On the Geography of Global Value Chains*

Pol Antràs
Harvard University and NBER
pantras@fas.harvard.edu

Alonso de Gortari
Dartmouth College
alonso.degortari@gmail.com

January 9, 2020

Abstract

This paper develops a multi-stage general-equilibrium model of global value chains (GVCs) and studies the specialization of countries within GVCs in a world with barriers to international trade. With costly trade, the optimal location of production of a given stage in a GVC is not only a function of the marginal cost at which that stage can be produced in a given country, but is also shaped by the proximity of that location to the precedent and the subsequent desired locations of production. We show that, other things equal, it is optimal to locate relatively downstream stages of production in relatively central locations. We also develop and estimate a tractable, quantifiable version of our model that illustrates how changes in trade costs affect the extent to which various countries participate in domestic, regional or global value chains, and traces the real income consequences of these changes.

*We thank Arnaud Costinot and Iván Werning for useful conversations during the most upstream stages of this project. Roc Armenter, Rob Johnson, Myrto Kalouptsidi, Sam Kortum, Marc Melitz, Eduardo Morales, Felix Tintelnot, and Kei-Mu Yi provided very useful feedback on preliminary versions of the paper. We are also grateful for the editorial guidance of Fabrizio Zilibotti and the valuable input from five anonymous referees. We thank seminar audiences at Princeton, the North American Econometric Society Summer Meeting in Philadelphia, the SED in Toulouse, the Federal Reserve Bank of Dallas, Geneva, Chicago Booth, MIT, Stanford, Harvard, Brown, Clark, UCLA, Wisconsin, CREI-UPF, Autónoma in Barcelona, Toulouse, CEMFI, Toronto, HKU, HKUST, Rochester, Waseda, University of Tokyo, New Economic School, Notre Dame, Stockholm, Copenhagen, and UC Davis for useful feedback. Evgenii Fadeev, Le Kang, Daniel Ramos, BooKang Seol, Maria Voronina, and Diana Zhu provided outstanding research assistance at various stages of this paper's production chain. Antràs acknowledges support from the NSF (proposal #1628852). All errors are our own.

1 Introduction

In recent decades, technological progress and falling trade barriers have allowed firms to slice up their value chains, retaining within their domestic economies only a subset of the stages in these value chains. The rise of global value chains (GVCs) has dramatically changed the landscape of the international organization of production, placing the specialization of countries *within* GVCs at the center stage.

This paper studies how the comparative advantage of countries in specific segments of GVCs is determined in a world with barriers to international trade. The role of trade barriers on the geography of GVCs has been relatively underexplored in the literature, largely due to the technical difficulties that such an analysis entails. More specifically, characterizing the allocation of production stages to countries is not straightforward because the optimal location of production of a given stage in a GVC is not only a function of the marginal cost at which that stage can be produced in a given country, but is also shaped by the proximity of that location to the precedent and the subsequent desired locations of production. The aim of this paper is to develop tools to operationalize the study of the geography of GVCs in both partial equilibrium and general equilibrium environments.

We start off our analysis in section 2 by developing a simple partial equilibrium framework of multi-stage production in which a *lead firm* optimally chooses the location of its various production stages in an environment with costly trade. A key insight from our partial-equilibrium framework is that the relevance of geography (or trade costs) in shaping the location of the various stages of a GVC is more and more pronounced as one moves towards more and more downstream stages of a value chain. Intuitively, whenever trade costs are largely proportional to the gross value of the good being transported, these costs compound along the value chain, thus implying that trade costs erode more value added in downstream relative to upstream stages. In a parameterized example of our framework, this differential effect of trade costs takes the simple form of a stage-specific 'trade elasticity' that is increasing in the position of a stage in the value chain. The fact that trade costs are proportional to gross value follows from our iceberg formulation of these costs, a formulation that is not only theoretically appealing, but is also a reasonable approximation to reality.[1]

Having characterized the key properties of the solution to the lead-firm problem, we next show how it can be 'decentralized'. More specifically, we consider an environment in which there is no lead firm coordinating the chain, and instead stand-alone producers of the various stages in a GVC make cost-minimizing sourcing decisions by purchasing the good completed up to the prior stage from their least-cost source. The partial equilibrium of this decentralized economy coincides with the solution to the lead-firm problem – and in fact can be recast as a dynamic programming formulation of the lead-firm problem – but it is dramatically simpler to compute. For a chain entailing N stages with each of these stages potentially being performed in one of J countries, characterizing the J optimal GVCs that service consumers in each country requires only $J \times N \times J$ computations, instead of the

[1] The fact that import duties and insurance costs are approximately proportional to the value of the goods being shipped should be largely uncontroversial. For shipping costs, weight and volume are naturally also relevant, but as shown by Brancaccio, Kalouptsidi and Papageorgiou (2017), search frictions in the shipping industry allow shipping companies to extract rents from exporters by charging shipping fees that are increasing in the value of the goods in transit.

lead firm having to optimize over J^N potential paths for each of the J locations of consumption (for a total of $J \times J^N$ computations).

Although the results of our partial equilibrium model suggest that more central countries should have comparative advantage in relatively downstream stages within GVCs, formally demonstrating such a result requires developing a general-equilibrium model of GVCs in which production costs are endogenously determined and also shaped by trade barriers. With that goal in mind and also to explore the real income implications of changes in trade costs, in section 3 we develop a simple Ricardian model of trade in which the combination of labor productivity and trade cost differences across countries shapes the equilibrium position of countries in GVCs. More specifically, we adapt the Eaton and Kortum's (2002) Ricardian model of trade to a multi-stage production environment and derive sharp predictions for the *average* participation of countries in different segments of GVCs.

Previous attempts to extend the Ricardian model of trade to a multi-stage, multi-country environment (e.g., Yi, 2003, 2010, Fally and Hillberry, 2018, Johnson and Moxnes, 2019) have focused on the quantification of relatively low-dimensional models with two stages or a small number of countries. Indeed, as we describe in section 3, it is not obvious how to exploit the extreme-value distribution results developed by Eaton and Kortum (2002) in a multi-stage environment in which cost-minimizing location decisions are a function of the various cost 'draws' obtained by producers worldwide at various stages in the value chain. The reason for this is that neither the sum nor the product of Fréchet random variables are themselves distributed Fréchet, and thus previous approaches have been forced to resort to numerical analyses and simulated method of moments estimation.

We propose three alternative approaches to restore the tractability of the Eaton and Kortum (2002) framework in a Ricardian model with multi-stage production. The first approach consists in simply treating the *overall* (i.e., chain-level) unit cost of production of a GVC flowing through a sequence of countries as a draw from a Fréchet random variable with a location parameter that is a function of the states of technology and wage levels of *all* countries involved in that GVC, as well as of the trade costs incurred in that chain.[2]. The second approach maintains the standard assumption that labor productivity is stage-specific and drawn from a Fréchet distribution, but instead considers a decentralized equilibrium in which producers of a particular stage in a GVC have incomplete information about the productivity of certain suppliers upstream from them. More specifically, we assume that firms know their productivity and that of the suppliers immediately upstream from them (i.e., their tier-one suppliers) when they commit to sourcing from a particular supplier, but they do not know the precise productivity of their suppliers' suppliers (i.e., tier-two suppliers, tier-three suppliers, and so on). Finally, we develop an alternative decentralized approach inspired by the work of Oberfield (2018), in which technology is again specified at the stage level (rather than at the chain level), but in which productivity is buyer-seller specific. By appropriate choice of functional forms, we follow Oberfield (2018) in showing that this formulation can also deliver a Fréchet distribution of productivity *at the chain level*.

Interestingly, we find that these three alternative approaches are isomorphic, in the sense that

[2] A recent paper by Allen and Arkolakis (2019) adopts a similar path-specific representation of productivity in a very different setting

they yield the exact same equilibrium equations. More specifically, regardless of the microfoundation one chooses to invoke, we show in section 4 that our model generates a closed-form expression for the probability of any potential path of production constituting the cost-minimizing path to service consumers in any country. These probabilities are analogous to the trade shares in Eaton and Kortum (2002), and indeed our model nests their framework in the absence of multi-stage production. Our Ricardian multi-stage framework also delivers a simple formula relating real income to the relative prevalence of purely domestic value chains, a formula that generalizes the 'gains from trade' formula in Arkolakis et al. (2012). Although our set of general-equilibrium equations is a bit more cumbersome than in Eaton and Kortum (2002), we show how the proof of existence and uniqueness in Alvarez and Lucas (2007) can be easily (though tediously) adapted to our setting. Finally, we formally establish the existence of a centrality-downstreamness nexus, by which the average downstreamness of a country in GVCs should be increasing in this country's centrality (holding other determinants of comparative advantage constant). We close section 4 by providing suggestive empirical evidence for this centrality-downstreamness nexus and for a key mechanism of the model – namely, the fact that the elasticity of trade flows to distance is larger for downstream stages than for upstream stages.

In section 5, we generalize our framework along several dimensions, which permits our model to nest and better compare to various other Ricardian models of trade. More specifically, we introduce multiple industries and rich input-output links across these industries, and we allow for technologies that depend on whether an industry's output is used as final consumption or as an intermediate input for different industries. Exploiting properties of the distribution of final-good and input prices produced by the model, we show that the various versions of our model deliver closed-form expressions for final-good and input trade flows across countries, which can easily be mapped to the various entries of a world Input-Output table (or WIOT, for short). Various versions of these type of world Input-Output tables have become available in recent years, including the World Input Output Database, the OECD's TiVA statistics, and the Eora MRIO database.

In section 6, we leverage the tractability of our framework to back out the model's fundamental parameters from data on the various entries of a WIOT, when aggregated at the country level.[3] Our empirical approach constitutes a blend of calibration and estimation. First, we show that when abstracting from variation in domestic costs across countries, our equilibrium conditions unveil a simple way to back out the matrix of bilateral trade costs across countries from data on bilateral trade flows within and across countries. Our approach is akin to that in Head and Ries (2001), but it requires the use of only final-good trade flows. We also fix a key parameter that governs the shape of the Fréchet distributions of productivity to (roughly) match the aggregate trade elasticity implied by our model. Conditional on a set of countries J and a number of stages N, we then estimate the remaining parameters of the model via a generalized method of moments (GMM), in which we target all the entries of a WIOT. We perform this exercise using 2014 data from the World Input-Output Database, a source which is deemed to provide relatively high-quality and reliable data on intermediate input and final-good bilateral trade flows across countries for a sample of 43 countries

[3] In section 7.4, we estimate a two-sector version of our model. Estimating our full multi-industry model is not feasible with current computational constraints, as explained in section 6 (especially, footnote 34).

and the rest of the world. We find that the model is able to match the data very well.

Armed with estimates of the fundamental parameters of the model, we conclude the paper in section 7 by performing counterfactual exercises that illustrate how changes in trade barriers affect the extent to which various countries participate in domestic, regional or global value chains, and traces the real income consequences of these changes. We find that the gains from trade (i.e., the income losses from reverting to autarky) emanating from our model are, on average, 60% larger than those obtained from a version of our model without multi-stage production. This variant of our model is a generalization of the Eaton and Kortum (2002) model – akin to the work of Alexander (2017) – estimated to match all the entries of the WIOT. Similarly, we find that the real income losses from a hypothetical U.S.-China trade war are, for those two countries, almost 50% higher with sequential production than without it, though such a war may well benefit third countries.

Our paper most closely relates to the burgeoning literature on GVCs. On the theoretical front, in recent years a few theoretical frameworks have been developed highlighting the role of the sequentiality of production for the global sourcing decisions of firms. Among others, this literature includes the work of Harms, Lorz, and Urban (2012), Baldwin and Venables (2013), Costinot *et al.* (2013), Antràs and Chor (2013), Kikuchi *et al.* (2014), Fally and Hillberry (2018), and Tyazhelnikov (2016).[4] A key limitation of this body of theoretical work is that it either completely abstracts from modeling trade costs or it introduces such barriers in highly stylized ways, for instance assuming common trade costs across all country-pairs as in Baldwin and Venables (2013) or section 6.1 in Costinot *et al.* (2013). On the empirical front, a growing body of work, starting with the seminal work of Johnson and Noguera (2012), has been concerned with tracing the value-added content of trade flows and using those flows to better document the rise of GVCs and the participation of various countries in this phenomenon (see Koopman et al., 2014, Johnson, 2014, Timmer et al., 2014, de Gortari, 2019).[5] A parallel empirical literature has developed indices of the relative positioning of industries and countries in GVCs (see Fally, 2012, Antràs et al., 2012, Alfaro et al., 2015). There is also a prior body of work estimating the differential sensitivity of input trade and final-good trade to trade barriers, as exemplified by Bergstrand and Egger (2010) and Baldwin and Taglioni (2014), among others. On the quantitative side, and as mentioned above, our work builds on and expands on previous work by Yi (2003, 2010), Fally and Hillberry (2018), and Johnson and Moxnes (2019). Other authors, including Caliendo and Parro (2015), Alexander (2017), Antràs and Chor (2018), and Baqaee and Farhi (2019) have developed quantitative frameworks with Input-Output linkages across countries, but in models with a roundabout production structure without an explicit sequentiality of production. The connection between our framework and these previous contributions is further explored in de Gortari (2019), who blends several strands of this literature by generalizing the formulas on value-added content and downstreamness within the context of a multi-sector Armington model with sequential production. Finally, some implications of the rise of offshoring and GVCs for trade policy have been studied by Antràs and Staiger (2012) and Blanchard et al. (2018), but in

[4]This literature is in turn inspired by earlier contributions to modeling multi-stage production, such as Dixit and Grossman (1982), Sanyal and Jones (1982), Kremer (1993), Yi (2003) and Kohler (2004).

[5]An important precursor to this literature is Hummels et al. (2001), who combined international trade and Input-Output data to construct indices of vertical specialization.

much more stylized frameworks than studied in this paper.

The rest of the paper is structure as follows. Section 2 develops our partial equilibrium model and highlights some of its key features. Section 3 describes the assumptions of the general equilibrium model, and section 4 characterizes its equilibrium and provides suggestive empirical evidence for some of the key features of our model. Section 5 develops several extensions of our framework. Section 6 covers the estimation of our model and section 7 explores several counterfactuals. All proofs and several details on data sources and the estimation are relegated to the Supplementary Appendix and the Online Appendix.

2 Partial Equilibrium: Interdependencies and Compounding

In this section, we develop a simple model of firm behavior that formalizes the problem faced by a firm choosing the location of its various production stages in an environment with costly trade. For the time being, we consider the problem of a firm (or, more precisely, of a competitive fringe of firms) producing a particular good following a strictly sequential process. We defer a discussion of more general processes and of the general equilibrium aspects of the model to sections 3 and 5.

2.1 Environment

There are J countries in which consumers derive utility from consuming a final good. The good is produced combining N stages that need to be performed sequentially. The last stage of production can be interpreted as final assembly and is indexed by N. We will often denote the set of countries $\{1, ..., J\}$ by $\mathcal{J}$ and the set of production stages $\{1, ..., N\}$ by $\mathcal{N}$.

