and are informative about the true bounds.

I relegate the formal description of this optimization problem to Appendix Section X and begin by discussing here the main numerical challenges. First, the counterfactual gains in general depend on the full supply chain network whereas in the autarky experiments the gains depend only on the domestic supply chain network. Hence, while the multi-industry autarky bounds relied on solving $\mathcal{K} \times \mathcal{K}$ separate linear optimization problems of size $\mathcal{J} \times \mathcal{J}$ each, computing bounds for an arbitrary shock to trade barriers requires solving a single global optimization problem with $\mathcal{J} \times \mathcal{K} \times \mathcal{J} \times \mathcal{K} \times \mathcal{J}$ endogenous variables. Second, and relatedly, this implies that the objective function is given by the full formula in equation (24) which is nonlinear. Third, the constraints are highly nonlinear since they include fixed points for the changes in unit prices and wages. These three points imply the increase in dimensionality makes this problem hard to solve numerically unless the data is aggregated to few countries and industries.

All is not lost though since, first, computing general counterfactuals for a given supply chain network is straightforward and, second, the specialized inputs framework described above provides a simple way of constructing supply chain networks conditional on a given input-output dataset. In the special case of the specialized inputs Armington microfoundation, we need only find a set of input shares that satisfy equation (17) for all pairs of country-sectors. Further, since these constraints are linear then if two supply

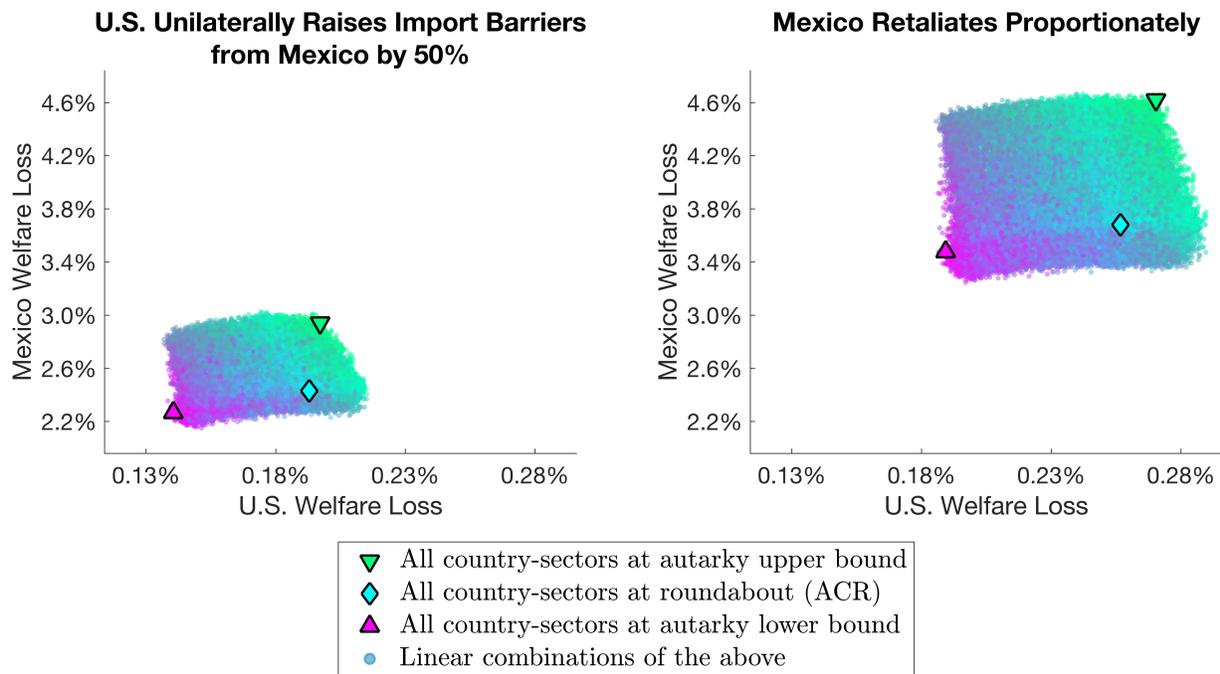


Figure 8: Approximate Bounds of a NAFTA Trade War: The left panel plots the U.S. and Mexico welfare loss following a 50% increase in trade barriers for goods shipped from Mexico to the U.S across 10,000 supply chain networks that replicate the WIOD in 2014. Each circle is a supply chain network constructed as a random linear combination, within each country-sector, of the supply chain networks corresponding to the autarky specialized inputs lower bound, the roundabout solution, and the autarky specialized inputs upper bound underlying Figure 7. The right panel plots the welfare losses when Mexico retaliates and trade barriers also increase by 50% for goods shipped from the U.S. to Mexico. Note the true bounds on these counterfactuals are wider than these sets.

chain networks replicate a given input-output dataset then any linear combination of these two supply chain networks will also do so.

Figure 8 looks at the effects of a NAFTA trade war between Mexico and the U.S. across the set of supply chain networks that can be constructed as linear combinations of the supply chain networks underlying the autarky lower bounds, the roundabout solution, and the autarky upper bounds in figure 7. Specifically, the left panel of figure 8 first takes each of these three benchmark networks, then computes the observed equilibrium (which replicates the 2014 WIOD), and then recomputes the equilibrium following a 50% increase in the trade barriers Mexico faces when exporting goods to the U.S. Then, I take 10,000 linear combinations of these supply chain networks and redo this exercise while plotting each simulation as a dot; the color of each dot proxies how close it is to the three benchmark networks. The right panel does the same exercise except that it assumes that Mexico retaliates and trade barriers for goods shipped from the U.S. to Mexico also increase by 50%.

The sets depicted in figure 8 are narrower than the true bounds, by definition, but show that one can easily find sizable deviations to the predictions of a roundabout model.¹⁹ In this particular example, the

¹⁹The curious reader may wonder why the points in these sets are not convex combinations of the three benchmark networks. There are two reasons. First, the linear combinations are within industry pairs so, for example, there are some supply chain

roundabout model tends to underpredict the costs of NAFTA trade war for Mexico and overpredict for the U.S. relative to the overall set. Of course, this may not be true in reality since the true supply chain network corresponds to only one point which may be anywhere in the set and have little to do with the average network. Further, even if we made the heroic assumption that this model represents the true data generating process in reality, I cannot say whether the true supply chain network is contained within the depicted sets since these only correspond to approximate bounds and the true bounds may be much wider.

As a word of caution, the magnitudes from this exercise are not to be taken at face value. The above model is very stylized and does not incorporate elements such as tariff income, fixed costs of production, nor monopoly power. Rather, the goal is, first, to illustrate in the simplest possible way how quantitative results may differ drastically depending on how researchers interpret input-output datasets. Though outside the scope of this paper, I conjecture that more complex and richer microfoundations deliver qualitatively similar results. Second, the goal is to motivate the need for further measurement so that input-output datasets can be interpreted accurately. By doing this, structural models can then be disciplined by supply chain networks that are informed by data, and are thus hopefully as close as possible to the true underlying networks, instead of relying on the theoretically-built roundabout networks.

4 GVCs and Measures of Globalization

A second strand of the GVC literature is concerned with measuring the extent of fragmentation of production across borders and stages of the supply chain. This literature has developed various measures to better capture this fragmentation than those based on traditional statistics such as gross trade flows. The most popular of these statistics are those based on value-added trade (Hummels et al. 2001, Johnson and Noguera 2012, Koopman et al. 2014) which aim to capture where value is created rather than where value is shipped from. A second set of popular statistics are those based on average upstreamness (Fally 2012, Antràs et al. 2012, Antràs and Chor 2013) which aim to measure a country's average position along the value chain. There are many others such as measures of the factor content of trade (Trefler and Zhu 2010), value-added exchange rates (Bems and Johnson 2017), international inflation spillovers (Auer et al. 2017), and business cycle synchronization (di Giovanni and Levchenko 2010, Johnson 2014b, Duval et al. 2016, di Giovanni et al. 2017). I will refer to all of these measures generically as *measures of globalization*.

This section shows how measures of globalization vary drastically depending on the supply chain network even within a given input-output dataset. In order to sharpen the analysis, I focus throughout on decompositions of value-added trade but the same ideas hold generally for any of the measures of globalization listed above. I proceed in three steps. First, I show that any measure of globalization can be defined using the general theory of GVCs from section 2. This contrasts with the standard approach in which these measures are defined with a specific equilibrium theory of production, mainly the roundabout solution, in mind. Defining these more generally proves useful since this permits the comparison of different theories

networks in which the U.S. is close to the lower bound input shares in some industries but close to the upper bound in other industries. Second, in the autarky case the gains from trade are monotonic in the domestic input shares and so there is a one-on-one relation between linear combinations (at the country level) and the gains from trade. More generally, this is not the case since the counterfactual equilibrium depends nonlinearly (and possibly non-monotonically) on the observed equilibrium through the fixed point in unit prices and wages.

of production in terms of their implications on measures of globalization. Second, I show how to compute bounds on these measures when imposing the specialized inputs solution and compare them to the conventional measures given by the roundabout solution. Third, and finally, I illustrate the latter empirically with the WIOD and showing how the bounds on the share of U.S. value-added in imported Mexican final goods and the U.S.-China value-added deficit, two widely studied measures in the GVC literature, are very wide and thus imply that the conventional roundabout estimates may be highly mismeasured.

A key difference between this section on measures of globalization and the previous section on counterfactuals is that here we still need to take a stand on a theory of production in order to do empirical analysis but we do not need to take a stand on a specific microfoundation. That is, here we still need to assume that the theory of production is given by either the roundabout or specialized inputs solution. However, we need only assume that the theory of production delivers a certain type of supply chain network in equilibrium but can disregard the specific microfoundation that delivers such equilibrium. In this sense, this section is much more flexible than the previous one in that it requires much fewer assumptions.

4.1 Decomposing Value-Added Trade

Decomposing final good consumption into where each dollar of value-added was produced is useful in order to understand how final consumption in some country, say the U.S., is linked to the production capacity of another, say China, through final good exports of a third country, say Mexico. Further, this decomposition can be used to construct value-added trade imbalances by taking the difference between the aggregate value of, say, Chinese value-added consumed in the U.S. arriving through final good exports of any country and the aggregate value of U.S. value-added that is eventually consumed in China. More specifically, in the most general form, the dollar value of value-added from s'' that arrives through final good exports of s' and is consumed in country j is defined as

$$\begin{aligned} \text{VA}(s'' | s', j) = & 1_{[s''=s']} \left[\mathcal{G}^1(s', j) - \sum_{s^2 \in \mathcal{S}} \mathcal{G}^2(s^2, s', j) \right] \\ & + \sum_{N=3}^{\infty} \sum_{s^{N-2} \in \mathcal{S}} \cdots \sum_{s^2 \in \mathcal{S}} \left[\mathcal{G}^{N-1}(s'', s^{N-2}, \dots, s^2, s', j) - \sum_{s^N \in \mathcal{S}} \mathcal{G}^N(s^N, s', s^{N-2}, \dots, s^2, s', j) \right]. \end{aligned} \quad (29)$$

The first term imputes value-added created directly at the assembly stage, appearing only if $s'' = s'$, while the remaining terms impute value-added created by s'' at all further upstream stages of production and which eventually arrives, through any possible sequence, to s' to be shipped to consumers in j .

Value-added trade can be rewritten in terms of a model's equilibrium supply chain network once one takes a stand on the theory of production solving the GVC challenge in (1). To exemplify this, I begin by showing how this decomposition simplifies once one assumes the specialized input solution in (9) and (10), and then show how it relates to the other solutions. To make the exposition clearer, I derive the decomposition separately for the value-added created at each stage N . First, the value-added by s'' at the

most downstream stage, $N = 1$, into final good sales of s' to j equals

$$VA^1 (s'' | s', j) = 1_{[s''=s']} \beta_F (s', j) F (s', j), \quad (30)$$

where the super-index on VA is meant to index the stage at which this value-added is produced. Clearly, since the most downstream stage is that of final production, s'' adds value-added to final good sales of s' if and only if $s'' = s'$, and the decomposition is given by the share of value-added $\beta_F (s', j)$ in each dollar of output times the sales of final goods. Second, the value-added generated at the $N = 2$ upstream stage is given by

$$VA^2 (s'' | s', j) = \beta_X (s'', s') \alpha_F (s'' | s', j) F (s', j), \quad (31)$$

and the decomposition is now given by the value-added share in intermediate inputs $\beta_X (s'', s')$ times the level of inputs from s'' used in the final good sales from s' to j . Third, and finally, the decomposition of value-added for any further upstream stage $N \geq 3$ is given by

$$VA^N (s'' | s', j) = \sum_{s^{N-1} \in \mathcal{S}} \dots \sum_{s^2 \in \mathcal{S}} \beta_X (s'', s^{N-1}) \left[\prod_{n=3}^N \alpha_X (s^n | s^{n-1}, s^{n-2}) \right] \alpha_F (s^2 | s', j) F (s', j), \quad (32)$$

with $s^N = s''$ and $s^1 = s'$. Hence, the total value-added of s'' embedded in final good sales of s' to j is given by the sum of value-added by s'' created at all stages of production

$$VA (s'' | s', j) = \sum_{N=1}^{\infty} VA^N (s'' | s', j).$$

While writing the decomposition in terms of summations across stages of production is useful for illustrating the intuition behind it, in practice it is tedious to implement numerically. I now show how to write this decomposition compactly with linear algebra. First, organize final good flows $F (s, j)$ into a vector \mathbf{F} of size $1 \times \mathcal{S}\mathcal{J}$. Second, organize the intermediate input shares $\alpha_X (s'' | s', s)$ into a matrix $\mathbf{\alpha}_X$ stacked as

$$\mathbf{\alpha}_X = \begin{bmatrix} \alpha_X (1|1, 1) & \alpha_X (1|1, 2) & \dots & \alpha_X (1|1, \mathcal{S}) & \alpha_X (1|2, 1) & \dots & \alpha_X (1|\mathcal{S}, \mathcal{S}) \\ \alpha_X (2|1, 1) & \alpha_X (2|1, 2) & \dots & \alpha_X (2|1, \mathcal{S}) & \alpha_X (2|2, 1) & \dots & \alpha_X (2|\mathcal{S}, \mathcal{S}) \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \alpha_X (\mathcal{S}|1, 1) & \alpha_X (\mathcal{S}|1, 2) & \dots & \alpha_X (\mathcal{S}|1, \mathcal{S}) & \alpha_X (\mathcal{S}|2, 1) & \dots & \alpha_X (\mathcal{S}|\mathcal{S}, \mathcal{S}) \end{bmatrix},$$