At each stage $n > 1$, production combines a local composite factor (which encompasses primitive factors of production and a bundle of materials), with the good finished up to the previous stage $n-1$. Production in the initial stage $n = 1$ only uses the composite factor. The cost of the composite factor varies across countries and is denoted by c_i in country i.[6] Countries also differ in their geography, as captured by a $J \times J$ matrix of iceberg trade coefficients $\tau_{ij} \geq 1$, where τ_{ij} denotes the units of the finished or unfinished good that need to be shipped from i for one unit to reach j. Firms are perfectly competitive and the optimal location $\ell(n) \in \mathcal{J}$ of the different stages $n \in \mathcal{N}$ of the value chain is dictated by cost minimization. Because of marginal-cost pricing, we will somewhat abuse notation and denote by $p^n_{\ell(n)}$ the unit cost of production of a good completed up to stage n in country $\ell(n)$. That good is available in country $\ell(n+1)$ at a cost $p^n_{\ell(n)} \tau_{\ell(n)\ell(n+1)}$.

Although the main results of this section extend to more general specifications of technology, in the main text we will restrict the analysis to the following functional form for the sequential cost

[6] For now we take this cost as given, but in the general equilibrium analysis in section 3, we will break c_i into the cost of labor and of a bundle of intermediate inputs we call *materials*. This will allow our model to encompass previous Ricardian models – and most notably Eaton and Kortum (2002) – featuring roundabout production. On the downside, this assumption precludes a study of the effects of GVC integration on wage inequality, as in the work of Basco and Mestieri (2019) or Lee and Yi (2019).

function associated with a path of production $\boldsymbol{\ell} = \{\ell(1), \ell(2), ..., \ell(N)\}$:

$$p^n_{\ell(n)}(\boldsymbol{\ell}) = \left(a^n_{\ell(n)} c_{\ell(n)}\right)^{\alpha_n} \left(p^{n-1}_{\ell(n-1)}(\boldsymbol{\ell}) \tau_{\ell(n-1)\ell(n)}\right)^{1-\alpha_n}, \text{ for all } n \in \mathcal{N}, \tag{1}$$

where $\alpha_n \in (0,1)$ denotes the cost share of the composite factor at stage n and $a^n_{\ell(n)}$ is the unit factor requirement at stage n in country $\ell(n)$. Because the initial stage of production uses solely the local composite factor, we have $\alpha_1 = 1$. Notice that the cost function in (1) is *Ricardian* in nature, in the sense that cross-country differences in technology are associated with differences in the efficiency with which the local composite factor is used in different stages. The choice of this specification will permit a particularly sharp characterization of the key results of this section, and is important for tractability in the general-equilibrium version of the model in section 3.

Note that equation (1) also applies to the final assembly stage N, and a good completed in $\ell(N)$ after following the path $\boldsymbol{\ell}$ is available in any country j at a cost $p^F_j(\boldsymbol{\ell}) = p^N_{\ell(N)}(\boldsymbol{\ell}) \tau_{\ell(N)j}$ (we use the superscript F to denote finished goods). For each country $j \in \mathcal{J}$, the goal of the firm is then to choose the optimal path of production $\boldsymbol{\ell}^j = \left\{\ell^j(1), \ell^j(2), ..., \ell^j(N)\right\} \in \mathcal{J}^N$ that minimizes the cost $p^F_j(\boldsymbol{\ell})$ of providing the good to consumers in that country j. The remainder of this section will seek to characterize the solution to this problem.

2.2 *Lead-Firm* Problem

We consider first the problem of a *lead firm* choosing the location of production of all stages $n \in \mathcal{N}$, in order to minimize the overall cost of serving consumers in a given country j. Using $p^F_j(\boldsymbol{\ell}) = p^N_{\ell(N)}(\boldsymbol{\ell}) \tau_{\ell(N)j}$ and iterating (1), this problem reduces to:

$$\boldsymbol{\ell}^j = \arg\min_{\boldsymbol{\ell} \in \mathcal{J}^N} p^F_j(\boldsymbol{\ell}) = \arg\min_{\boldsymbol{\ell} \in \mathcal{J}^N} \left\{ \prod_{n=1}^{N} \left(a^n_{\ell(n)} c_{\ell(n)}\right)^{\alpha_n \beta_n} \times \prod_{n=1}^{N-1} \left(\tau_{\ell(n)\ell(n+1)}\right)^{\beta_n} \times \tau_{\ell(N)j} \right\} \tag{2}$$

where

$$\beta_n \equiv \prod_{m=n+1}^{N} (1 - \alpha_m), \tag{3}$$

and where we use the convention $\prod_{m=N+1}^{N} (1-\alpha_m) = 1$. Thus, α_n is the share of stage-n composite factor in stage-n production, while $\alpha_n \beta_n$ is the share of the stage-n composite factor in the production of the finished good (i.e., its share in the global value chain). Note that $\sum_{n=1}^{N} \alpha_n \beta_n = 1$.

It is worth highlight two important features of program (2). First, notice that when trade costs are identical for all country-pairs (i.e., $\tau_{ij} = \tau$ for all i and j), the last two terms reduce to a constant that is independent of the path of production. In such a case, we can break the cost-minimization problem in (2) into a sequence of N independent cost-minimization problems in which the optimal location of stage n is simply given by $\ell^j(n) = \arg\min_i \{a^n_i c_i\}$, and is thus independent of the country of consumption j. Notice, however that this result requires no differences between internal and external trade costs (i.e., $\tau_{ij} = \tau$ also for $i = j$), and thus this case is isomorphic, up to a productivity shifter, to an environment with costless trade. With a general geography of trade costs, a lead firm can no longer perform cost minimization independently stage by stage, and instead it needs to optimize over the whole path of production. Intuitively, the location $\ell(n)$ minimizing production costs $a^n_{\ell(n)} c_{\ell(n)}$ might not be part of a firm's optimal path if the optimal locations for stages $n-1$ and $n+1$ are sufficiently far from $\ell(n)$. A direct implication of this result is that the presence of arbitrary trade costs turns a problem of dimensionality $N \times J$ into J much more complex problems of dimensionality J^N each. As we will see below, however, the dimensionality of program (2) can be dramatically reduced using dynamic programming.

A second noteworthy aspect of the minimand in equation (2) is that the trade-cost elasticity of the unit cost of serving consumers in country j increases along the value chain. More specifically, note from equation (3) that, because $\alpha_n > 0$ for all n, we have $\beta_1 < \beta_2 < ... < \beta_N = 1$. The reason for this compounding effect of trade costs stems from the fact that the costs of transporting goods have been modeled (realistically, as we argued in the Introduction) to be proportional to the gross value of the good being transacted, rather than being assumed proportional to the value added at that stage. Thus, as the value of the good rises along the value chain, so does the amount of resources used to transport the goods across locations.[7]

An implication of this compounding effect is that, in choosing their optimal path of production, firms will be more concerned about reducing trade costs in relatively downstream stages than in relatively upstream stages. As we will demonstrate when exploring the general equilibrium of our model, this feature of the cost function will tend to generate a centrality-downstreamness nexus by which, *ceteris paribus*, relatively more central countries will tend to gain comparative advantage and specialize in relatively downstream stages. At this point, however, this result may not appear to be entirely trivial because if α_n were to rise sufficiently fast with n, one could envision firms being more concerned with reducing the cost of the composite factor than with reducing trade costs for the most downstream stages. This would work against the centrality-downstreamness nexus if, plausibly, central locations tend to feature a higher price c_i for their local composite factor.

Although we have derived this compounding effect of trade costs for the case of Ricardian technological differences and Cobb-Douglas cost functions, we show in Appendix A.1.1 that the same result applies for arbitrary constant-returns-to-scale technologies. More specifically, denoting by β_n the elasticity of $p^F_j(\boldsymbol{\ell})$ with respect to $\tau_{\ell(n)\ell(n+1)}$, we show that β_n is again necessarily non-decreasing in n even when these elasticities are not pinned down by exogenous parameters. Thus, the result that firms will be more concerned about minimizing trade costs in downstream stages than in upstream stages is quite general.[8] We summarize this result as follows:

Proposition 1 *The elasticity of the overall cost of production with respect to trade costs incurred at stage n increases along the chain (i.e., increases with n).*

[7] For the particular case in which overall value added is a symmetric Cobb-Douglas aggregator of the value added of all stages (i.e., $\alpha_n \beta_n = 1/N$, for all n), the trade-cost elasticity equals $\beta_n = n/N$ and thus increases linearly with the downstreamness n of a stage.

[8] For example, for the case of a symmetric Leontief technology and production costs equal to 1 in all countries and stages (i.e., $a^n_{\ell(n)} c_{\ell(n)} = 1$ for all n and $\ell(n)$), we obtain

$$p^F_j(\boldsymbol{\ell}) = \tau_{\ell(N)j} + \tau_{\ell(N)j}\tau_{\ell(N-1)\ell(N)} + \tau_{\ell(N)j}\tau_{\ell(N-1)\ell(N)}\tau_{\ell(N-2)\ell(N-1)} + \tau_{\ell(N)j}\tau_{\ell(N-1)\ell(N)}\tau_{\ell(N-2)\ell(N-1)}\tau_{\ell(N-3)\ell(N-2)} + \cdots,$$

which again illustrates the larger relative importance of downstream trade costs.

2.3 Decentralization and Dynamic Programming

We have so far characterized the problem of a *lead firm* with full information on the productivity of the various potential worldwide producers of each stage n. This characterization relies on strong informational assumptions, so we now consider an alternative environment in which no individual firm coordinates the whole value chain. Instead, we assume that a value chain consists of a series of stage-specific producers that simply minimize their cost of production taking into account their composite factor cost, their productivity, and the cost at which they can obtain the good finished up to the immediately preceding stage. Similarly, consumers in country j simply purchase the final good from whichever stage-N producer worldwide can provide the finished good at the lowest price.

From equation (1), a producer of stage n in country $\ell(n)$ would choose to procure the good finished up to stage $n-1$ from the location $\ell(n-1) \in \mathcal{J}$ that solves

$$p^n_{\ell(n)} = \min_{\ell(n-1)\in\mathcal{J}} \left\{ \left(a^n_{\ell(n)} c_{\ell(n)}\right)^{\alpha_n} \left(p^{n-1}_{\ell(n-1)} \tau_{\ell(n-1)\ell(n)}\right)^{1-\alpha_n} \right\}, \tag{4}$$

where $p^{n-1}_{\ell(n-1)}$ is the minimum (free-on-board) price charged by producers of stage $n-1$ in country $\ell(n-1)$. Importantly, notice that the problem in (4) amounts to minimizing sourcing costs $p^{n-1}_{\ell(n-1)}\tau_{\ell(n-1)\ell(n)}$ regardless of the composite factor cost $c_{\ell(n)}$ and productivity $a^n_{\ell(n)}$, and of the future path of the good after flowing through $\ell(n)$ at stage n. Producers of the initial stage only use their local composite factor and thus $p^1_{\ell(1)} = a^1_{\ell(1)} c_{\ell(1)}$. In sum, in this decentralization, we have $J \times N$ agents computing the cost of J possible sources of goods and picking the minimum one.

Once we have solved these $J \times N$ simple problems, it is then straightforward to use forward induction to solve for the resulting equilibrium path $\boldsymbol{\ell}^j = \left\{\ell^j(1), \ell^j(2), ..., \ell^j(N)\right\} \in \mathcal{J}^N$ for each destination market $j \in \mathcal{J}$. To see this, begin by computing $p^1_{\ell(1)} = a^1_{\ell(1)} c_{\ell(1)}$ for each possible initial location of stage $n=1$. Then, for each of the J potential locations of stage $n=2$, we can use (4) to solve for corresponding optimal stage-1 location and associated cost up to stage $n=2$. Importantly, the same solution applies regardless of the final destination of consumption j. We can then use these J resulting costs (up to stage $n=2$) to proceed to stage $n=3$ and solve for the optimal location of stage $n=2$ for each potential location of $n=3$. We can then move to $n=4$, to $n=5$, and so on, until reaching final consumption, which delivers the optimal location of final assembly $\ell^j(N)$ for each destination market j. More formally, if we define $\boldsymbol{\ell}^j_n \in \mathcal{J}^n$ as the optimal sequence for delivering the good completed up to stage n to producers in country j – with a particular sequence given by $\boldsymbol{\ell}^j_n = \left\{\ell^j(1), \dots, \ell^j(n)\right\}$ – this sequence can be found recursively for all $n = 2, \dots, N$ by simply solving

$$\ell^j(n) = \arg\min_{k\in\mathcal{J}} \left\{ p^n_k\left(\boldsymbol{\ell}^k_{n-1}\right)\tau_{kj}\right\},$$

with $\boldsymbol{\ell}^j_n = \left\{\boldsymbol{\ell}^{\ell^j(n)}_{n-1}, \ell^j(n)\right\}$. For $n=1$, we have $\boldsymbol{\ell}^k_0 = \varnothing$ for all $k \in \mathcal{J}$ and the price depends only the composite factor cost: $p^1_k(\varnothing) = a^1_k c_k$.

A natural question is then how the solution to this decentralization problem relates to that obtained from the *lead firm* problem in equation (2). It turns out that as long as technology features constant returns to scale (as in our particular Cobb-Douglas formulation in (1)), these two solutions will in fact coincide (see Appendix A.1.2). Intuitively, with constant returns to scale, the identity of the specific firms making these decisions along the chain is immaterial, so the recursive formulation of the problem in (4) is entirely consistent with our previous *lead firm* using dynamic programming to solve for the optimal path of production leading to consumption in each country $j \in \mathcal{J}$. More specifically, instead of solving program (2) in a brute force manner, the *lead firm* breaks the problem into a series of stage- and country-specific optimal sourcing problems, and then solves the problem via forward induction (starting in the most upstream stage). Equation (4) constitutes the Bellman equation associated with this problem, and invoking the principle of optimality, we can then establish that the resulting optimal path of production $\boldsymbol{\ell}^j = \boldsymbol{\ell}_N^j = \left\{\ell^j(1), \ell^j(2), ..., \ell^j(N)\right\} \in \mathcal{J}^N$ that minimizes the cost $p_j^F(\boldsymbol{\ell})$ in this decentralized formulation of the problem will coincide with the one obtained solving the *lead-firm* problem in (2) by exhaustive search.

A key advantage of this decentralized or dynamic programming approach is that it only requires $J \times N \times J$ computations to obtain the optimal production path for *all* destinations of final consumption, instead of having to optimize over J^N potential paths for each country j (see Tyazhelnikov, 2016, for a contemporaneous derivation of this result). For example, with 200 countries and 5 stages, this amounts to only $200,000$ computations rather than 64 trillion computations.[9] Although it might be clear from our discussion above, it is worth stressing that the isomorphism between the lead-firm problem and the decentralized problem holds true for *any* constant-returns-to-scale technology, and not only for the Cobb-Douglas one in (1).[10] We summarize this result as:

Proposition 2 *The solution $\boldsymbol{\ell}^j$ to the lead firm cost-minimization problem in (2) coincides with the equilibrium path of a decentralized production chain in which $J \times N$ agents solve the stage cost-minimization problem in (4).*

Before turning to the general equilibrium of our framework, it is important to stress that Proposition 2 relies crucially on the assumption that technology features constant returns to scale. With variable returns to scale, matters would become significantly more complicated. First, the *lead firm* problem could not be solved independently for each destination market j, because whether a location ℓ constitutes a cost-minimizing location for stage n in a particular chain ending in j will be a function of the scale of this production node, and the latter is shaped by the overall level of production flowing through this node (potentially involving chains ending in destination markets other than j). As a result, dynamic programming ceases to be a powerful tool to simplify the problem (see de Gortari, 2020). Second, and relatedly, in a decentralized equilibrium, agents will fail to internalize the effect of their sourcing decisions on marginal costs (via scale effects), and the equilibrium paths might deviate from the ones the *lead firm* would choose.