of size $\mathcal{S} \times \mathcal{S}^2$, and let $\mathbf{\alpha}_F$ be an analogous matrix of elements $\alpha_F (s'' | s', j)$ but of size $\mathcal{S} \times \mathcal{S}\mathcal{J}$. Third, let β_X and β_F be vectors of elements $\beta_X (s', s)$ and $\beta_F (s', j)$ and of size $1 \times \mathcal{S}^2$ and $1 \times \mathcal{S}\mathcal{J}$. Finally, denote the Kronecker product with \otimes and the Khatri-Rao, or column-wise Kronecker, product with $*$ to define the following auxiliary matrices. Let $\tilde{\mathbf{F}} = \mathbf{F} * (\mathbb{I}_{\mathcal{S}\mathcal{J} \times \mathcal{S}\mathcal{J}})$ be of size $\mathcal{S}\mathcal{J} \times \mathcal{S}\mathcal{J}$, $\tilde{\mathbf{\alpha}}_X = \mathbf{\alpha}_X * (\mathbb{I}_{\mathcal{S} \times \mathcal{S}} \otimes \mathbf{1}_{1 \times \mathcal{S}})$ of size $\mathcal{S}^2 \times \mathcal{S}^2$, $\tilde{\mathbf{\alpha}}_F = \mathbf{\alpha}_F * (\mathbb{I}_{\mathcal{S} \times \mathcal{S}} \otimes \mathbf{1}_{1 \times \mathcal{J}})$ of size $\mathcal{S}^2 \times \mathcal{S}\mathcal{J}$, $\tilde{\beta}_X = \beta_X * (\mathbb{I}_{\mathcal{S} \times \mathcal{S}} \otimes \mathbf{1}_{1 \times \mathcal{S}})$ of size $\mathcal{S} \times \mathcal{S}^2$, and $\tilde{\beta}_F = \beta_F * (\mathbb{I}_{\mathcal{S} \times \mathcal{S}} \otimes \mathbf{1}_{1 \times \mathcal{J}})$ of size $\mathcal{S} \times \mathcal{S}\mathcal{J}$. The matrix with elements $VA (s'' | s', j)$ of size $\mathcal{S} \times \mathcal{S}\mathcal{J}$ is given

by

$$\mathbf{VA} = \tilde{\boldsymbol{\beta}}_F \tilde{\mathbf{F}} + \tilde{\boldsymbol{\beta}}_X [\mathbb{I} - \tilde{\boldsymbol{\alpha}}_X]^{-1} \tilde{\boldsymbol{\alpha}}_F \tilde{\mathbf{F}}. \quad (33)$$

The relation between matrix and full notation is that the first term $\tilde{\boldsymbol{\beta}}_F \tilde{\mathbf{F}}$ summarizes the value-added created at the most downstream stage $\text{VA}^1(s''|s', j)$ and is the matrix representation of (30). Analogously, $\tilde{\boldsymbol{\beta}}_X (\tilde{\boldsymbol{\alpha}}_X)^{N-2} \tilde{\boldsymbol{\alpha}}_F \tilde{\mathbf{F}}$ is the matrix representation of $\text{VA}^N(s''|s', j)$ for $N \geq 2$ so that the second term in (33), given by $\sum_{N=2}^{\infty} \tilde{\boldsymbol{\beta}}_X (\tilde{\boldsymbol{\alpha}}_X)^{N-2} \tilde{\boldsymbol{\alpha}}_F \tilde{\mathbf{F}} = \tilde{\boldsymbol{\beta}}_X [\mathbb{I} - \tilde{\boldsymbol{\alpha}}_X]^{-1} \tilde{\boldsymbol{\alpha}}_F \tilde{\mathbf{F}}$, is the matrix representation of (31) and (32).

I have so far derived the decomposition of value-added by source when imposing the specialized inputs solution, but since the roundabout solution is a special case it is nested within these formulas. Indeed, remember that in the roundabout solution the input mix is independent of the use or destination of output and given by $\alpha_X(s''|s', s) = \alpha_F(s''|s', j) = \alpha(s''|s')$ for all $s \in \mathcal{S}$ and $j \in \mathcal{J}$. Hence, imposing the GVC mapping in (5) implies that the value-added decomposition in (29) becomes

$$\mathbf{VA} = \boldsymbol{\beta} [\mathbb{I} - \boldsymbol{\alpha}]^{-1} \tilde{\mathbf{F}}, \quad (34)$$

where now $\tilde{\mathbf{F}} = \mathbf{F} * (\mathbb{I}_{\mathcal{S} \times \mathcal{S}} \otimes \mathbf{1}_{1 \times \mathcal{J}})$, $\boldsymbol{\beta}$ is a diagonal matrix of size $\mathcal{S} \times \mathcal{S}$ with elements $\beta(s)$, and $\boldsymbol{\alpha}$ is the square matrix of technical coefficients $\alpha(s''|s')$ and of size $\mathcal{S} \times \mathcal{S}$. This decomposition of value-added trade is the standard decomposition used in the GVC literature and mirrors the formulas in [Johnson and Noguera \(2012\)](#) and [Koopman et al. \(2014\)](#).

The key difference between the specialized inputs and roundabout decomposition of value-added trade is that the former depends on an inverse matrix $[\mathbb{I} - \tilde{\boldsymbol{\alpha}}_X]^{-1}$ of size $\mathcal{S}^2 \times \mathcal{S}^2$ while the latter depends on the Leontief inverse matrix $[\mathbb{I} - \boldsymbol{\alpha}]^{-1}$ of size $\mathcal{S} \times \mathcal{S}$. The former is much larger since it summarizes the larger set of information contained in the specialized inputs technical coefficients in which input shares vary depending on the use and destination of output. Finally, decomposing value-added by source in the world of ‘only trade in final goods’ is trivial since, by construction, all value-added is created at the assembly stage. That is, imposing the GVC mapping in (4) implies that the value-added decomposition in (29) becomes

$$\text{VA}(s''|s', j) = 1_{[s''=s']} F(s', j).$$

This discussion illustrates the value of having developed a general theory of GVCs in section 2. While measures of globalization such as the decomposition of value-added trade have been traditionally defined directly in terms of roundabout networks as in (34), here I have defined this measure for any theory of production in (29). Once a researcher takes a stand on a specific theory, then the additional assumptions can be used to simplify the decomposition in (29). Hence, a general theory provides a unifying framework for comparing different theories of production in terms of their implications on measures of globalization.²⁰

As a final point note that imposing a theory of production in this case does not require imposing a specific microfoundation, but rather, only requires assuming how the theory behaves in equilibrium. For example, the decomposition (34) requires assuming the roundabout solution. But it does not require taking

²⁰Further, defining concepts cleanly at this general level should also prove useful for resolving outstanding debates in the literature based on specific theories of production. For example, there is an ongoing discussion about how to define certain value-added measures among [Koopman et al. \(2014\)](#), [Los et al. \(2016\)](#), [Johnson \(2017\)](#), and [Koopman et al. \(2018\)](#).

a stand on whether this equilibrium is achieved through a roundabout Armington model (as described in (18)) or roundabout Ricardian model (as in Eaton and Kortum 2002) or any other microfoundation.