[9] Though the dimensionality of the lead firm's problem is huge, for the particular case with Cobb-Douglas technologies, in Appendix A.1.2 we show that the problem can also be written as a zero-one integer programming problem, for which many extremely quick and efficient algorithms are available (see, for instance, http://www.gurobi.com).

[10] In Online Appendix B.1, we illustrate some of the salient and distinctive features of our partial model via a simple example featuring four countries ($J = 4$) and four stages ($N = 4$).

3 General Equilibrium Model

We next embed our model of firm behavior into a general equilibrium model of trade along the lines of the multi-country Ricardian model of trade of Eaton and Kortum (2002).

3.1 Environment

We continue to assume a world with J countries (indexed by i or j) where consumers now derive utility from consuming a continuum of final-good varieties (indexed by z). Preferences are CES and given by

$$u\left(\left\{y_i^N(z)\right\}_{z=0}^1\right) = \left(\int_0^1 \left(y_i^N(z)\right)^{(\sigma-1)/\sigma} dz\right)^{\sigma/(\sigma-1)}, \quad \sigma > 1. \tag{5}$$

Production of each of the final-good varieties is as described in the previous section: all production processes entail N sequential stages (indexed by n) and are characterized by the Ricardian, Cobb-Douglas specification in (1). We let countries differ in three key aspects: (i) their technological efficiency, as determined by the unit composite factor requirements $a_i^n(z)$, (ii) their geography, as captured by a $J \times J$ matrix of iceberg trade cost $\tau_{ij} \geq 1$, and (iii) their size, as reflected by the measure L_i of 'equipped' labor available for production in each country i (labor is inelastically supplied and commands a wage w_i). In section 5, we will demonstrate that our framework remains tractable even when relaxing many of these assumptions.

The local composite factor used at each stage comprises labor and an aggregator of final-good varieties that corresponds exactly to the CES aggregator in (5). This amounts to assuming that part of final-good production is not absorbed by consumers, but rather by firms that use those goods as a bundle of materials. More specifically, we let the cost c_i of the composite factor in country i be captured by a Cobb-Douglas aggregator $c_i = (w_i)^\gamma (P_i)^{1-\gamma}$, where P_i is the ideal price index associated with the CES aggregator in (5). This roundabout structure of production is standard in recent quantitative Ricardian models (see Eaton and Kortum, 2002, Alvarez and Lucas, 2007, or Caliendo and Parro, 2014), and constitutes a particularly convenient way for these models to introduce intermediate input trade flows and a gross-output to value-added ratios higher than one. Although our model features intermediate input flows across countries even in the absence of these production 'loops', we still adopt this formulation for comparability (especially when evaluating our model quantitatively in section 7).[11]

This completes the discussion of the structure of our general-equilibrium model. In principle, given values for the unit composite factor requirements $a_i^n(z)$ and all other primitive parameters, the equilibrium of the model could be computed by (i) solving for the cost-minimizing path of production for each good z and each destination of consumption j given a vector of wages, and (ii) invoking labor-market clearing to reduce equilibrium wages to the solution of a fixed point problem.

[11] In recent work, Fally and Sayre (2018) and Baqaee and Farhi (2019) have argued that the elasticity of substitution between value added and materials is likely to be lower than one, especially in the short run. For the same reason, one could argue that our Cobb-Douglas sequential production process in (1) features too much substitution across stages. Nevertheless, like most trade models, ours is a model of the long run, so it is much less clear to us that a Cobb-Douglas assumption biases our quantitative estimates in any particular direction.

Such an approach, however, would not be particularly useful in order to formally characterize certain features of the equilibrium or to estimate the model in a computationally feasible and transparent manner. With that in mind, we next explore a particularly convenient parameterization of the unit factor requirements $a_i^n(z)$.

3.2 Technology

Building on the seminal work of Eaton and Kortum (2002), we propose a probabilistic specification of the unit factor requirements $a_i^n(z)$ that delivers a remarkably tractable multi-stage, multi-country Ricardian model. We are certainly not the first ones to explore such a multi-stage extension of the Eaton and Kortum (2002) framework. Yi (2010) and Johnson and Moxnes (2019), for instance, consider a 'natural' extension in which each productivity parameter $1/a_i^n(z)$ is assumed stochastic and drawn independently (across goods and stages) from a type II (or Fréchet) extreme-value probability distribution, as in Eaton and Kortum (2002). A key limitation of their approach is that the minimum cost associated with a given GVC path is not characterized by a particularly tractable distribution. The reason for this is that, although the minimum of a series of Fréchet draws is itself distributed Fréchet, the product of Fréchet random variables in our Cobb-Douglas cost function in (2) is *not* distributed Fréchet.[12] As a result, these papers need to resort to numerical methods to approximate the solution of their models, even when restricting the analysis to two-stage chains. We instead develop three alternative approaches that all permit a sharp and exact characterization of some of the features of the equilibrium for an arbitrary number of stages, and that will be amenable to structural, generalized method of moments estimation using world Input-Output tables.

A. *Lead-Firm* Approach with Chain-Level Productivity

We begin by revisiting the problem of a *lead firm* choosing the location of the various stages of production with full knowledge of the realized unit requirements $a_i^n(z)$ for each stage $n \in \mathcal{N}$ and each country $i \in \mathcal{J}$. The key conceptual innovation we propose, relative to Eaton and Kortum (2002) and previous multi-stage extensions of that framework, is to think about productivity at the *chain level* rather than at the stage level. More specifically, a given production path $\boldsymbol{\ell} = \{\ell(1), \ell(2), ..., \ell(N)\} \in \mathcal{J}^N$ will be associated with a *chain-level* production cost that is naturally a function of trade costs, composite factor costs and the state of technology of the various countries involved in the chain. Yet, two chains flowing across the same countries in the exact same order may not achieve the same overall productivity due to (unmodeled) idiosyncratic factors, such as compatibility problems, production delays, or simple mistakes. More formally, and building on the cost function in (2), we assume that the overall productivity of a given chain $\boldsymbol{\ell}$ is characterized by

$$\Pr\left(\prod_{n=1}^{N}\left(a_{\ell(n)}^{n}(z)\right)^{\alpha_n\beta_n} \geq a\right) = \exp\left\{-a^{\theta}\prod_{n=1}^{N}\left(T_{\ell(n)}\right)^{\alpha_n\beta_n}\right\}, \quad (6)$$

[12] Assuming a Leontief cost function (i.e., perfect complementarity) does not provide tractability either because the sum of Fréchet random variables is not distributed Fréchet either.

which amounts to assuming that $\prod_{n=1}^{N}\left(a^n_{\ell(n)}(z)\right)^{-\alpha_n\beta_n}$ is distributed Fréchet with a shape parameter given by θ, and a location parameter that is a function of the states of technology in all countries in the chain, as captured by $\prod_{n=1}^{N}\left(T_{\ell(n)}\right)^{\alpha_n\beta_n}$. A direct implication of this assumption is that the unit cost associated with serving consumers in a given country j via a given chain $\boldsymbol{\ell}$ is also distributed Fréchet. More precisely, denoting by $p^F_j(\boldsymbol{\ell}, z)$ the price paid by consumers in j for a good z produced following the path $\boldsymbol{\ell}$, we have

$$\Pr\left(p^F_j(\boldsymbol{\ell}, z) \geq p\right) = \exp\left\{-p^{\theta} \times \prod_{n=1}^{N}\left(\left(c_{\ell(n)}\right)^{-\theta} T_{\ell(n)}\right)^{\alpha_n\beta_n} \times \prod_{n=1}^{N-1}\left(\tau_{\ell(n)\ell(n+1)}\right)^{-\theta\beta_n} \times \left(\tau_{\ell(N)j}\right)^{-\theta}\right\}, \tag{7}$$

independently of the final good z under consideration. This result will be key for neatly characterizing the equilibrium, as we will show in section 4.

B. Decentralized Approaches with Stage-Level Productivity

Although we adopt the 'Fréchet-in-the-chain' specification in (6) for tractability, we next outline two alternative approaches that deliver an identical set of equilibrium conditions to those we will derive under the specification in (6). In both cases, we explore an environment akin to the decentralized equilibrium developed in section 2.3, in which stage-specific producers at each stage $n \in \mathcal{N}$ simply attempt to minimize the sourcing cost $p^{n-1}_{\ell(n-1)}$ of the good completed up to the prior stage $n-1$. To save space, we relegate the technical details of these alternative approaches to Appendix A.2.

The first approach is close in spirit to the stage-specific productivity randomness in Yi (2010) and Johnson and Moxnes (2016), and assumes that $1/a^n_{\ell(n)}(z)$ is drawn independently (across goods and stages) from a Fréchet distribution. If stage-specific producers were to decide on their source of inputs with precise information on the alternative sourcing costs $p^{n-1}_{\ell(n-1)}$ available to them, the isomorphism result in Proposition 2 would apply, and the resulting distribution of final-good prices $p^F_j(\boldsymbol{\ell}, z)$ would be a function of the product of Fréchet distributions, and thus hard to characterize.[13] In order to make this alternative approach tractable, we assume that producers at each stage n and location $\ell(n)$ do not observe realized upstream prices $p^{n-1}_{\ell(n-1)}$ before making sourcing decisions, and can only forecast these prices based on information on the productivity levels of their potential direct (or tier-one) suppliers in various countries. These tier-one supplier productivity levels are not sufficient statistics for sourcing prices because upstream marginal costs also depend on the productivity of suppliers further upstream (i.e., tier-two suppliers, tier-three suppliers and so on). The idea behind this formulation is that firms need to pre-commit to purchase from particular suppliers based on information they gather from inspecting (e.g., through factory visits) all their potential immediate suppliers. Yet, ex-post, a supplier's marginal cost might be higher or lower than expected because this supplier may face unexpectedly high or low sourcing costs itself.[14]

[13] By Proposition 2, tractability would evidently be restored if one assumed the chain-level technology in (6). The goal of this section, however, is to explore variants of our model in which technology is specified at the stage level.

[14] The approach of building some form of incomplete information (or ex-ante uncertainty) into the Eaton and Kortum (2002) framework is similar in spirit to the one pursued by Tintelnot (2017) and Antràs, Fort and Tintelnot (2017).

Although the pre-commitment to buy from a particular source naturally affects the nature of ex-post competition, we assume that buyers have all the bargaining power and continue to be able to source upstream inputs at marginal cost. As we show in the Appendix, in this decentralized economy with incomplete information, the equilibrium price $p_j^F(\boldsymbol{\ell}, z)$ paid by consumers in j for a good z produced following the path $\boldsymbol{\ell}$ is Fréchet distributed. Furthermore, by appropriate choice of the shape and scale parameter of the stage-specific distributions of $1/a_{\ell(n)}^n(z)$, this alternative formulation delivers an expression for the equilibrium distribution of prices $p_j^F(\boldsymbol{\ell}, z)$ that is identical to that obtained in equation (7) in our *lead-firm* approach.

The second decentralized approach is inspired by recent work of Oberfield (2018).[15] This approach again begins by specifying a distribution of productivity $1/a_{\ell(n)}^n(z)$ at the stage level, but assumes that this productivity is now buyer-seller specific (or *match* specific) and characterized by a Pareto distribution with shape parameter θ. Furthermore, each producer (or buyer) chooses the best match among a pool of potential suppliers at $n-1$, with the number of available potential suppliers in each source country k being characterized by a Poisson distribution with a mean related to the parameter $T_{\ell(n)}$. For the case of a closed economy with an infinite loop of stages (i.e., when stage 1 producers buy inputs from other producers), Oberfield (2018) showed that this formulation delivers a Fréchet distribution for chain-level productivity. In Appendix A.2, we show that his result generalizes to our multi-country environment with an arbitrary finite number of stages, provided that the distribution of productivity in the initial stage ($n = 1$) is Fréchet distributed. Furthermore, as long as all transactions are priced at marginal cost, by appropriate choice of the parameters governing the Pareto and Poisson distributions at $n > 1$ and the Fréchet distribution at $n = 1$, this alternative decentralized approach again delivers a distribution for the final-good prices $p_j^F(\boldsymbol{\ell}, z)$ that is Fréchet and given by the same expression (7) as in our *lead-firm* approach (see Appendix A.2 for details).

4 Characterization of the Equilibrium

Having presented and discussed our assumptions on technology, in this section we characterize the general equilibrium of our model. We proceed in five steps. First, we leverage our extreme-value representation of GVC productivity to obtain closed-form expressions for the relative prevalence (in value terms) of different GVCs in the world equilibrium, and for the aggregate price index in each country. Second, we study the existence and uniqueness of the general equilibrium. Third, we derive an expression for the gains from trade in our model and compare it to the one obtained by Eaton and Kortum (2002). Fourth, we formalize the link between downstreamness and centrality that we hinted at in section 2. Finally, we present some suggestive evidence for this relationship and for a key mechanism in the model. Throughout this section, we focus on deriving our results for the 'Fréchet-in-the-chain' specification in (6), with the understanding (formalized in Appendix A.2) that our two alternative decentralized approaches lead to the exact same equilibrium conditions.[16]

[15] We are grateful to Sam Kortum for suggesting this alternative approach to us.

[16] As pointed out by a referee, it is also possible to recast our one-sector model into a multi-industry model with input-output links à la Caliendo and Parro (2015), in which different stages are treated as separate industries. Nevertheless, because the output of these stage-specific industries is not observed in the data, this isomorphism is of limited use in

4.1 Relative Prevalence of Different GVCs and Equilibrium Prices

Consider the lead-firm version of our model, in which the price paid by consumers in j for a good produced following the path $\boldsymbol{\ell} \in \mathcal{J}^N$ is given by the Fréchet distribution in (7). In such a case, we can readily invoke some key results in Eaton and Kortum (2002) to characterize the equilibrium prices and the relative prevalence of different GVCs. First, it is straightforward to verify that the probability of a given GVC $\boldsymbol{\ell}$ being the cost-minimizing production path for serving consumers in j is given by

$$\pi_{\boldsymbol{\ell}j} = \frac{\prod_{n=1}^{N-1} \left(\left(T_{\ell(n)}\right)^{\alpha_n} \left(\left(c_{\ell(n)}\right)^{\alpha_n} \tau_{\ell(n)\ell(n+1)} \right)^{-\theta} \right)^{\beta_n} \times \left(T_{\ell(N)}\right)^{\alpha_N} \left(\left(c_{\ell(N)}\right)^{\alpha_N} \tau_{\ell(N)j} \right)^{-\theta}}{\Theta_j}, \tag{8}$$

where

$$\Theta_j = \sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \prod_{n=1}^{N-1} \left(\left(T_{\ell(n)}\right)^{\alpha_n} \left(\left(c_{\ell(n)}\right)^{\alpha_n} \tau_{\ell(n)\ell(n+1)} \right)^{-\theta} \right)^{\beta_n} \times \left(T_{\ell(N)}\right)^{\alpha_N} \left(\left(c_{\ell(N)}\right)^{\alpha_N} \tau_{\ell(N)j} \right)^{-\theta}, \tag{9}$$

and where remember that $c_i = (w_i)^{\gamma} (P_i)^{1-\gamma}$. With a unit measure of final goods, $\pi_{\boldsymbol{\ell}j}$ also corresponds to the share of GVCs ending in j for which $\boldsymbol{\ell}$ is the cost-minimizing production path.[17]

Second, and as already anticipated in equation (7), the distribution of final-good prices $p_j^F(\boldsymbol{\ell}, z)$ paid by consumers in j satisfies $\Pr\left(p_j^F(\boldsymbol{\ell}, z) \leq p\right) = 1 - \exp\left\{-\Theta_j p^{\theta}\right\}$ and is thus independent of the path of production $\boldsymbol{\ell}$. Consequently, the probabilities $\pi_{\boldsymbol{\ell}j}$ in equation (8) above also constitute the shares of country j's income spent on final goods produced under all possible paths $\boldsymbol{\ell} \in \mathcal{J}^N$.

As is clear from equation (8), GVCs that involve countries with higher states of technology T_i or lower composite factor costs c_i will tend to feature disproportionately in production paths leading to consumption in j. Furthermore, and consistently with Proposition 1, high trade costs penalize the participation of countries in GVCs, but such an effect is disproportionately large for downstream stages relative to upstream stages. This is captured by the fact that the 'trade elasticity' associated with stage n is given by $\theta\beta_n$, and β_n is increasing in n with $\beta_N = 1$.

Following the same steps as in Eaton and Kortum (2002), we can further solve for the exact ideal price index P_j in country j associated with (5):

$$P_j = \kappa \left(\Theta_j\right)^{-1/\theta}, \tag{10}$$

where $\kappa = \left[\Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)\right]^{1/(1-\sigma)}$ and Γ is the gamma function. For the price index to be well defined, we impose $\sigma - 1 < \theta$.

So far, we have just described how to adapt the Eaton and Kortum (2002) toolbox to derive an explicit formula for the relative prevalence of various production paths (or GVCs) in delivering consumption goods to a given country. We will next demonstrate that this characterization is

estimating our model and performing counterfactuals.

[17] Note that when $N = 1$, we necessarily have $\alpha_N = 1$, and the formulas (8) and (9) collapse to the well-know trade share formulas in Eaton and Kortum (2002).

sufficient to derive interesting implications for the gains from trade (and the costs of trade wars) from our framework, and also to formalize a link between centrality and GVC position. Nevertheless, it is important to emphasize that these 'GVC trade shares' are not observable in the data, thus creating a challenge for the estimation of the model's parameters. In section 5, we will demonstrate that it is relatively straightforward to map objects derived from the model (e.g., final-good and intermediate input flows across countries) to the type of information available in world Input-Output tables.

4.2 General Equilibrium

We have thus far characterized the relative prevalence of different GVCs as a function of the vectors of equilibrium wages $\boldsymbol{w} = (w_1, ..., w_J)$ and input bundle costs $\boldsymbol{P} = (P_1, ..., P_J)$ in all countries. We next describe how these vectors are pinned down in general equilibrium.

Notice first that invoking (10) and $c_i = (w_i)^{\gamma} (P_i)^{1-\gamma}$, we can solve for the vector $\boldsymbol{P}$ as a function of the vector $\boldsymbol{w}$ from the system of equations:

$$P_j = \kappa \left(\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \prod_{n=1}^{N-1} \left(\left(T_{\ell(n)}\right)^{\alpha_n} \left(\left(c_{\ell(n)}\right)^{\alpha_n} \tau_{\ell(n)\ell(n+1)} \right)^{-\theta} \right)^{\beta_n} \times \left(T_{\ell(N)}\right)^{\alpha_N} \left(\left(c_{\ell(N)}\right)^{\alpha_N} \tau_{\ell(N)j} \right)^{-\theta} \right)^{-1/\theta}, \tag{11}$$

for all $j \in \mathcal{J}$.

To solve for equilibrium wages, notice that for all GVCs, stage-n value added (or labor income) accounts for a share $\gamma\alpha_n\beta_n$ of the value of the finished good emanating from that GVC. Furthermore, total spending in any country j is given by the sum of final-good spending ($w_j L_j$) and spending in the intermediate input bundle ($w_j L_j \times (1-\gamma)/\gamma$). The share of that spending by j going to GVCs in which country i is in position n is given by $\Pr(\Lambda_i^n, j) = \sum_{\boldsymbol{\ell} \in \Lambda_i^n} \pi_{\boldsymbol{\ell} j}$, where $\Lambda_i^n = \left\{ \boldsymbol{\ell} \in \mathcal{J}^N \mid \ell(n) = i \right\}$ and $\pi_{\boldsymbol{\ell} j}$ is given in equation (8). It thus follows that the equilibrium wage vector is determined by the solution of the following system of equations

$$w_i L_i = \sum_{j \in \mathcal{J}} \sum_{n \in \mathcal{N}} \alpha_n \beta_n \times \Pr(\Lambda_i^n, j) \times w_j L_j. \tag{12}$$

The system of equations is nonlinear because $\Pr(\Lambda_i^n, j)$ is a nonlinear function of wages themselves, and of the vector $\boldsymbol{P}$, which is in turn a function of the vector of wages $\boldsymbol{w}$. When $N = 1$, we have that $\alpha_N\beta_N = 1$ and $\Pr(\Lambda_i^n, j) = \pi_{ij} = (\tau_{ij} c_i)^{-\theta} T_i / \sum_k (\tau_{kj} c_k k)^{-\theta} T_k$. The equilibrium then reduces to the general equilibrium in Eaton and Kortum (2002) and Alvarez and Lucas (2007).

In Online Appendix B.2, we build on Alvarez and Lucas (2007) to show that, given a vector of wages $\mathbf{w}$, the system of equations in (11) delivers a unique vector of input bundle costs $\boldsymbol{P}$ even when $N > 1$. In that Appendix, we also demonstrate the existence of a solution $\mathbf{w}^* \in \mathbb{R}_{++}^J$ to the system of equations in (12) – with (11) plugged in – and we derive a set of sufficient conditions that ensure that this solution is unique.

4.3 Gains from Trade

We next study the implications of our framework for how changes in trade barriers affect real income in all countries. Consider a 'purely-domestic' value chain that performs all stages in a given country j to serve consumers in the same country j. Let us denote this domestic chain by $\boldsymbol{j} = (j, j, ..., j)$. From equation (8), such a value chain would capture a share of country j's spending equal to

$$\pi_{\boldsymbol{j}j} = \frac{(\tau_{jj})^{-\theta \sum_{n=1}^{N} \beta_n} \times (c_j)^{-\theta} T_j}{\Theta_j},$$

where we have used the fact that $\sum_{n=1}^{N} \alpha_n \beta_n = 1$. Combining this equation with (10) and $c_j = (w_j)^{\gamma} (P_j)^{1-\gamma}$, we can express real income in country j as

$$\frac{w_j}{P_j} = \left(\kappa \left(\tau_{jj}\right)^{\sum_{n=1}^{N} \beta_n} \right)^{-1/\gamma} \left(\frac{T_j}{\pi_{\boldsymbol{j}j}} \right)^{1/(\theta\gamma)}.$$

Consider now a change in trade costs, while holding all other technological parameters (α_n, β_n, T_j, γ, θ, κ) and domestic trade costs τ_{jj} unchanged. Letting 'hats' denote gross changes in variables (i.e. $\hat{x} = (x + dx)/x$), the change in welfare following any change to the matrix of trade barriers equals

$$\hat{W}_j = \left(\hat{\pi}_{\boldsymbol{j}j}\right)^{-1/(\theta\gamma)}. \tag{13}$$

This formula is particularly useful when considering a prohibitive increase in trade barriers. Because under autarky $\pi_{\boldsymbol{j}j} = 1$, we can conclude that the (percentage) real income losses of going to autarky, relative to an initial trade equilibrium, are given by $\hat{G}_j = 1 - (\pi_{\boldsymbol{j}j})^{1/(\theta\gamma)}$. This formula is analogous to the one that applies in the Eaton and Kortum (2002) framework (and the wider class of models studied by Arkolakis et al., 2012). An important difference, however, is that $\pi_{\boldsymbol{j}j}$ is *not* the aggregate share of spending on domestic intermediate or final goods (which we denote Π_{jj}^X and Π_{jj}^F, respectively, and which are readily available in input-output datasets), but rather the share of spending on goods that are produced *entirely* through domestic supply chains. That is, unlike Π_{jj}^X and Π_{jj}^F, $\pi_{\boldsymbol{j}j}$ cannot be directly observed in the data, and as a result, the sufficient statistic approach advocated by Arkolakis et al. (2012) is not feasible in our setting. Instead, one needs a model to structurally back out $\pi_{\boldsymbol{j}j}$ from available data. For a similar reason, the hat algebra approach to counterfactual analysis proposed by Dekle et al. (2008) is not feasible in our setting either.

Importantly, the share $\pi_{\boldsymbol{j}j}$ is necessarily lower than Π_{jj}^F (and increasingly so, the larger number of stages), and thus the gains from trade emanating from our model are expected to be larger on this account. This result is similar to the one derived by Melitz and Redding (2014) in an Armington framework with sequential production, and also bears some resemblance to Ossa's (2015) argument that the gains from trade can be significantly larger in a multi-sector models, with stages in our model playing the role of sectors in his framework. We will elaborate on this comparison below. Before doing so, however, it should be noted that the values of γ and θ that are appropriate for our model might be different from those appropriate for a model without multi-stage production.

First, remember that our model features an additional type of intermediate input flows relative to a model with roundabout production. In order to match the empirical ratio of value added to gross output in each country, our model will thus require setting relatively higher values of γ, which other things equal, will lead to lower gains from trade. As for the parameter θ governing the elasticity of trade flows to iceberg trade costs, we can no longer invoke estimates from standard gravity equation specifications to back out that parameter, an issue we will return to in section 6. Overall, whether our model generates larger or smaller gains from trade than models without multi-stage production is an empirical question, and one which we will explore in section 7.

Having said this, we next briefly study two special parameterizations of our model that might help better understand the implications of sequential production for the gains from trade. In our first exercise, we focus on the impact of heterogeneity in comparative advantage across stages of the supply chain. Specifically, suppose that the state of technology T_j of countries is now stage-specific and denoted by T_j^n, and countries only vary along this dimension. Although our baseline model does not feature such heterogeneity, it is evident that equation (13) would continue to apply in such a case. Our first result (proved in Appendix A.1.3) is then:

Proposition 3 *If $\alpha_n \beta_n = 1/N$ for all n, $\frac{1}{N} \sum_{n \in \mathcal{N}} \left(T_j^n\right)^{1/N} = \frac{1}{J} \sum_{j \in \mathcal{J}} \left(T_j^n\right)^{1/N} = \overline{T}$ for all j and n, and all other parameters are symmetric across countries, then starting from free trade, the losses of reverting to autarky ($\hat{G}_j = 1 - \hat{W}_j$) are given by*

$$\hat{G}_j = 1 - \left(\frac{1}{J} \times \frac{\text{GeometricMean}_n \left[\left(T_j^n\right)^{1/N} \right]}{\text{ArithmeticMean}_n \left[\left(T_j^n\right)^{1/N} \right]} \right)^{N/\gamma\theta}.$$

This proposition illustrates that higher dispersion in productivity across stages of production leads to higher gains from trade because countries gain more from specializing along specific segments of the value chain. To see this, note that a mean preserving spread of $\left(T_j^n\right)^{1/N}$ leaves the arithmetic mean of these parameters constant (by construction) but decreases their geometric mean. While this result is not entirely novel − similar insights are found in multi-sector models with cross-industry heterogeneity in productivity (Ossa 2015) − this precise characterization in terms of arithmetic and geometric means is new, and it showcases how international trade enables countries to leverage their comparative advantage in specific segments of the value chain *within* a given industry.

In a our second parameterization, we focus on the interplay between geography and production fragmentation. In particular, we find (see Appendix A.1.4 for a proof):

Proposition 4 *If all countries are symmetric in all respects, and $\tau_{ij} = \tau$ for $i \neq j$ and $\tau_{ij} = 1$ for $i = j$, then the losses of reverting to autarky are given by*

$$\hat{G}_j = 1 - \left(\prod_{n=1}^{N} \left(1 + (J-1)\, \tau^{-\theta\beta_n} \right) \right)^{-1/\gamma\theta}.$$

This result is useful for showing that our model echoes Melitz and Redding (2014)'s result that the gains from trade may become unboundedly large as production is sliced into more and more stages of production, i.e., $\lim_{N\to\infty} \hat{G}_j = 1$. There are, however, two important differences relative to the result in Melitz and Redding (2014). First, while they assume that all value is added at the most upstream stage and that each subsequent production stage entails only re-shipping the upstream input to the next stage without adding further value, our model lets us generalize this insight to richer production settings, with arbitrary value-added shares along the value chain (as long as $\alpha_n < 1$ for all $n > 1$). Second, under the assumptions in Melitz and Redding (2014), the value-added to gross output ratio is given by GDP/GO $= 1/N$ and thus goes to zero as $N \to \infty$. In our model this need not be the case. For example, let $\Psi \in (0,1)$ be some target GDP to gross output ratio and define $\alpha_n = (n-1)/N$ for $n > 1$, so $\beta_n = (N-n)!/N^{N-n}$. Because $\sum_{n=1}^{N} \beta_n \in [1,2)$ and $\lim_{N\to\infty} \sum_{n=1}^{N} \beta_n = 1$, we can just set $\gamma = \Psi$ and this ensures that $\lim_{N\to\infty}$ GDP/GO $= \Psi$. Crucially, one can verify that $\lim_{N\to\infty} \hat{G}_j = 1$. In plain words, the fragmentation of the value chain into arbitrarily many stages of production can drive the gains from trade to infinity without necessarily pushing the GDP/GO ratio to zero.

4.4 The Centrality-Downstreamness Nexus

We now exploit the tractability of our framework to explore the role of a country's geography (and, in particular, its centrality) in shaping its average position in GVCs. In order to formalize a centrality-downstreamness nexus, let us define the average upstreamness of production of a given country i as

$$U(i) = \sum_{n=1}^{N} (N-n+1) \times \frac{\sum_{j\in\mathcal{J}} \alpha_n \beta_n \times \Pr\left(\Lambda_i^n, j\right) \times w_j L_j}{w_i L_i}. \tag{14}$$

Recall that the term inside the summation in the numerator – $\alpha_n \beta_n \times \Pr\left(\Lambda_i^n, j\right) \times w_j L_j$ – represents how much of country j's total consumption of finished varieties, both by firms and final consumers, is value-added produced directly in country i at stage n in GVCs. The summation across all countries j thus represents the worldwide amount of i's value-added produced at stage n, while the denominator is country i's GDP. Thus, the index $U(i)$ in (14) delivers the average stage at which country i produces value, with the general equilibrium equation (12) implying that the weights add up to one. This is thus a closely related measure to the index of upstreamness proposed by Antràs et al. (2012).