4.2 Bounding Value-Added Trade

Conditional on an input-output dataset and a theory of production, measures of globalization can be bounded. In particular, when imposing the specialized inputs solution in (9) and (10) the bounds on the value-added from some country-sector t'' embedded in the final goods shipped from some country-sector t' to consumers in some country i are given by

$$\begin{aligned}
& \min/\max \quad \sum_{N=1}^{\infty} VA^N(t'' | t', i), \\
& \text{subject to} \quad X(s'', s') = \sum_{s \in \mathcal{S}} a_X(s'' | s', s) X(s', s) + \sum_{j \in \mathcal{J}} a_F(s'' | s', j) F(s', j), \quad \forall s'', s', \\
& \quad \sum_{s'' \in \mathcal{S}} a_X(s'' | s', s) \leq 1, \quad \forall s', s \\
& \quad \sum_{s'' \in \mathcal{S}} a_F(s'' | s', j) \leq 1, \quad \forall s', j, \\
& \quad a_X(s'' | s', s), a_F(s'' | s', j) \geq 0, \quad \forall s'', s', s, j.
\end{aligned} \tag{35}$$

The endogenous variables are the destination-specific input shares for the production of both intermediate inputs and final goods. Similarly to the autarky bounds optimization problems (27) and (28), the constraints are all linear and rely only on ensuring that the constructed supply chain network replicate the observed input-output data. In contrast to the autarky bounds, however, this optimization searches globally over the full supply chain network. That is, whereas the autarky optimization can be done separately within each sector s' , here the objective function depends on how value-added flows across all stages and sequences of production and thus needs to be studied as a whole.

In practice, this problem is hard to solve exactly with current computing power since the objective function is highly nonlinear and the problem is highly dimensional. Specifically, note that $VA^N(t'' | t', i)$ is a polynomial of order N in the endogenous variables and so the objective function is the infinite sum of polynomials of every order. Further, with \mathcal{J} countries and \mathcal{K} industries per country, $\mathcal{J}\mathcal{K} \times \mathcal{J}\mathcal{K} \times (\mathcal{J}\mathcal{K} + \mathcal{J})$ endogenous variables need to be solved for.

A straightforward alternative is to, instead, truncate the objective function and implement an approximate bounds approach. Define the \bar{N} -th order bounds as the solutions to the optimization problem

$$\min / \max \quad \sum_{N=1}^{\bar{N}} VA^N(s'' | s', j), \tag{36}$$

subject to the same constraints in (35). Though a solution to (36) does not deliver the bounds on (35) when $\bar{N} < \infty$, the bounds are likely to be close as long as \bar{N} is large enough. This can be seen by noting that in the limit, when $\bar{N} \rightarrow \infty$, the approximate bounds converge to the true bounds. Further, since value is

created at each stage of production, this implies that value created in very upstream stages of production represents a relatively small share of final good consumption. Hence, the terms corresponding to higher-order polynomials beyond \bar{N} are relatively unimportant when \bar{N} is large enough.

The approximate bounds approach is more easily implemented numerically since the first-order bounds, when $\bar{N} = 1$, are characterized by a linear programming optimization problem while the second-order bounds, when $\bar{N} = 2$, are defined by a quadratic programming optimization problem. Both can be feasibly solved in high dimensions. The problem can be further simplified by focusing on the heterogeneity in input shares while keeping constant value-added shares. That is, when imposing that $\beta_X(s', s) = \beta_F(s', j) = \text{GDP}(s') / \text{GO}(s')$, the second-order bounds are solved by a linear program while the third-order bounds, with $\bar{N} = 3$, are solved by a quadratic program.²¹ In practice, it will become clear below that even the second-order bounds are quite informative about the true bounds.

4.3 Value-Added Trade in the World Input-Output Database

4.3.1 U.S. Value-Added Returned Home Through Imported Mexican Final Goods

One of the most important features, if not the most important one, of trade in the NAFTA region is that supply chains have become deeply integrated. A simple statistic that has received widespread attention recently, especially revolving around the current NAFTA renegotiation discussion, is the amount of U.S. value-added that returns home through final good imports from its NAFTA partners. This statistic matters because, first, it says something about how the different countries are exploiting their comparative advantage by specializing on specific segments of the supply chain instead of on specializing on different goods and, second, because it informs how changes in trade barriers ripple across country borders.²²

But how much U.S. value actually returns home through, say, imported Mexican final goods? Figure 9 shows that current estimates might be off by a wide margin. Specifically, figure 9 provides estimates for the U.S. content in imported Mexican manufacturing final goods in 2014. The roundabout point estimates correspond to the benchmark numbers used in both academia and policy and show that, for example, about 17% of the \$118 billion of imported Mexican manufactures correspond to U.S. value-added created at upstream stages of production. Figure 9 also computes the second-order bounds on these estimates when using the specialized inputs framework in (33) together with the optimization problem in (36). This shows that the true share of U.S. value-added in imported Mexican manufactures may be as low as 6% or as high as 47%. Intuitively, the shares vary drastically because, conditional on the level of imported Mexican goods, the upper bound corresponds to networks in which Mexico uses a lot of U.S. inputs to produce these goods while the lower bound corresponds to networks in which Mexico uses very few U.S. inputs.

Of course, the true bounds are wider and thus the true estimates may not even be contained within these second-order bounds. However, the latter are so wide that I hope this is sufficient to convince the

²¹In this case the linear inequality constraints in (35) are replaced by $\sum_{s'' \in \mathcal{S}} \alpha_X(s'' | s', s) = \sum_{s'' \in \mathcal{S}} \alpha_F(s'' | s', j) = 1 - \beta(s')$.

²²The typical thinking in the policy world is that a higher share of U.S. content in imports from, say, Mexico implies that supply chains are more integrated and so disrupting these would be more costly. The exact quantitative effects, of course, will depend on elasticities of substitution and the costs of relocating supply chains across countries. However, Blanchard et al. (2016) have shown that this basic intuition holds formally. Specifically, they show that countries, such as the U.S., should set lower tariffs on imports that contain a high share of their own domestic content.