Centrality is defined similarly as the average distance to all countries weighted by each country's aggregate consumption of finished varieties

$$C(i) = \sum_{j\in\mathcal{J}} \tau_{ij} \times \frac{w_j L_j}{\sum_{k\in\mathcal{J}} w_k L_k}. \tag{15}$$

We seek to establish a connection between the measures of upstreamness $U(i)$ and centrality $C(i)$. As in section 2, the structure of equation (8) already hints at a negative association between the two, since high values of trade costs (high τ_{ij}) in relatively downstream stages (high n) have a disproportionately negative effect on the likelihood of a given permutation of countries forming an equilibrium value chain. We begin by showing that the centrality-downstreamness nexus holds perfectly under specific, restrictive assumptions, and then show that it remains a strong force in more general scenarios.

In order to develop a precise formulation of this result, let us assume that the ease of trade between any two countries i and j can be decomposed as $\tau_{ij} = \rho_i \rho_j$, where we take $\rho_i > 1$ to be an index negatively related to country i's *centrality.* Notice that if country i is more central than country j, i.e. $\rho_i < \rho_j$, then it is cheaper to ship from i to any other country in the world than it is to ship from country j. Under these strong conditions ρ_i entirely summarizes a country's centrality, that is if $\rho_i < \rho_j$ then $C(i) < C(j)$. This is a rather strong notion of centrality but it has the virtue of providing the following stark result (which we prove in Appendix A.1.5):

Proposition 5 *If trade costs are log-separable and $\alpha_n \beta_n = 1/N$ for all n then the more central a country is the lower is the average upstreamness of this country in global value chain production. In other words, $\rho_i < \rho_j$ implies that $U(i) < U(j)$.*

The condition in the Proposition $\alpha_n \beta_n = 1/N$ serves to isolate the role of geography in shaping GVC positioning. Without this assumption, technology would not be symmetric in the value added originated at different stages, and thus the state of technology T_i of a country would affect different stages differentially, thereby generating *technological* comparative advantage.[18]

The assumption $\alpha_n \beta_n = 1/N$ implies, however, that $\alpha_n = 1/n$, and thus the value-added intensity of production decreases along the chain. This condition might then appear not to be entirely innocuous for our result. More specifically, because more central countries tend to feature higher wages in general equilibrium, it would appear that if value-added intensity instead rose sufficiently fast along the value chain, the link between downstreamness and centrality might be broken. Nevertheless, the perfect correlation between centrality and downstreamness continues to apply for *any* path of α_n, as long as trade costs are log-separable and countries are symmetric with $T_j = T$ and $L_j = L$. While we have been unable to show this result formally, we have run millions of simulations that strongly indicate that this statement is true.[19]

Of course, in the real world, trade costs are not log-separable and countries are not symmetric. With that in mind, we now explore via simulations whether the centrality-downstreamness nexus remains an important force even in more general scenarios. Specifically, conditional on a number of countries J and stages of production N we take random draws for all of the model's parameters, including trade costs which are only restricted to be symmetric ($\tau_{ij} = \tau_{ji}$) and zero domestically ($\tau_{ii} = 1$), and compute the Spearman rank correlation between centrality $C(i)$ and average up-

[18]For instance, if downstream stages contributed more to overall value added than upstream stages, we would obtain a prediction analogous to that in Costinot et al. (2013), namely that countries with better technologies T_i have comparative advantage in downstream stages. There is an obvious connection between our centrality-downstreamness result and that key result in Costinot et al. (2013). In both cases, if a country characteristic translates into a share of output being 'wasted', this proves to be more costly downstream than upstream. Conceptually, however, that there is an important distinction stemming from the fact that centrality affects the ability of countries to produce for export but also to import for consumption. More generally, beyond the log-separable case $\tau_{ij} = \rho_i \rho_j$, centrality is not just a country-characteristic, but it depends in a complex manner on country-pair trade costs, as further explored below.

[19]At the end of the proof of Proposition 5, we show that a sufficient condition for the result to hold true is that $c_i^{-\theta} \rho_i$ be increasing in centrality ρ_i. Our simulations indeed show that, regardless of the value of θ, this condition is met for *all* possible paths of α_n whenever countries are symmetric in all respects except for their centrality.

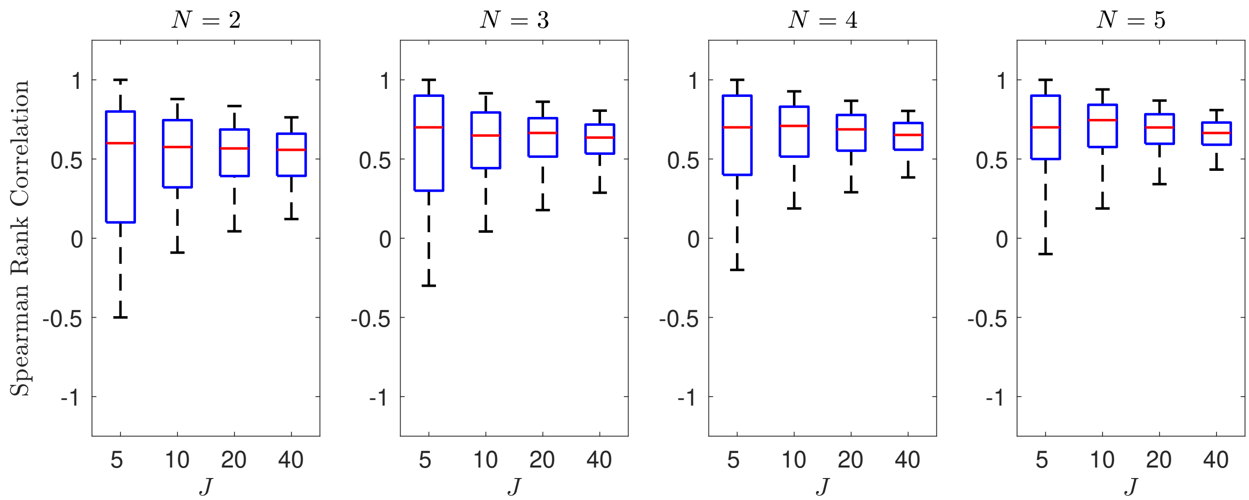

Figure 1: Centrality Downstreamness Simulations.

streamness $U(i)$.[20] That is, in each simulation we compute the equilibrium and rank countries from 1 to J according to both centrality and average upstreamness and compute the correlation between these two rankings. Under the strong conditions of Proposition 5, this correlation is perfect and the correlation coefficient equals 1. In contrast, if the downstreamness ranking were the reverse of the centrality ranking the correlation coefficient would equal -1.

Figure 1 presents the distribution of correlation coefficients across 10,000 simulations of the model for various combinations of J and N. In order to keep results tidy, we summarize the distributions using box plots in which the middle line corresponds to the median, the box to the range 25-75% of values, and the whiskers to the range 5-95% of values. As can be seen, the rank correlation is positive in the overwhelming majority of cases, indicating that more central countries tend to be more downstream. Furthermore, the association is quite strong with the median simulation featuring an average correlation well above 0.5. Finally, the centrality-downstreamness nexus becomes stronger when focusing on longer sequences of production (higher N) and the distribution of values tends to become tighter when including more countries (higher J). We will shortly provide suggestive empirical evidence consistent with this prediction.

4.5 Suggestive Evidence

We complete this section by exploring the empirical relevance of two of our model's key ingredients, namely the fact that the trade elasticity is larger for downstream than for upstream stages, and the centrality-downstreamness nexus discussed above. These empirical tests are reduced-form in nature and not structurally related to our model, but we nonetheless deem them to be informative.

We begin by studying empirically the compounding effect of trade costs. A crude way to assess the differential sensitivity of trade flows to trade costs at different stages of the value chain is to

[20] We sample $\alpha_n\beta_n$ from a uniform distribution while normalizing $\sum_{n=1}^{N}\alpha_n\beta_n = 1$, T_j and L_j from lognormal distributions, and sample $\tau_{ij} = (1 + u_{ij})(1 + u_{ji})$ where u_{ij} is a uniform random variable. The variances of these distributions are chosen so that the standard deviations of all variables are roughly comparable.

compare the elasticity of intermediate input and final good flows to various proxies for trade costs τ_{ij}. In particular, and building on the gravity equation literature, consider projecting the bilateral trade cost parameters τ_{ij} on a vector of pair-specific variables including distance, contiguity and a common language indicator. More specifically, let

$$\ln \tau_{ij} = \ln \kappa + \delta_{dist} \ln Distance_{ij} + \delta_{con} Contiguity_{ij} + \delta_{lang} SameLanguage_{ij}.$$

As long as the coefficients δ_{dist}, δ_{con} and δ_{lang} are common for intermediate inputs and final goods, then any difference in the sensitivity of final good versus intermediate input trade flows to these bilateral gravity variables indicates a differential sensitivity of 'upstream' versus 'downstream' trade flows to trade costs. To assess the plausibility of this approach, consider the case of the distance elasticity δ_{dist}. Our key identification assumption in this case is that trade costs, as a percentage of the value of the good being shipped, are identical regardless of whether the good is an input or a final good. If we observe final good trade being more sensitive to distance than input trade, we will then conclude that final good trade is more sensitive to trade costs than input trade is.

We implement this test in Table I using bilateral trade flows. Columns (1) and (2) report the results of a standard gravity specification in which the log of *aggregate* shipments from country i to country j are run on exporter and importer fixed effects, as well as the log of distance between i and j, and dummy variables for whether i and j share a contiguous, share a common border, or are the same country (i.e., domestic trade). Our shipments data are from 2011 and correspond to the Eora MRIO database. The data cover 190 countries and include information on domestic shipments (i.e., sales from i to i). The gravity variables are from the CEPII dataset for the year 2006 (the most recent one available), and the merge between these two data sources leaves us with information on 180 countries (and thus $180 \times 180 = 32,400$ observations).[21]

Our results in columns (1) and (2) are fairly standard. Distance reduces trade flows with an elasticity of around -1, while contiguity, common language, and domestic trade, have a sizable positive effect on bilateral flows. Starting in column (3), we exploit a key advantage of the Eora MRIO database, namely the fact that it reports separately bilateral shipments of intermediate inputs (X_{ij}) and of finished goods (F_{ij}). In columns (3) and (4), we re-run the specifications in columns (1) and (2) but restricting the analysis to final-good flows, while in columns (5) and (6) we do the same focusing on intermediate input flows. As is apparent from the table, the elasticity of trade flows to distance is significantly larger for final good trade (-1.229) than for intermediate input trade (-1.058). The difference is sizeable, highly statistically significant, and robust to controlling for alternative determinants of trade flows. Furthermore, comparing columns (4) and (6), we see that the positive effect of contiguity, common language, and within-border trade on trade flows is attenuated when focusing on the intermediate input component of trade, with the effects being highly statistically significant for the case of common language and domestic trade.

[21]To proxy for distance between i and j, we use the variable "dist" from the CEPII database. For $i \neq j$, this corresponds to the geodesic distance between the most populated cities in these countries. For internal distance ($i = j$), the CEPII database follows the bulk of literature in constructing an area-based measure as follows $dist_{ii} = 0.67\sqrt{area/\pi}$ (see Mayer and Zignago, 2006).

Table I: Trade Cost Elasticities for Final Goods and Intermediate Inputs

	Total Flows		Final Good Flows		Input Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-1.111***	-0.718***	-1.229***	-0.810***	-1.058***	-0.681***
	(0.019)	(0.013)	(0.021)	(0.015)	(0.018)	(0.013)
Contiguity		1.149***		1.161***		1.154***
		(0.090)		(0.098)		(0.088)
Language		0.402***		0.499***		0.358***
		(0.024)		(0.028)		(0.023)
Domestic		5.320***		5.610***		5.061***
		(0.168)		(0.191)		(0.159)
Observations	32,400	32,400	32,400	32,400	32,400	32,400
R^2	0.980	0.983	0.966	0.970	0.980	0.983

Notes: Standard errors are clustered at the country-pair level. ***, **, and * denote 1, 5, and 10 percent significance levels. All regressions include exporter and importer fixed effects.

Taken together, the results in Table I are highly suggestive of trade barriers impeding trade more severely in downstream stages than in upstream stages.[22] In Online Appendix B.3, we further show that our results are not materially affected when pooling data from all years (1995-2013) for which the Eora dataset is available (instead of just using 2011 data).[23] We further show that we obtain very similar results when restricting the analysis to manufacturing trade flows. We also repeat our tests using data from the two releases of the WIOD database, which cover a smaller and more homogeneous set of countries. The results with the 2013 release of the WIOD continue to indicate a significantly lower distance elasticity and lower 'home bias' in intermediate-input relative to final-good trade. Nevertheless, with the 2016 release of the same dataset, we only find support for the second differential effect (see Online Appendix B.3 for details). This last result makes us interpret our results with caution. Another important caveat with the evidence above is that it is based on gravity-style specifications that, strictly speaking, are inconsistent with our theoretical framework. Specifically, in section 5.3 we will show that bilateral trade flows will typically be affected by trade costs associated with third countries (see Morales et al., 2014, and Adao et al,. 2017, for recent evidence of these third-market effects).

We next turn to examining the empirical relevance of the downstreamness-centrality nexus formalized in Proposition 5. For that purpose, we build on Antràs et al. (2012) who propose a measure of the positioning of countries in GVCs and study how this measure correlates with various country-level variables. More specifically, Antràs et al. (2012) propose a measure of industry "upstreamness" (or average distance of an industry's output from final use) similar to that in equation (14), and then compute the average upstreamness of a country's export vector using trade flow data from the BACI

[22] Our results are consistent with the findings of Baldwin and Taglioni (2012), who estimate an elasticity of trade flows to trade costs – as implied by the ratio of cif to fob trade flows – that is almost twice as large for final goods than for intermediate. Using a significantly different specification, Bergstrand and Egger (2010) estimate fairly similar distance elasticities for final goods and for inputs.

[23] In fact, we obtain extremely stable results when running these regressions year-by-year for this same period.

Table II: Export Upstreamness and Centrality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Centrality (GDP weighted)			-0.173^{***}		-0.233^{***}		-0.155^{***}	
			(0.065)		(0.061)		(0.044)	
Centrality (pop. weighted)				-0.228^{***}		-0.239^{***}		-0.238^{***}
				(0.084)		(0.081)		(0.081)
Log(Y/L)	0.083	0.082	0.102	0.046	0.083^{*}	0.006		
	(0.142)	(0.142)	(0.138)	(0.148)	(0.046)	(0.037)		
Rule of Law	−0.029	−0.026	0.010	0.010				
	(0.103)	(0.104)	(0.105)	(0.105)				
Credit/Y	-0.437^{***}	-0.440^{***}	-0.375^{***}	-0.407^{***}				
	(0.136)	(0.137)	(0.130)	(0.135)				
Log(K/L)	0.156	0.159	0.163	0.188				
	(0.131)	(0.132)	(0.129)	(0.132)				
Schooling	-0.085^{***}	-0.085^{***}	-0.083^{***}	-0.094^{***}				
	(0.031)	(0.031)	(0.030)	(0.029)				
Observations	120	118	118	118	118	118	118	118
R^2	0.154	0.153	0.194	0.199	0.083	0.056	0.054	0.054

Notes: Robust standard errors reported. ***, **, and * denote 1, 5 and 10 percent significance levels.

dataset for the year 2002. Column (1) of Table II reproduces exactly their baseline specification, which includes 120 countries, and correlates a country's upstreamness with its GDP per capita, rule of law, financial development, capital-labor ratio and human capital (schooling).[24] Only financial development and schooling have a statistically significant partial correlation with upstreamness.

In order to assess the relationship between upstreamness and centrality, we simply add a measure of centrality to the core specification in column (1). In particular, for each country j we compute $Centrality_j^{GDP} = \sum_i (GDP_i/Distance_{ji})$ and $Centrality_j^{pop} = \sum_i (Population_i/Distance_{ji})$, which capture a country's proximity to other countries with either large GDP or large population (or both). We are able to compute these measures for only 118 of the original 120 countries in Antràs et al. (2012), so for completeness, column (2) reproduces the results of running the same specification as in column (1) with only those 118 countries. Clearly, the results are not materially affected. More interestingly, in columns (3) and (4) we document a highly statistically significant negative relationship between upstreamness and each of the two measures of centrality. This partial correlation is not driven by the presence of the other covariates: columns (5) and (6) show that it persists when only controlling for GDP per capita, and columns (7) and (8) demonstrate that it holds even unconditionally. In Online Appendix B.3, we plot this relationship and show that it is not driven by any outliers. Though these correlations cannot be interpreted causally, they are again suggestive of the empirical relevance of the nexus between centrality and downstreamness highlighted in Proposition 5.

[24] The source for each of these variables is discussed in Antràs et al. (2012).

5 Extending the Model and Mapping it to Input-Output Data

While our discussion so far has been centered on a stylized model of sequential production, we now demonstrate the flexibility and applicability of our tools by extending our framework in ways that incorporate other key features present in standard international trade models. In fact, we show that this extended model nests many of the literature's existing Ricardian models. At the end of this section, we also show how to map our model to the observable datapoints contained in input-output tables.

5.1 Extending the Model

We generalize our model along three main dimensions. First, we assume there are K industries indexed by $k \in \mathcal{K}$. Second, we assume that the production function depends not only on the country, industry, and stage of production but also on the use of output. For example, we assume that the production of electronics may vary depending on whether these will be used by firms in the form of intermediate inputs or by consumers in the form of final goods. Further, we allow the production of intermediate inputs to vary depending on the purchasing industry − so, for example, the production of electronics sold to the textile industry might differ from the production of electronics sold to the machinery industry. As we will show below, this flexibility will be useful for comparing our sequential production model to the literature's 'state-of-the-art' roundabout production models.

With these two properties in mind, define the n-stage price of an industry k intermediate input variety z produced for industry k' through a supply chain $\boldsymbol{\ell}$ as

$$p_{\ell(n)}^{n,k,k'}(\boldsymbol{\ell}, z) = \left(a_{\ell(n)}^{n,k,k'}(z)\, c_{\ell(n)}^{k}\right)^{\alpha_n^{k,X}} \left(p_{\ell(n-1)}^{n-1,k,k'}(\boldsymbol{\ell}, z)\, \tau_{\ell(n-1)\ell(n)}^{k}\right)^{1-\alpha_n^{k,X}}.$$

Producing an industry k good thus requires N stages of sequential production with $1-\alpha_n^{k,X}$ denoting the expenditure share on upstream sequential inputs at stage n (with $1-\alpha_1^{k,X}=0$, $\beta_{n-1}^{k,X} = \left(1-\alpha_n^{k,X}\right)\beta_n^{k,X}$, and the superscript X denoting that these shares correspond to the sequential production of intermediate inputs).[25] Crucially, note that we assume that the production function depends on the use of output (i.e., the industry k' that will purchase this intermediate input) only through the productivity shifters $a_j^{n,k,k'}(z)$ for all n. Following the lead-firm approach, we assume that the overall productivity of chain $\boldsymbol{\ell}$ in industry k in the production of intermediate inputs sold to k' is characterized as

$$\Pr\left(\prod_{n=1}^{N}\left(a_{\ell(n)}^{n,k,k'}(z)\right)^{\alpha_n^{k,X}\beta_n^{k,X}} \geq a\right) = \exp\left\{-a^{\theta^k}\prod_{n=1}^{N}\left(T_{\ell(n)}^{k,k'}\right)^{\alpha_n^{k,X}\beta_n^{k,X}}\right\}. \tag{16}$$

Chain-level productivity is thus governed by the industry-specific comparative advantage parameter θ^k and the producing-industry- and buying-industry-specific absolute advantage parameter $T_j^{k,k'}$. To

[25] Note that an industry k can effectively require only $N^k \leq N$ stages of sequential production if $1-\alpha_n^{k,X}=0$ for $n = 1+\left(N-N^k\right)$. Hence, instead of making notation heavier and assuming that N^k varies at the industry-level, we use a single common N and assume that the effective number of required stages of sequential production are determined by the expenditure shares $\alpha_n^{k,X}$.

aid the reader's intuition, Appendix Section A.4 depicts this production structure graphically.

Final good production is analogous with $p_{\ell(n)}^{n,k,F}(\boldsymbol{\ell},z)$ denoting the n-stage price of an industry k final good variety z produced through supply chain $\boldsymbol{\ell}$. As in the case for intermediate inputs, producing each final good variety requires N stages of production governed by the stage expenditure shares $\alpha_n^{k,F}$ and the country-industry-stage-specific productivity shifters $a_{\ell(n)}^{n,k,F}(z)$ shaped by the absolute technology parameters $T_j^{k,F}$ in a way analogous to (16).

Finally, as a third extension of our framework, we introduce input-output linkages. Specifically, we assume that the unit cost of production in industry k' equals

$$c_j^{k'} = (w_j)^{\gamma_j^{k'}} \prod_{k\in\mathcal{K}} \left(P_j^{k,k'}\right)^{\gamma_j^{k,k'}}, \tag{17}$$

where $\gamma_j^{k'}$ denotes the value-added share and $\gamma_j^{k,k'}$ the expenditure on industry k finished intermediate inputs (with $\gamma_j^{k'} + \sum_{k\in\mathcal{K}} \gamma_j^{k,k'} = 1$ for all k'). This unit cost depends on wages w_j and on the price indices $P_j^{k,k'}$ associated with the unit cost of a CES bundle of industry k varieties purchased by industry k' in country j.[26] Analogously, we denote by $P_j^{k,F}$ the price index of a CES bundle of industry k varieties purchased by final consumers in country j. As is standard, we also assume that consumer preferences are Cobb-Douglas across industries with ζ_j^k denoting the expenditure share on industry k final goods, and $P_j^F = \prod_{k\in\mathcal{K}} \left(P_j^{k,F}\right)^{\zeta_j^k}$ the consumer price index.[27]

Invoking the famous terminology developed by Baldwin and Venables (2013), these three extensions turn our benchmark model of 'snakes' (i.e., of purely sequential value chains), into a richer production structure that incorporates 'spider'-like features, with each production stage sourcing parts and components from several industries. Note, however, that the 'spider legs' in our model are composite bundles of snakes, rather than individual, customized snakes.

The multi-industry equilibrium can be characterized using derivations similar to the single-industry case. Following the lead-firm approach, the share of industry k intermediate input varieties sourced through $\boldsymbol{\ell}$ by industry k' in country j equals

$$\pi_{\boldsymbol{\ell}j}^{k,k'} = \frac{\prod_{n=1}^{N-1} \left(\left(T_{\ell(n)}^{k,k'}\right)^{\alpha_n^{k,X}} \left(\left(c_{\ell(n)}^{k}\right)^{\alpha_n^{k,X}} \tau_{\ell(n)\ell(n+1)}^{k} \right)^{-\theta^k} \right)^{\beta_n^{k,X}} \times \left(T_{\ell(N)}^{k,k'}\right)^{\alpha_N^{k,X}} \left(\left(c_{\ell(N)}^{k}\right)^{\alpha_N^{k,X}} \tau_{\ell(N)j}^{k} \right)^{-\theta^k}}{\Theta_j^{k,k'}}.$$

The share of industry k final good varieties sourced through $\boldsymbol{\ell}$ by consumers in j is defined analogously and denoted by $\pi_{\boldsymbol{\ell}j}^{k,F}$.

[26] Note that industry k' in j buys each variety z from the supply chain $\boldsymbol{\ell}$ that minimizes $p_{\ell(N)}^{N,k,k'}(\boldsymbol{\ell},z)\,\tau_{\ell(N)j}^{k}$.

[27] While our model is fairly general, it is straightforward to generalize even further and let the production parameters vary across stages and depending on the use of output by defining $\gamma_j^{n,k,k'}$, $\alpha_n^{k,k',X}$, $T_j^{n,k,k'}$, $T_j^{n,k,F}$.

5.2 Relationship to Previous Ricardian Trade Models

Let us now describe the relationship between our extended sequential production model and other Ricardian models of trade through the lens of the the 'gains-from-trade' formula. Analogous to (13), the change in real income following any change to trade barriers in the extended model can be written in terms of a set of changes in the domestic expenditure shares as

$$\hat{W}_j = \prod_{k\in\mathcal{K}} \left(\hat{\pi}^{k,F}_{\boldsymbol{j}j}\right)^{-\frac{1}{\theta^k}\zeta^k_j} \times \prod_{k\in\mathcal{K}} \prod_{k'\in\mathcal{K}} \prod_{k''\in\mathcal{K}} \left(\hat{\pi}^{k,k'}_{\boldsymbol{j}j}\right)^{-\frac{1}{\theta^k}\gamma^{k,k'}_j \delta^{k',k''}_j \zeta^{k''}_j} . \tag{18}$$

where, as before, a 'hat' indicates the (gross) change in a variable and where $\boldsymbol{j}$ is a purely domestic chain of size N. Additionally, the auxiliary parameters $\delta^{k,k'}_j$ equal the gross value of industry k inputs used in the production of one dollar of industry k' goods across all stages of production.[28]

The intuition for this formula is straightforward and depends on three forces. First, the gains from trade depend on the expenditure shares on domestic supply chains since these are an inverse measure of how much a country exploits comparative advantage through international trade. Second, the powers on the expenditure shares indicate the importance of each supply chain for domestic consumption in autarky. That is, relative to one dollar of aggregate final consumption, ζ^k_j indicates the gross value of industry k finished final goods that need to be produced while $\sum_{k',k''\in\mathcal{K}} \gamma^{k,k'}_j \delta^{k',k''}_j \zeta^{k''}_j$ captures the gross value of industry k finished intermediate inputs that need to be produced for industry k' across all stages of the supply chain (including all input-output linkages) to satisfy that dollar of final consumption. Third, the trade elasticities θ^k regulate the sensitivity of the losses from increases in trade barriers due to the interaction between how much a country is already producing along a specific supply chain in the trade equilibrium (the domestic shares) and the importance of a given supply chain for final consumption (the gross value weights).

This extension nests many of the previously developed Ricardian models (this can be seen graphically in the flow charts in Appendix Section A.4). On the one hand, this model nests the single industry sequential production model developed above when $K = 1$ and when restricting production technology T_j and α_n to be the same for intermediate inputs and final goods, in which case equation (18) reduces to (13). On the other hand, our model nests 'state-of-the-art' Ricardian models featuring roundabout production. Specifically, letting $N = 1$, the gains from trade are given exactly by the formula in (18) but depend on the aggregate domestic expenditure on industry k final goods given by $\Pi^{k,F}_{jj}$ and the aggregate domestic expenditure on industry k intermediate inputs by each industry k' given by $\Pi^{k,k'}_{jj}$, as in the recent work by Alexander (2017). Further restricting the model to $T^{k,k'}_j = T^{k,F}_j = T^k_j$ for all k' delivers exactly the influential roundabout model of Caliendo and Parro (2015), which includes multiple industries and input-output linkages but which features a single aggregate domestic expenditure share Π^k_{jj} for each industry. In addition, restricting to the single industry case and a common value-added share across countries delivers the classic Eaton and Kortum (2002) model. In sum, the formula in (18) illustrates how the sequential production,

[28]Formally, $\delta^{k,k'}_j$ is an element of $\boldsymbol{\delta}_j = \left[\mathbb{I} - \boldsymbol{\gamma}_j\right]^{-1}$ where $\boldsymbol{\gamma}_j$ is the $K \times K$ matrix of elements $\gamma^{k,k'}_j$.

		Input use & value added			Final use			Total use
		Country 1	$\cdots$	Country J	Country 1	$\cdots$	Country J	
Output supplied	Country 1	$\boldsymbol{X}_{11}$	$\cdots$	$\boldsymbol{X}_{1J}$	$\boldsymbol{F}_{11}$	$\cdots$	$\boldsymbol{F}_{1J}$	$\boldsymbol{GO}_1$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	Country J	$\boldsymbol{X}_{J1}$	$\cdots$	$\boldsymbol{X}_{JJ}$	$\boldsymbol{F}_{J1}$	$\cdots$	$\boldsymbol{F}_{JJ}$	$\boldsymbol{GO}_J$
Value added		$\boldsymbol{GDP}_1$	$\cdots$	$\boldsymbol{GDP}_J$				
Gross output		$\boldsymbol{GO}_1$	$\cdots$	$\boldsymbol{GO}_J$				

Figure 2: A Schematic World Input-Output Table.

multi-industry, and input-output linkages elements interact and shape the gains from trade.[29]

5.3 Mapping the Model to Observables

So far, we have described how to extend standard Ricardian models to a multi-stage environment that delivers sharp predictions for the 'trade shares' $\pi_{\boldsymbol{\ell}j}^{k,k'}$ and $\pi_{\boldsymbol{\ell}j}^{k,F}$ in terms of the relative prevalence of specific production paths (or GVCs) rather than in terms of gross trade volumes. Unfortunately, these 'GVC trade shares' are not directly observable in the data, so we next describe how to map the model to the type of information available in World Input-Output Tables (WIOTs)

First, let us describe the datapoints contained in WIOTs while aiding the exposition with the schematic representation in Figure 2. This matrix is split into two blocks with the left one containing data on bilateral intermediate input trade flows across countries and the right one containing the data for final good trade flows. Specifically, each $\boldsymbol{X}_{ij} = \left[X_{ij}^{kk'}\right]$ is the $K \times K$ matrix of bilateral intermediate input flows from industry k in country i sold to industry k' in country j. Analogously, $\boldsymbol{F}_{ij} = \left[F_{ij}^{k}\right]$ is the $K \times 1$ vector of bilateral final good sales of industry k in country i to consumers in country j. Hence, rows across the WIOT represent the sales of a country-industry to other country-industries and to consumers with the sum across a row equaling a country-industry's gross output. Meanwhile, columns in the left block represent the purchases of intermediate inputs by country-industries (with the sum across a column equaling gross output minus value added) and columns in the right block represent the final good purchases of final good consumers in each country (with the sum across a column equaling aggregate final good consumption).[30]

Second, consider the implications of our model for final good consumption. Notice that for final goods to flow from a given source country i to consumers in a given destination country j, it must be the case that country i is in position N in a chain producing final goods for country j. Remember from section 4.2 that Λ_i^n represents the set of GVCs flowing through i at position n so that Λ_i^N corresponds to the set of chains in which final assembly is carried out in i. With this notation, the

[29] Arkolakis et al. (2012) show that analogous formulas apply under Armington frameworks with roundabout production. This is also the case in our family of models with sequential production. For example, de Gortari (2019) develops a supply-chain Armington model which delivers a gains from trade formula identical to (18) when imposing $N = 2$.