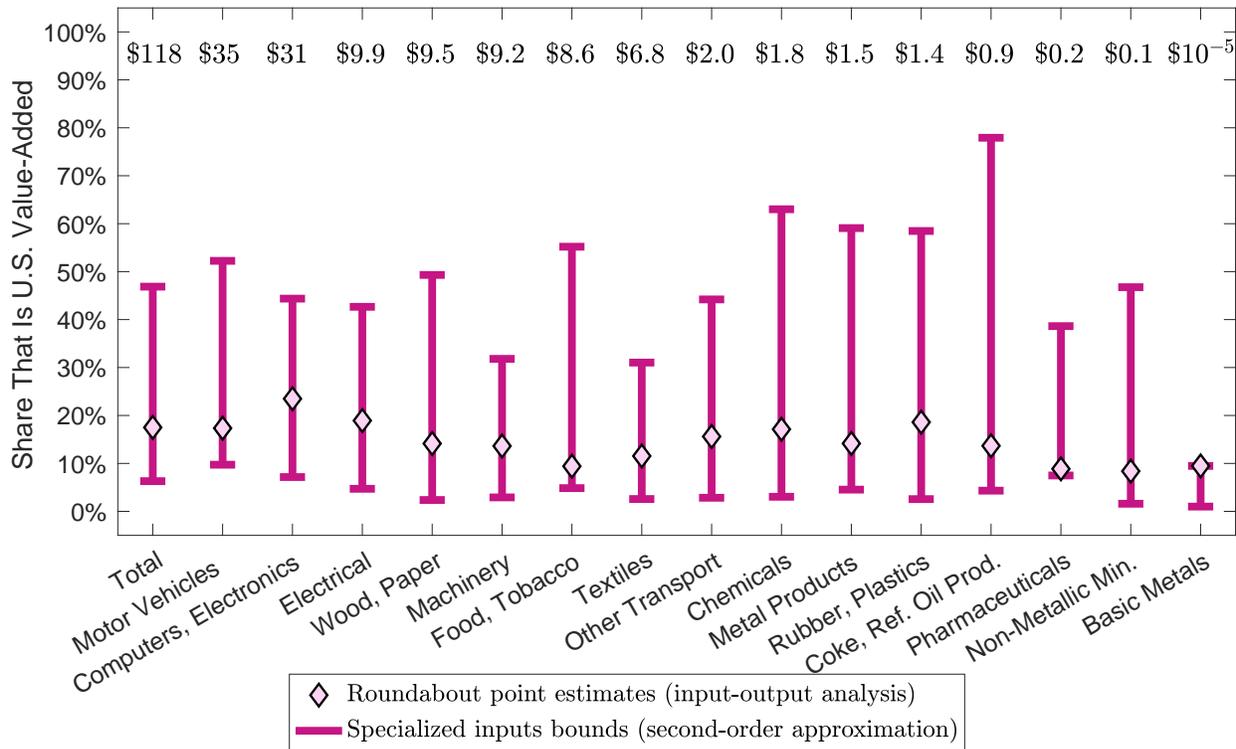


Figure 9: Share of U.S. Value-Added in Imported Mexican Final Goods: Point estimates are obtained through the roundabout decomposition in (34). The bounds correspond to the second-order bounds on the specialized inputs decomposition in equation (33) obtained by solving the optimization problem in (36) with $\bar{N} = 2$ and common value-added shares. The numbers at the top are the dollar value in billions of imported Mexican final goods in each manufacturing category. Data is from the WIOD for 2014.

reader that, in practice, the share of U.S. value in imported Mexican goods may be highly mismeasured.

4.3.2 U.S.-China Trade Imbalances

Another value-added trade statistic that has received widespread attention both in the academic and policy worlds is the value-added trade balance between the U.S. and China. In their seminal contribution, [Johnson and Noguera \(2012\)](#) showed the trade deficit looks less extreme if it is computed as the difference between the U.S. value consumed in China and the Chinese value consumed in the U.S. instead of the difference in gross exports between the two countries. The reasoning is that the former provides a better measure of how both economies are relatively linked, and one that does so in terms of the final consumer's perspective, since gross exports from, say China, in principle may say little about China's importance to those exports (i.e. how much value-added China contributes).

But is it really true that the U.S.-China trade deficit is smaller when computed in value-added terms or is this just an artifice of how input-output datasets are conventionally interpreted? Figure 10 plots the U.S.-China trade balance both in gross and value-added terms between 2000-2014. The difference between the gross trade balance (circles) and the value-added trade balance based on the roundabout solution (di-

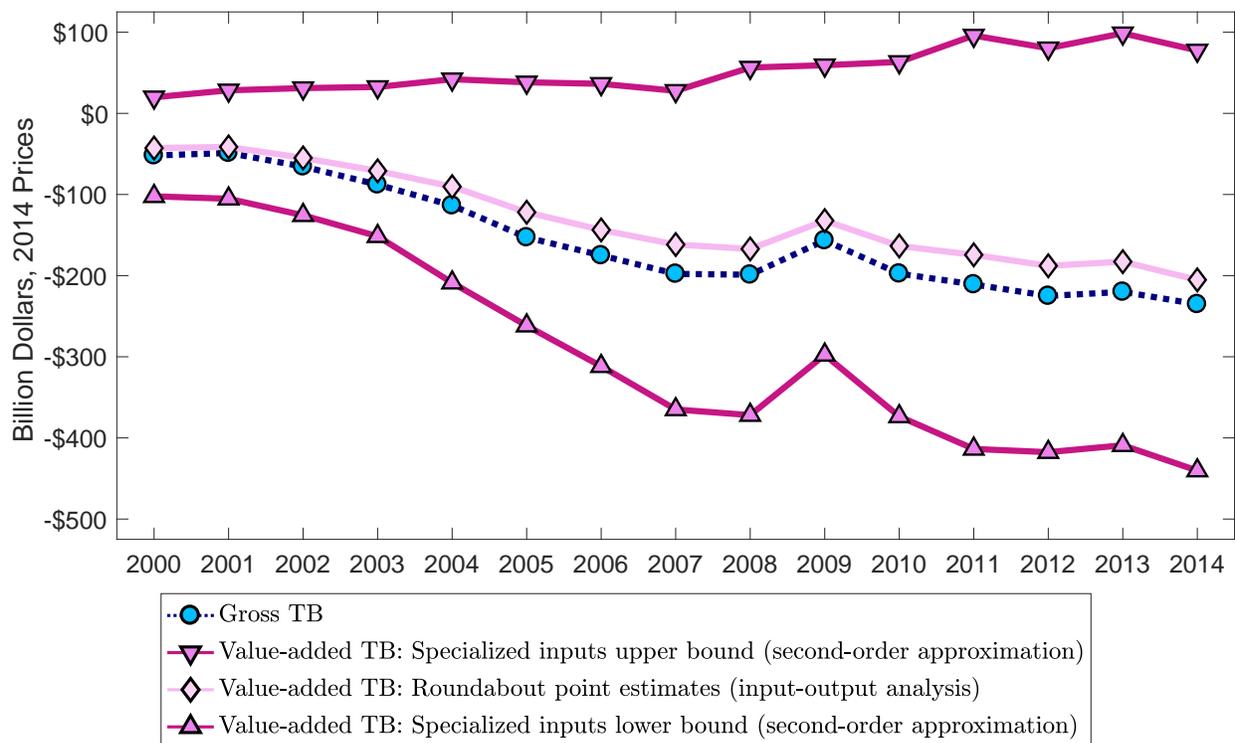


Figure 10: U.S.-China Trade Imbalances: The series with circles corresponds to the gross trade balance. The other three series correspond to the value-added trade balance. The middle-series (diamonds) corresponds to the roundabout point estimates, which can be computed with input-output analysis, and are based on the decomposition in (34). The upper and lower series (triangles) are second-order approximations to the bounds on the specialized inputs decomposition in (33) and computed through the optimization problem (36) with common value-added shares. Data is from the WIOD.

amonds) replicate the key finding in [Johnson and Noguera \(2012\)](#) that the gross deficit in 2004 overstates the value-added deficit by about 25%. Furthermore, the evolution of both series across time looks exactly like the third figure in [Johnson \(2014a\)](#). However, specialized inputs tell a potentially very different story. The second-order bounds show these findings may actually be more pronounced in that the value-added balance at the upper bound is a surplus. Alternatively, the lower bound shows these findings may actually be reversed in that the value-added deficit at the lower bound is larger than the gross deficit.²³ Intuitively, a value-added surplus means that in reality China is exporting back to the U.S. a lot more U.S. value than is currently accounted for. Meanwhile, a larger value-added deficit means that in reality China is exporting back to the U.S. so little U.S. value that this imbalance is much larger than when studied in gross terms.

²³Since the value-added trade balance is based on the decomposition of value-added trade $VA(s''|s',j)$, the bounds are found by replacing the objective function in (35) with

$$\sum_{N=1}^{\infty} \sum_{t'' \in \text{USA} \times \mathcal{K}} \sum_{t' \in \mathcal{J} \times \mathcal{K}} VA^N(t''|t', \text{CHINA}) - \sum_{N=1}^{\infty} \sum_{t'' \in \text{CHINA} \times \mathcal{K}} \sum_{t' \in \mathcal{J} \times \mathcal{K}} VA^N(t''|t', \text{USA}),$$

and the approximate bounds are computed analogously to (36).

Again, as before, figure 10 does not necessarily imply that the widely-held view that the value-added deficit is smaller than the gross deficit is wrong. But it does illustrate the point that even facts which are now part of conventional wisdom may be pure artifice of how supply chain networks have been measured so far. The key point is that determining facts regarding whether the value-added deficit is or is not smaller than the gross deficit requires more knowledge about the supply chain networks underlying input-output datasets in order to make either statement convincingly. Indeed, the bounds in figure 10 deliver a very stark message: The difference between the conventional measures of value-added and gross trade balances are dwarfed by the differences in the potential mismeasurement in the former statistic.

5 GVCs and Measurement: Bringing in New Sources of Information

This last section is devoted to a third strand of the GVC literature concerned with measuring GVC flows. The motivation is twofold. First, section 2 provided evidence from firm-level data and domestic input-output tables that is not consistent with the roundabout interpretation of multi-country input-output data (at the current level of industrial aggregation). This evidence thus suggests that the specialized inputs solution is more appropriate for studying GVCs and that this type of information can be used to measure GVCs more accurately. Second, sections 3 and 4 showed that both the bounds on counterfactuals and measures of globalization, conditional on a given input-output dataset, are wide. Hence, better measurement is needed in order to obtain more accurate quantitative answers to both questions.

5.1 Disciplining GVCs with New Sources of Information

As discussed previously, specific GVC networks can be found as the solution to optimization problems that minimize some objective function while searching over the set of GVC networks consistent with a given input-output database. For example, the optimization problems (27) and (28) minimized the welfare gains from trade while (35) minimized a measure of value-added trade. However, these optimization problems did not minimize a feature of the network per se but rather a statistic of interest based on this statistic.

I now show how to use additional information, beyond that contained in input-output datasets, to construct a GVC network while also exploiting the information contained in the input-output data. This approach is useful when a researcher has some useful information about the GVCs underlying input-output data but that is insufficient for fully characterizing the flows directly. For example, a researcher with access to the universe of firm-to-firm transaction data for a whole country, say Belgium, could build the GVC flows linking Belgian industries to other industries in Belgium and the world directly and there would be no need to complement this data with the input-output data (i.e. the input-output data is redundant in this scenario). However, in practice, there are many firm-level datasets that are informative about GVC flows but are too limited for measuring the latter directly. For example, most countries construct customs-level datasets but these lack the universe of domestic transactions and so do not include the whole GVC picture.

In this type of scenarios, GVCs can be measured with the following optimization problem

$$\begin{aligned}
\min \quad & h \left(\{ \alpha_X (s'' | s', s) \}_{s'' \in \mathcal{S}, s \in \mathcal{S}}, \{ \beta_X (s', s) \}_{s \in \mathcal{S}}, \{ \alpha_F (s'' | s', j) \}_{s'' \in \mathcal{S}, j \in \mathcal{J}}, \{ \beta_F (s', j) \}_{j \in \mathcal{J}} \right), \\
\text{subject to} \quad & X (s'', s') = \sum_{s \in \mathcal{S}} \alpha_X (s'' | s', s) X (s', s) + \sum_{j \in \mathcal{J}} \alpha_F (s'' | s', j) F (s', j), \quad \forall s'', \\
& \sum_{s'' \in \mathcal{S}} \alpha_X (s'' | s', s) + \beta_X (s', s) = 1, \quad \forall s, \\
& \sum_{s'' \in \mathcal{S}} \alpha_F (s'' | s', j) + \beta_F (s', j) = 1, \quad \forall j, \\
& \alpha_X (s'' | s', s), \alpha_F (s'' | s', j), \beta_X (s', s), \beta_F (s', j) \geq 0, \quad \forall s'', s, j.
\end{aligned} \tag{37}$$

The objective function $h(\cdot)$ depends on the endogenous variables, the input and value-added expenditure shares from the specialized inputs framework, and (potentially) some exogenous parameters chosen by the researcher. For example, a simple and tractable objective function is given by targeting some exogenous value for each share and minimizing the weighted sum of the squared deviations

$$\begin{aligned}
h(\cdot) = & \sum_{s'' \in \mathcal{S}} \sum_{s \in \mathcal{S}} \omega_X^0 (s'' | s', s) [\alpha_X (s'' | s', s) - \alpha_X^0 (s'' | s', s)]^2 \\
& + \sum_{s'' \in \mathcal{S}} \sum_{j \in \mathcal{J}} \omega_F^0 (s'' | s', j) [\alpha_F (s'' | s', j) - \alpha_F^0 (s'' | s', j)]^2 \\
& + \sum_{s \in \mathcal{S}} \omega_X^0 (s', s) [\beta_X (s', s) - \beta_X^0 (s', s)]^2 + \sum_{j \in \mathcal{J}} \omega_F^0 (s', j) [\beta_F (s', j) - \beta_F^0 (s', j)]^2.
\end{aligned} \tag{38}$$

In this case, $\alpha_X^0 (s'' | s', s)$, $\alpha_F^0 (s'' | s', j)$, $\beta_X^0 (s', s)$, and $\beta_F^0 (s', j)$ are targets chosen by the researcher, disciplined by the information in customs data for example, while $\omega_X^0 (s'' | s', s)$, $\omega_F^0 (s'' | s', j)$, $\omega_X^0 (s', s)$, and $\omega_F^0 (s', j)$ correspond to the weights on each target and are also chosen by the researcher. Solving the optimization with this objective function thus delivers the GVC network closest to the researcher's targets, in the sense of minimizing the weighted sum of squared deviations, among all of the GVC networks consistent with an input-output dataset as characterized by the constraints in (37).

Careful inspection of (37) reveals that the optimization problem is defined for a specific country-sector s' . That is, this problem delivers values for the inputs shares from all sources s'' in the production of intermediates sold to all country-sectors s and countries j . This is a larger problem than the multi-industry counterfactual bounds in (28) since the latter has more structure because the cross-industry Cobb-Douglas production function requires only searching for the input shares of the domestic k' industry from suppliers of industry k'' . However, it is a smaller problem than the multi-industry value-added bounds in (35) which requires searching over the full supply chain network and thus for the input shares across all country-industry bounds s' simultaneously. While the measurement problem (37) is also large and nonlinear, I define it more restrictively in order to get exact solutions.

Finally, in the following empirical application I use the quadratic objective function since this is the easiest nonlinear function that can be solved in high dimensions given that it has linear first-order conditions. Hence, measuring a full GVC network using (37) with (38) requires taking a stand on the weights

and targets and solving \mathcal{S} optimization problems, one for each $s' \in \mathcal{S}$, of size $\mathcal{S} \times (\mathcal{S} + \mathcal{J})$ each. Needless to say, measurement can be done with many other objective function functional forms in future work.²⁴

5.2 Improving Measurement with Mexican Customs Shipment-Level Data

I now implement this approach while disciplining the targets using Mexican customs data. In particular, remember that this data delivers the distribution of foreign inputs use to produce each of Mexico's manufacturing exports but contains no information on domestic purchases. Hence, it cannot be mapped to overall input expenditure shares without further assumptions. I take the stand that Mexico only does processing trade - i.e. that imported inputs are only used to produce exports. While this is a very strong assumption, it is not too far-fetched for Mexico since it is, together with China, one of two large economies in which processing trade is widely prevalent (De La Cruz et al. 2011).