[30] Note that the difference between aggregate final consumption and value added is the trade deficit or surplus. These deficits are nontrivial for certain countries and are taken into account in both our estimation and our counterfactual exercises, as discussed below.

dollar value of final goods shipped from industry k in country i to consumers in country j equals

$$F_{ij}^{k} = \sum_{\boldsymbol{\ell} \in \Lambda_i^N} \pi_{\boldsymbol{\ell} j}^{k,F} \times \zeta_j^k w_j L_j, \tag{19}$$

where $\zeta_j^k w_j L_j$ is country j's total consumption of industry k final goods.[31]

Third, note that computing intermediate input flows between any two countries i and j is more intricate, but equally straightforward. The difficulty lies in the fact that country j potentially buys different shares of inputs from i at different stages of the supply chain, yet WIOTs only report the aggregate amount of intermediate inputs traded across all stages of the supply chain. As we now show, this challenge can be overcome by introducing a set of auxiliary variables ϑ_j^k capturing the value-added created in each country j and industry k and then solving jointly for these variables and the bilateral intermediate input flows.

To begin, note there are two broad types of traded inputs. The first type are finished N-stage intermediate inputs which are traded both within and across industries−in other words, the intermediates traded through the input-output linkages characterized by $\gamma_j^{k,k'}$. The second type are n-stage sequential inputs (with $n < N$) which are only traded within industries−and disciplined by the sequential production expenditure shares $\alpha_n^{k,X}$ and $\alpha_n^{k,F}$.

Cross-industry input flows are made entirely of inputs of the first type such that the input flow sold from industry k in country i to industry $k' \neq k$ in country j equals

$$X_{ij}^{kk'} = \sum_{\boldsymbol{\ell} \in \Lambda_i^N} \pi_{\boldsymbol{\ell} j}^{k,k'} \times \frac{\gamma_j^{k,k'}}{\gamma_j^{k'}} \vartheta_j^{k'}. \tag{20}$$

Since $\vartheta_j^{k'}$ is a value-added measure, then $\gamma_j^{k,k'}/\gamma_j^{k'}$ imputes the aggregate amount of N-stage inputs purchased from industry k used in this production and then $\sum_{\boldsymbol{\ell} \in \Lambda_i^N} \pi_{\boldsymbol{\ell} j}^{k,k'}$ imputes the share of inputs produced through supply chains in which i produces the Nth stage.

Within-industry input flows, on the other hand, equal the sum of the finished N-stage inputs plus the sequential inputs used to produce both finished N-stage intermediate inputs and finished N-stage final goods

$$X_{ij}^{kk} = \underbrace{\sum_{\boldsymbol{\ell} \in \Lambda_i^N} \pi_{\boldsymbol{\ell} j}^{k,k} \times \frac{\gamma_j^{k,k}}{\gamma_j^{k}} \vartheta_j^{k}}_{N\text{-stage intermediates}} + \sum_{n=1}^{N-1} \sum_{\boldsymbol{\ell} \in \Lambda_{i \to j}^n} \sum_{h \in \mathcal{J}} \Bigg(\underbrace{\beta_n^{k,X} \sum_{k' \in \mathcal{K}} \pi_{\boldsymbol{\ell} h}^{k,k'} \times \frac{\gamma_h^{k,k'}}{\gamma_h^{k'}} \vartheta_h^{k'}}_{n\text{-stage intermediates}} + \underbrace{\beta_n^{k,F} \pi_{\boldsymbol{\ell} h}^{k,F} \times \zeta_h^k w_h L_h}_{n\text{-stage final goods}} \Bigg). \tag{21}$$

The first term on the right-hand side is analogous to the cross-industry flow in (20) and captures the N-stage inputs traded within industry k across countries i and j. The second term is similar, but instead captures the n-stage inputs traded within industry k across countries i and j used in

[31] Notice that we are interpreting the probability function $\pi_{\boldsymbol{\ell} j}^{k,F}$ as an expenditure share. This is justified by the fact that all finished varieties in country j command the same expected price regardless of the actual chain of production, exactly as in Eaton and Kortum (2002). For the same reason, below we will also treat $\pi_{\boldsymbol{\ell} j}^{k,k'}$ as an expenditure share.

the production of N-stage intermediate input varieties consumed in all markets $h \in \mathcal{J}$. In terms of notation, remember that $\Lambda^n_{i\to j} = \left\{\boldsymbol{\ell} \in \mathcal{J}^N \,|\, \ell(n) = i \text{ and } \ell(n+1) = j\right\}$ accounts for all supply chains through which i and j produce sequentially the nth and $(n+1)$th stages of production, respectively. This term also includes a $\beta^{k,X}_n$ which accounts for the fact that for every dollar of N-stage varieties bought by some country, only $\beta^{k,X}_n$ cents are traded at stage n. Finally, the third term captures the total amount of i's output shipped to j at stage n in final good varieties that are eventually consumed as N-stage final goods across all countries $h \in \mathcal{J}$.[32]

Having defined input flows in terms of value added, the latter can in turn be defined in terms of the former as the difference between gross output and aggregate intermediate input purchases

$$\vartheta^k_i = \sum_{j\in\mathcal{J}} \left(\sum_{k'\in\mathcal{K}} X^{k,k'}_{ij} + F^k_{ij} \right) - \sum_{j\in\mathcal{J}} \left(\sum_{k'\in\mathcal{K}} X^{k',k}_{ji} \right). \tag{22}$$

Equations (20), (21), and (22) thus pin down both the bilateral intermediate input flows and value-added terms as a function of wages. Finally, the general equilibrium equation is such that aggregate value added is equal to the return on the factors of production: $w_j L_j = \sum_{k\in\mathcal{K}} \vartheta^k_j$.

In sum, and although computing these flows has proved somewhat cumbersome, a key feature of our model is that it provides explicit formulas that link the various entries of a WIOT to fundamental parameters of the model, as well as to endogenous variables determined by these same parameters.[33]

6 Estimation

We now describe how we leverage the fact that our model delivers explicit analytical counterparts to the entries of a WIOT in order to structurally estimate the model's key parameters. For computational reasons, we will mostly focus on an application in which the various industries in a WIOT are collapsed into a single sector. Nevertheless, following the lead of Alexander (2017), we will allow for asymmetries in the technology to produce final goods and inputs. It is undeniable that an estimated multi-industry version of our model could deliver estimates with significant different implications for the counterfactuals studied in section 7, so readers are entitled to treat the results in the next two sections as a proof of concept, rather than as a definitive quantitative exercise. Having said this, in section 7, we will briefly describe the empirical results of an extension of our model featuring two sectors, manufacturing and services.[34]

[32] Note the difference between finished and final goods. Finished refers to stage-N goods which are traded both as intermediates inputs used to produce further downstream goods and as final goods sold to consumers.

[33] We have demonstrated how trade flows within our model's GVCs aggregate into the bilateral input-output trade flows observed in the data. De Gortari (2019) studies the inverse question of how to use the aggregate data to disentangle the shape of the GVCs underlying these flows.

[34] We do not go beyond this two-industry model for four main reasons. First, the number of parameters to estimate increases quadratically with the number of industries. The parameters to estimate in each industry are $\left\{\alpha^{k,X}_n, \beta^{k,X}_n, T^{k,k'}_j, \alpha^{k,F}_n, \beta^{k,F}_n, T^{k,F}_j, \gamma^{k,k'}_j, \gamma^k_j\right\}$ implying $\sum_k 2\left(N^{k,X} - 1 + N^{k,F} - 1\right) + 2JK\left(K+1\right)$ parameters in total. Second, computing the model's general equilibrium requires solving for $\left\{w_j, P^{k,k'}_j, P^{k,F}_j, \vartheta^k_j\right\}$ or $J\left(1 + K\left(K+1\right)\right)$ endogenous variables, and thus becomes increasingly complicated as more industries are added. Third, our MPEC estimation procedure requires deriving and computing the model's Jacobian matrix which is itself

Before delving into the details of the estimation, we briefly describe our data source. Note that building a WIOT of the type in Figure 2 is a formidable endeavor because it requires collecting trade and production data from many different sources, including national and supra-national statistical offices, but also because it necessarily requires assumptions and data analysis in order to make the data comparable. In this paper we work, for the most part, with the World Input Output Database (or WIOD for short), the outcome of a project that was carried out by a consortium of 12 research institutes headed by the University of Groningen in the Netherlands (see Timmer et al., 2015). We choose this dataset for our estimation because we believe that the assumptions put into its construction are less heroic than those contained in other sources. The main limitation of the WIOD is that it only covers 43 relatively developed countries, and includes no African country and only two countries in Latin America (Brazil and Mexico).[35]

The data contained in WIOTs (when aggregated at the country level) is a set of aggregate bilateral intermediate input flows denoted by $\overline{X}_{ij}$ and a set of aggregate bilateral final good flows denoted by $\overline{F}_{ij}$. Our notation is such that 'bars' define observed datapoints. It will often be useful to describe the data using aggregate expenditure shares which we define as

$$\overline{\Pi}_{ij}^{X} = \frac{\overline{X}_{ij}}{\sum_{i' \in \mathcal{J}} \overline{X}_{i'j}}, \quad \text{and} \quad \overline{\Pi}_{ij}^{F} = \frac{\overline{F}_{ij}}{\sum_{i' \in \mathcal{J}} \overline{F}_{i'j}}. \tag{23}$$

In order to describe how we use the data to discipline our model in our estimation procedure, let us first outline the parameters needed to estimate or calibrate for a given number J of countries and N of stages. Geography is pinned down by the $J \times J$ matrix of iceberg trade costs τ_{ij}, which are assumed common for inputs and for final goods. Production depends, first, on the labor value-added shares γ_j, which we allow to be country-specific but common across stages. Second, production depends on the Cobb-Douglas input expenditure shares α_n^X and α_n^F, which are allowed to vary across the production of intermediate input and final good varieties, and which are stage-specific but common across countries. Third, labor productivity depends on the country-specific state-of-technology levels T_j^X and T_j^F, which are also allowed to differ across intermediate input and final good varieties (as in Alexander, 2017). The parameter θ governing the strength of comparative advantage is instead assumed common across stages and countries. Overall, production thus depends on $J + 2(N-1) + 2J$ parameters together with the N-stage trade elasticity θ. Finally, although countries are also allowed to vary in terms of their supply of equipped labor, the particular values of L_j only affect the estimates of absolute productivity and of equilibrium wages, but have no bearing

of size $\left[\sum_k 2\left(N^{k,X} - 1 + N^{k,F} - 1\right) + JK(J+3)(K+1) + J\right] \times \left[\sum_k (N^{k,X} - 1 + N^{k,F} - 1) + JK(J+1)(K+1)\right]$. Each dimension of the Jacobian matrix increases quadratically with K thus implying that the number of elements increases at the rate of K^4. These three reasons imply that increasingly more computational power is needed in each iteration of the estimation algorithm when more industries are added. Finally, a fourth reason is that the problem appears to behave increasingly non-linear when more industries are added. This non-linerarity both increases the number of iterations required for the estimation algorithm to converge and decreases the reliability of the solution being at a global minimum (as is discussed further in the Online Appendix). We thus focus on simple quantitative explorations of our model, and hope that future work will uncover quicker and more efficient computational procedures for unleashing the model's quantitative power in more complex environments.

[35] Two releases of the WIOD are available. The 2016 release contains a WIOT covering 43 countries and the rest of the world for the period 2000-2014. A previous release in 2013 contained information for 40 countries and the rest of the world, for the period 1995-2011. See http://www.wiod.org.

on the counterfactuals discussed below. With that in mind, we simply normalize equipped labor as $L_j = (capital_j)^{\frac{1}{3}}(population_j)^{\frac{2}{3}}$, with capital and population drawn from the Penn World Tables, while we normalize T_j^X and T_j^F to sum up to 100 across countries.

With regard to the length of the sequentiality of production N, we will argue that this parameter cannot be properly identified off WIOTs because bilateral flows represent highly aggregated counterparts of the rich supply chain flows underlying world trade flows (as can be seen in (19), (20) and (21)). Hence, our goal will be to use all of the variation contained in the data in order to pin down the model's parameters conditional on a given N. We will then verify that increasing N necessarily improves the model's fit, since this increases its flexibility and permits it to better match the variation contained in the data.[36] This approach will then let us compare the model's parameterizations across different N's in order to study the effects of sequential production on our counterfactual experiments.

Let us now turn to a more detailed description of our estimation approach. To pin down trade costs, we follow the method proposed by Head and Ries (2001) and make the simplifying assumption that domestic trade costs are common across countries and normalized to 0, i.e., $\tau_{jj} = 1$ for all $j \in \mathcal{J}$. International trade costs, up to a power $-\theta$, can then be immediately read off the data through the use of equation (19) and our empirical analogs in (23):

$$\tau_{ij}^{-\theta} = \sqrt{\frac{\overline{\Pi}_{ij}^F}{\overline{\Pi}_{ii}^F}\frac{\overline{\Pi}_{ji}^F}{\overline{\Pi}_{jj}^F}}. \tag{24}$$

Trade costs are symmetric by construction, i.e., $\tau_{ij} = \tau_{ji}$, and in practice the triangle inequality (i.e., $\tau_{ij} \leq \tau_{ik}\tau_{kj}$) holds across more than 99.9% of triples.

A consequence of this approach to backing out trade costs is that the calibrated values for $\tau_{ij}^{-\theta}$ are unaffected by the particular value of θ chosen. Although the value of θ affects the equilibrium of our model beyond its effect on $\tau_{ij}^{-\theta}$, it turns out that the moments we employ for our structural estimation (see below) do *not* identify θ. More precisely, for every possible value of θ, there exists a re-normalization of T_j^X and T_j^F that yields the same equilibrium (conditional on the same set of parameters γ_j, α_n^X, and α_n^F). With that in mind, we simply set $\theta = 5$ in our estimation. This value is slightly higher than is typically assumed in the literature, but our model predicts that the trade elasticity for final goods (i.e., θ) should be larger than the elasticity one would estimate with overall trade flows (which is a weighted average of $\theta\beta_1^X$, $\theta\beta_2^X$,...,θ and $\theta\beta_1^F, \theta\beta_2^F, \ldots, \theta$).[37]

To reiterate, our empirical strategy will be to fix N and find the remaining set of parameters

[36] The only case in which this does not hold is when our model represents the true data generating process−in which case increasing N beyond its true value bears no more fruits. While our model might be a good approximation to the real world, it is certainly not the real world's exact data generating process and so this case never holds empirically (nonetheless, we explore this idea further in section 6.2).