Table 1 presents the share of U.S. value-added in Mexican manufacturing imports according to the roundabout solution (column I), the lower and upper bound on specialized inputs (columns II and III), and the point estimates obtained through the use of Mexican customs data (columns IV and V). The difference between the latter two columns is that column IV assumes common value-added shares within each manufacturing industry and across all output, while column V additionally targets that Mexican domestic sales have twice as much directly added Mexican value-added than exports. The latter is in line with De La Cruz et al. 2011, who show that processing exports also tend to be less intensive in domestic value-added.

The main takeaway is that Mexican-American supply chains are much more integrated than as suggested by conventional estimates. For example, in motor vehicles - the largest imported manufacturing industry - the U.S. tends to import back around 38% of its own domestic value whereas the roundabout estimates predict a much smaller share of only 17%. For overall manufacturing, the U.S. share is about 30% and also substantially higher than the roundabout estimate of 17%. These differences are in line with the input shares observed in figure 2: While the customs data show that Mexico uses a high share of American inputs to produce exports to the U.S., these estimates trace value across all stages of production and confirm that a large part of these exports is American value-added. In contrast, the roundabout approach waters down the share of U.S. content by assuming common input shares in all output.

6 Conclusion

TBD

²⁴The problem cannot be made smaller without further structure because the input shares across suppliers and destinations of s' are interlinked through the linear constraints. However, the problem can easily be made more general, for example by searching for the full supply chain network in a single optimization problem, by choosing an objective function $h(\cdot)$ featuring complementarities across input shares from different country-sectors $s' \in \mathcal{S}$.

	Roundabout	Specialized Inputs Common V.A.		Specialized Inputs Lower Export V.A.	
		Approx. Bounds Lower	Upper	Mex. Customs + Proc. Trade	Mex. Customs + Proc. Trade
U.S. Value-Added in Imports from Mexico (%)	I	II	III	IV	V
Total Manufactures	17	6	47	27	30
Basic Metals	9	1	9	5	5
Chemicals	17	3	63	33	37
Coke, Refined Oil Prod.	14	4	78	43	25
Computers, Electronics	23	7	44	13	16
Electrical	19	5	43	25	28
Food, Tobacco	9	5	55	37	39
Machinery	14	3	32	20	24
Metal Products	14	5	59	27	32
Motor Vehicles	17	10	52	37	39
Non-Metallic Minerals	8	2	47	24	17
Other Transport	16	3	44	24	30
Pharmaceuticals	9	7	39	15	24
Rubber, Plastics	19	3	58	34	41
Textiles	12	3	31	24	26
Wood, Paper	14	2	49	33	38

Table 1: U.S. Value-Added in Manufacturing Imports from Mexico: Column I is computed with the roundabout formula (34). Columns II and III correspond to the specialized inputs bounds depicted in figure 7. Columns IV and V correspond to the specialized inputs estimates obtained through (37) when disciplining the targets with Mexican customs data and a processing trade assumption. Column IV further assumes common value-added shares in each Mexican manufacturing industry whereas column V assume twice as much direct Mexican value-added in domestic sales relative to exports. Data is from the WIOD for 2014.

References

- ALLEN, T., C. ARKOLAKIS, AND Y. TAKAHASHI (2017): “Universal Gravity,” *NBER Working Paper 20787*.
- ANDERSON, J. E. (1979): “A Theoretical Foundation for the Gravity Equation,” *The American Economic Review*, 69, 106–116.
- ANDERSON, J. E. AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A Solution to the Border Puzzie,” *The American Economic Review*, 93, 170–192.
- ANTRÀS, P. AND D. CHOR (2013): “Organizing the Global Value Chain,” *Econometrica*, 81, 2127–2204.
- ANTRÀS, P., D. CHOR, T. FALLY, AND R. HILLBERRY (2012): “Measuring the Upstreamness of Production and Trade Flows,” *The American Economic Review*, 102, 412–416.
- ANTRÀS, P. AND A. DE GORTARI (2017): “On the Geography of Global Value Chains,” *NBER Working Paper 23456*.
- ANTRÀS, P., T. C. FORT, AND F. TINTELNOT (2017): “The Margins of Global Sourcing: Theory and Evidence from U.S. Firms,” *American Economic Review*, 107, 2514–2564.
- ANTRÀS, P. AND R. W. STAIGER (2012): “Offshoring and the Role of Trade Agreements,” *American Economic Review*, 102, 3140–3183.
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): “New Trade Models, Same Old Gains?” *American Economic Review*, 102, 94–130.
- AUER, R. A., A. A. LEVCHENKO, AND P. SAUR (2017): “International Inflation Spillovers Through Input Linkages,” *NBER Working Paper 23246*.
- BALISTRERI, E. J., R. H. HILLBERRY, AND T. F. RUTHERFORD (2011): “Structural estimation and solution of international trade models with heterogeneous firms,” *Journal of International Economics*, 83, 95–108.
- BARROT, J.-N. AND J. SAUVAGNAT (2016): “Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks,” *The Quarterly Journal of Economics*, 1543–1592.
- BASTOS, P. AND J. SILVA (2010): “The quality of a firm’s exports: Where you export to matters,” *Journal of International Economics*, 82, 99–111.
- BASTOS, P., J. SILVA, AND E. VERHOOGEN (2018): “Export Destinations and Input Prices,” *American Economic Review*, 108, 353–392.
- BEMS, R. (2014): “Intermediate inputs, external rebalancing and relative price adjustment,” *Journal of International Economics*, 94, 248–262.
- BEMS, R. AND R. C. JOHNSON (2017): “Demand for Value Added and Value-Added Exchange Rates,” *American Economic Journal: Macroeconomics*.
- BERNARD, A. B., A. MOXNES, AND Y. U. SAITO (2018): “Production Networks, Geography and Firm Performance,” *Journal of Political Economy*.
- BLANCHARD, E. J., C. P. BOWN, AND R. C. JOHNSON (2016): “Global Supply Chains and Trade Policy,” .
- BLAUM, J., C. LELARGE, AND M. PETERS (2017a): “Firm Size and the Intensive Margin of Import Demand,” .
- (2017b): “The Gains From Input Trade With Heterogeneous Importers,” *American Economic Journal: Macroeconomics*, Forthcomin.
- BOEHM, C. E., A. FLAAEN, AND N. PANDALAI-NAYAR (2018): “Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tohoku Earthquake,” *The Review of Economics and Statistics*, Forth-

comin.

- BRAMBILLA, I., D. LEDERMAN, AND G. PORTO (2012): “Exports, Export Destinations, and Skills,” *The American Economic Review*, 102, 3406–3438.
- BRAMBILLA, I. AND G. G. PORTO (2016): “High-income export destinations, quality and wages,” *Journal of International Economics*, 98, 21–35.
- CALIENDO, L. AND F. PARRO (2015): “Estimates of the Trade and Welfare Effects of NAFTA,” *Review of Economic Studies*, 82, 1–44.
- CALIENDO, L., F. PARRO, AND A. TSYVINSKI (2017): “Distortions and the Structure of the World Economy,” *NBER Working Paper 23332*.
- CARVALHO, V. M., M. NIREI, Y. U. SAITO, AND A. TAHBAZ-SALEHI (2016): “Supply Chain Disruptions: Evidence from the Great East Japan Earthquake,” .
- CONCONI, P., M. GARCÍA-SANTANA, L. PUCCIO, AND R. VENTURINI (2018): “From Final Goods to Inputs : the Protectionist Effect of Rules of Origin,” *American Economic Review*, Forthcomin.
- COSTINOT, A. AND A. RODRÍGUEZ-CLARE (2014): “Trade Theory with Numbers: Quantifying the Consequences of Globalization,” in *Handbook of International Economics*, Elsevier B.V., vol. 4, 197–261.
- COSTINOT, A., J. VOGEL, AND S. WANG (2012): “An Elementary Theory of Global Supply Chains,” *The Review of Economic Studies*, 80, 109–144.
- DE GORTARI, A. (2017): “Disentangling Global Value Chains,” .
- DE LA CRUZ, J., R. B. KOOPMAN, Z. WANG, AND S.-J. WEI (2011): “Estimating Foreign Value-Added in Mexico’s Manufacturing Exports,” *U.S. International Trade Commission Working Paper*.
- DEAN, J. M., K. C. FUNG, AND Z. WANG (2011): “Measuring Vertical Specialization: The Case of China,” *Review of International Economics*, 19, 609–625.
- DI GIOVANNI, J. AND A. A. LEVCHENKO (2010): “Putting the Parts Together: Trade, Vertical Linkages, and Business Cycle Comovement,” *American Economic Journal: Macroeconomics*, 2, 95–124.
- (2013): “Firm entry, trade, and welfare in Zipf’s world,” *Journal of International Economics*, 89, 283–296.
- DI GIOVANNI, J., A. A. LEVCHENKO, AND I. MEJEAN (2017): “The Micro Origins of International Business Cycle Comovement,” *NBER Working Paper 21885*.
- DORNBUSCH, R., S. FISCHER, AND P. A. SAMUELSON (1977): “Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods,” *The American Economic Review*, 67, 823–839.
- DUVAL, R., N. LI, R. SARAF, AND D. SENEVIRATNE (2016): “Value-added trade and business cycle synchronization,” *Journal of International Economics*, 99, 251–262.
- EATON, J. AND S. KORTUM (2002): “Technology, Geography, and Trade,” *Econometrica*, 70, 1741–1779.
- FALLY, T. (2012): “Production Staging: Measurement and Facts,” .
- FALLY, T. AND R. HILLBERRY (2016): “A Coasian Model of International Production Chains,” *NBER Working Paper 21520*, 1–53.
- FIELER, A. C., M. ESLAVA, AND D. Y. XU (2017): “Trade, Quality Upgrading, and Input Linkages: Theory and Evidence from Colombia,” *American Economic Review*, Forthcomin.
- GOPINATH, G. AND B. NEIMAN (2014): “Trade Adjustment and Productivity in Large Crises,” *American Economic Review*, 104, 793–831.

- HANSON, G. H., R. J. MATALONI, AND M. J. SLAUGHTER (2005): "Vertical Production Networks in Multinational Firms," *The Review of Economics and Statistics*, 87, 664–678.
- HATANAKA, M. (1952): "Note on Consolidation Within a Leontief System," *Econometrica*, 20, 301–303.
- HEAD, K. AND T. MAYER (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook," in *Handbook of International Economics*, Elsevier B.V., vol. 4, 131–195.
- HUMMELS, D., J. ISHII, AND Y. KEI-MU (2001): "The nature and growth of vertical specialisation in world trade," *Journal of International Economics*, 54, 75–96.
- JOHNSON, R. C. (2014a): "Five Facts about Value-Added Exports and Implications for Macroeconomics and Trade Research," *Journal of Economic Perspectives*, 28, 119–142.
- (2014b): "Trade in Intermediates and Business Cycle Comovement," *American Economic Journal: Macroeconomics*, 6, 39–83.
- (2017): "Measuring Global Value Chains," *NBER Working Paper 24027*.
- JOHNSON, R. C. AND A. MOXNES (2016): "Technology, Trade Costs, and the Pattern of Trade with Multistage Production," .
- JOHNSON, R. C. AND G. NOGUERA (2012): "Accounting for Intermediates: Production Sharing and Trade in Value Added," *Journal of International Economics*, 86, 224–236.
- KIKKAWA, A. K., G. MAGERMAN, AND E. DHYNE (2017): "Imperfect Competition and the Transmission of Shocks: The Network Matters," .
- KOOPMAN, R., Z. WANG, AND S.-J. WEI (2012): "Estimating domestic content in exports when processing trade is pervasive," *Journal of Development Economics*, 99, 178–189.
- (2014): "Tracing Value-Added and Double Counting in Gross Exports," *American Economic Review*, 104, 459–494.
- KOOPMAN, R., Z. WANG, S.-J. WEI, AND K. ZHU (2018): "A Comparison of Two Approaches to Gross Trade Decomposition," .
- KRUGMAN, P. AND A. J. VENABLES (1995): "Globalization and the Inequality of Nations," *The Quarterly Journal of Economics*, 110, 857–880.
- LEONTIEF, W. (1941): *The Structure of American Economy 1919-1929: An Empirical Application of Equilibrium Analysis*, Oxford University Press.
- LIM, K. (2017): "Firm-to-Firm Trade in Sticky Production Networks," .
- LINDEN, G., K. KRAEMER, AND J. DEDRICK (2011): "Innovation and Job Creation in a Global Economy: The Case of Apple's iPod," *Journal of International Commerce and Economics*, 3, 223–239.
- LOS, B., M. P. TIMMER, AND G. J. DE VRIES (2016): "Tracing Value-Added and Double Counting in Gross Exports: Comment," *The American Economic Review*, 106, 1958–1966.
- MANOVA, K. AND Z. YU (2016): "How firms export: Processing vs. ordinary trade with financial frictions," *Journal of International Economics*, 100, 120–137.
- MANOVA, K. AND Z. ZHANG (2012): "Export Prices and Destinations," *The Quarterly Journal of Economics*, 127, 379–436.
- McMANUS, M. (1956): "On Hatanaka's Note on Consolidation," *Econometrica*, 24, 482–487.
- NUNN, N. (2007): "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade," *The Quarterly Journal of Economics*, 122, 569–600.

- OBERFIELD, E. (2018): "A Theory of Input-Output Architecture," *Econometrica*, 86, 559–589.
- OSSA, R. (2015): "Why trade matters after all," *Journal of International Economics*, 97, 266–277.
- POPPER, K. R. (1959): *The Logic of Scientific Discovery*, London: Hutchinson.
- RAUCH, J. E. (1999): "Networks versus markets in international trade," *Journal of International Economics*, 48, 7–35.
- SAMUELSON, P. A. (1952): "Spatial Price Equilibrium and Linear Programming," *The American Economic Review*, 42, 283–303.
- SCHOTT, P. K. (2004): "Across-Product versus Within-Product Specialization in International Trade," *The Quarterly Journal of Economics*, 119, 647–678.
- TINTELNOT, F., A. K. KIKKAWA, M. MOGSTAD, AND E. DHYNE (2017): "Trade and Domestic Production Networks," .
- TREFLER, D. AND S. C. ZHU (2010): "The structure of factor content predictions," *Journal of International Economics*, 82, 195–207.
- YI, K.-M. (2010): "Can Multistage Production Explain the Home Bias in Trade?" *The American Economic Review*, 100, 364–393.