[37] Simonovska and Waugh (2014), in a widely cited study, find a range for the elasticity of trade between 2.47 and 5.51. Using U.S. import data, Antràs et al. (2017) estimate an elasticity of trade of 4.54. Finally, Caliendo and Parro (2014) find an aggregate elasticity of 4.49 using the previous release of the WIOD data. By simulating data from our model, we have indeed confirmed that the elasticity of aggregate bilateral trade flows to trade resulting from our model ranges from 4.47 for $N = 2$ to 4.51 for $N = 5$. By construction, when focusing on final-good flows, we recover a value of 5.

that best fit the data. Specifically, having pinned down $\tau_{ij}^{-\theta}$ and θ, we fix N and estimate the set of parameters γ_j, α_n^X, α_n^F, T_j^X, and T_j^F by targeting all of the WIOD's flows via the generalized method of moments, thus minimizing the difference between the observed and the model's simulated WIOT flows.[38] We estimate our model using a constrained optimization approach based on the MPEC algorithm of Su and Judd (2012) described in Appendix Section B.4. The next section will then show how quantitative counterfactual experiments vary when increasing N and thus will showcase the effects of incorporating multi-stage production. In particular, note that our estimated model with $N = 1$ corresponds to the roundabout model of Alexander (2017), so that this parameterized model will be a useful benchmark for comparing our sequential production model with $N > 1$ to roundabout production models.

Having described the estimation procedure let us now discuss the variation in the data that allows us to identify each set of parameters. In particular, Figure 3 depicts four salient characteristics of the 2014 WIOD data that we use to pin down the production function parameters (note that circle size is proportional to country j's GDP). As we shall now show, the variation in each type of chart will let us reject specific elements found in previous roundabout production models.

First, the northwest panel shows the relationship between a country's ratio of aggregate value added to gross output (VA/GO) and its share of final output relative to its gross output (F/GO). In a closed economy, these two ratios would naturally coincide and all observations would lie on the 45 degree line. In a globalized world, differences in these ratios provide a rough measure of the positioning of countries in GVCs. More specifically, for a given VA/GO, a high share of F/GO indicates that a country is relatively downstream in GVCs. Similarly, for a given F/GO, a low VA/GO ratio indicates that a country uses a relatively large amount of foreign inputs in production, which again suggests a relatively downstream position of this country in GVCs. Hence, in a world in which countries are in markedly different segments of GVCs, the ratios VA/GO and F/GO might be expected to be negatively correlated. With this background in mind, the figure indicates that although there are some deviations from the 45 degree line, cross-country variation in these ratios is much larger than within-country differences. To understand the implications of this for our estimation, envision the simplest version of our model (akin to Eaton and Kortum, 2002) with $N = 1$, $T_j^X = T_j^F$, and a common γ across countries. In this case, it is easy to see that VA/GO = F/GO = γ for all countries and so this variation rejects this particular paremeterization. Indeed, subsequent iterations of the Eaton and Kortum (2002) model, such as Caliendo and Parro (2014)'s work, address this issue by letting γ_j vary across countries. This has the virtue of making VA/GO exactly equal to γ_j and of generating a positive correlation between VA/GO and F/GO. In sum, when $N = 1$, γ_j is identified by the cross-country variation in the northwest panel of Figure 3. While analytical expressions for VA/GO and F/GO are more complex when $N > 1$, it turns out that γ_j is also well suited for

[38]Specifically, for a given N we minimize the weighted squared sum of differences between the observed and simulated WIOT flows. In order to guarantee that our model provides a proper quantitative evaluation of the general-equilibrium workings of the world economy, we place a higher weight on matching the empirical moments of larger trade flows. Hence, we weigh each moment using the square root of the targeted moment and minimize $\sum_{i,j\in\mathcal{J}} \sqrt{\overline{X}_{ij}} \left(X_{ij} - \overline{X}_{ij}\right)^2 + \sum_{i,j\in\mathcal{J}} \sqrt{\overline{F}_{ij}} \left(F_{ij} - \overline{F}_{ij}\right)^2$.

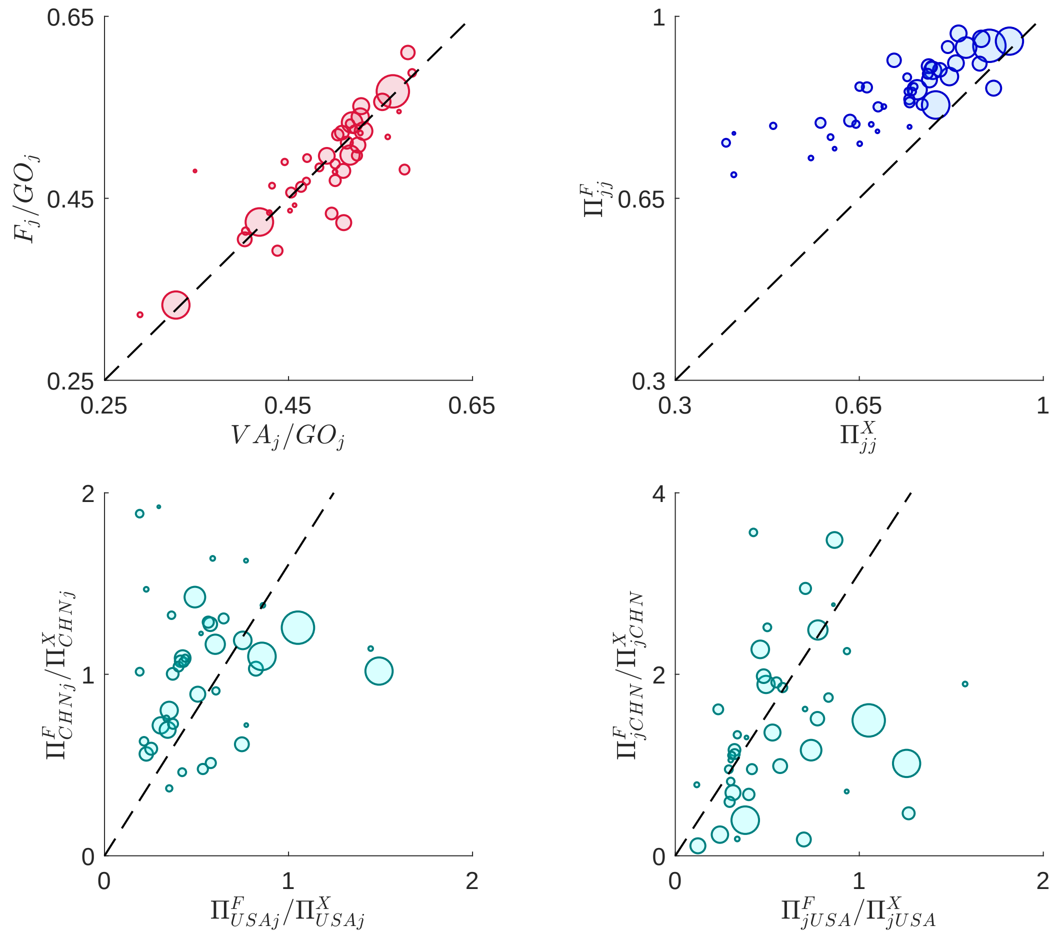

Figure 3: Some Key Features of the World Input Output Database

matching this variation in the sequential production model.[39]

Second, the northeast panel shows the relationship between a country's domestic expenditure share on intermediate inputs and final goods. As is clear from the graph, most observations lie above the 45 degree line, indicating that the observed input $\overline{\Pi}^X_{jj}$ and output $\overline{\Pi}^F_{jj}$ matrices are asymmetric, and that countries tend to rely on foreign sources more prevalently for inputs than for final goods.[40] Since standard roundabout models such as Eaton and Kortum (2002) and Caliendo and Parro (2014) assume that all output is produced in the same way regardless of its use (i.e., $\Pi^X_{jj} = \Pi^F_{jj}$), the WIOD rejects these *symmetric* parameterizations. There are at least two alternative ways in which this variation can be rationalized. The first, adopted by Alexander (2017), is that absolute productivity levels $T^X_j \neq T^F_j$ vary across the production of intermediate inputs and final goods so that $\Pi^X_{jj} \neq \Pi^F_{jj}$.

[39]To see this, imagine for simplicity $N = 2$ and that $\alpha^F_2 = \alpha^X_2 = \alpha_2$. As discussed, when $\alpha_2 \to 1$ the upstream stage of production is irrelevant for production and the VA/GO ratio equals γ_j. Conversely, when $\alpha_2 \to 0$, the downstream stage of production adds very little value, and so the VA/GO ratio is close to $\gamma_j/2$, since the same output is shipped twice but value is added essentially only once. In practice, for a general N, the VA/GO ratio features variation both because countries have different labor value-added shares but also because they find themselves at different degrees of upstreamness along the GVC; the interaction of both forces determine this statistic.

[40]This asymmetry is consistent with the difference in the regression coefficients on the domestic dummy of columns (4) and (6) in Table I.

A second explanation, is instead based on the central insight of this paper: Namely, the notion that trade costs are more detrimental for downstream versus upstream stages of production. To see this, note that $\Pi_{jj}^{X} < \Pi_{jj}^{F}$ even if $T_j^X = T_j^F$ as long as there is sequential production with $N > 1$. The reason for this is that the sequential input expenditure shares α_n^X and α_n^F determine how fast the trade elasticity increases along GVCs, and are thus crucial in shaping the observed differences between these two domestic input shares. Thus, our model incorporates both channels so that the variation in the northeast panel of Figure 3 helps identify T_j^X, T_j^F, α_n^X, and α_n^F. In addition, note that while this variation identifies the relative size of T_j^X and T_j^F within countries, the levels across countries are pinned down by the size of each country's aggregate value added and trade flows.

Third, and finally, the bottom two panels in Figure 3 depict the relation between the ratio of import shares of final goods relative to intermediate inputs from two source countries (China and the U.S., in this particular case) across all countries (southwest panel), and the relation between the ratio of export shares of final goods relative to intermediate inputs to those two destination countries (southeast panel). This is the type of variation that roundabout models cannot handle – not even the highly flexible model of Alexander (2017). The reason for this is that when $N = 1$, then

$$\frac{\Pi_{ij}^{F}}{\Pi_{ij}^{X}} = \frac{T_i^F}{T_i^X} \times \frac{\Theta_j^X}{\Theta_j^F},$$

and so the model predicts that the relative ratio of import shares from any two given countries or the relative ratio of export shares to a given pair of countries must be proportional across all importing or exporting countries. In other words, the model with $N = 1$ predicts that all observations in both of the two bottom panels depicted in Figure 3 should lie along the single ray depicted in these graphs.[41] Hence, this last variation in the data can only be fitted by the sequential production channel with, as before, the sequential expenditure input shares α_n^X and α_n^F delivering the degrees of freedom needed to pin this heterogeneity down. Crucially, increasing N delivers additional flexibility and this is why the model's fit improves as N grows.

Before turning to a discussion of our estimation results, we briefly comment on our treatment of trade imbalances. As mentioned above (see footnote 30), these imbalances are empirically nontrivial and correspond to the difference between aggregate final consumption and value added. Following a common approach in the trade literature (see, in particular, Costinot and Rodríguez-Clare, 2015), we treat these deficits as exogenous parameters, and we adjust our general-equilibrium equations to account for the difference between income and spending (see Appendix Section A.3).

6.1 Estimation Results

Table III presents our estimation results. The first five rows present the estimates of the sequential production input expenditure shares α_n^X and α_n^F and the mean value-added share γ_j across all countries for $N = 1, 2, 3, 4, 5$. We will refer to this set of estimates as asymmetric parameterizations since both the sequential input expenditure shares and absolute productivity levels depend on whether

[41] Note that the model predicts that the ray in the southwest panel is given by $\left(T_{CHN}^F / T_{CHN}^X\right) / \left(T_{USA}^F / T_{USA}^X\right)$, while the ray in the southeast panel is endogenous and given by $\left(\Theta_{CHN}^X / \Theta_{CHN}^F\right) / \left(\Theta_{USA}^X / \Theta_{USA}^F\right)$.

Table III: Estimation Results

N	Obj. Func.	α^X_{N-4}	α^X_{N-3}	α^X_{N-2}	α^X_{N-1}	α^X_N	α^F_{N-4}	α^F_{N-3}	α^F_{N-2}	α^F_{N-1}	α^F_N	Mean γ_j
Asymmetric Parameterization												
1	39.44					1.00					1.00	0.51
2	12.07				1.00	0.25				1.00	1.00	0.62
3	11.68			1.00	0.03	0.17			1.00	1.00	1.00	0.71
4	11.57		1.00	0.02	0.04	0.14		1.00	1.00	1.00	1.00	0.77
5	11.50	1.00	0.02	0.04	0.05	0.10	1.00	1.00	1.00	1.00	1.00	0.82
Symmetric Parameterization												
1	55.28					1.00					1.00	0.55
2	37.88				1.00	0.21				1.00	0.21	0.88
3	37.88			1.00	1.00	0.21			1.00	1.00	0.21	0.88

the N-stage goods are used as intermediate inputs or final goods. The last three rows of Table III instead present our estimation results when restricting $\alpha^X_n = \alpha^F_n$ and $T^X_j = T^F_j$, and we refer to these set of estimates as the symmetric parameterizations of our model. The full set of estimates can be found in Appendix Section A.5.

Let us first discuss the set of asymmetric parameterizations. Interestingly, we find that sequential production is only relevant for the production of intermediate inputs. That is, we find that $\alpha^F_N = 1$ for all N so that, effectively, the production of final good varieties requires a single stage of production even when we let the model have potentially more stages. In other words, our estimation procedure indicates that the model's fit does not improve when moving from roundabout to sequential production in final goods (i.e., from $N = 1$ to $N > 1$).

In contrast, we find that increasing the number of stages of sequential production used to produce intermediate inputs does improve the model's fit as $\alpha^X_2, \ldots, \alpha^X_N$ are all strictly less than one for $N > 1$. Crucially, it turns out that making use of this added flexibility has sizable effects on the levels of the estimated value-added shares γ_j. Intuitively, Table III shows that increasing the production length of intermediate inputs comes together with a high dependence on the upstream sequential inputs (i.e., high $1-\alpha^X_n$) which tends to decrease the VA/GO ratio. To continue to match the data, γ_j thus needs to increase in order to offset this effect. As we discuss below, this has important implications for the comparison of our counterfactual experiments across the model parameterized with different values of N. Having said this, we should note that the numerical fit of the model improves only very slightly when moving from $N = 2$ to larger N (i.e., the minimized objective function becomes increasingly flat when N increases beyond 2 as can be seen in Table III's second column). The reason is that, while variation in the country-level parameters γ_j, T^X_j, and T^F_j significantly enhance the model's flexibility, incorporating additional stage-level parameters α^X_n only enhance the model's flexibility marginally since this added flexibility has to be averaged across all countries. The fit of the model with $N = 2$ is, however, significantly better than with $N = 1$.

On the other hand, and perhaps surprisingly, the symmetric parameterizations indicate no need of additional flexibility beyond the sequential $N = 2$ case. That is, while the model's fit improves substantially between the roundabout and the two-stage model, estimating the symmetric model with $N > 2$ leads to a set of parameters that effectively shut down the more upstream stages (i.e., $\alpha_{N-1} = 1$ for all $N \geq 2$). As in the asymmetric case, the intuition for this result can be obtained from studying the estimated value-added shares γ_j. Since the symmetric roundabout model imposes $\Pi^X_{jj} = \Pi^F_{jj}$, the improvement in the model's fit achieved when moving from $N = 1$ to $N = 2$ is driven by the parameter α_2, which is estimated to be relatively low ($\alpha_2 = 0.21$) in order to fit the sharp differences across the observed domestic expenditure shares in final-goods and inputs. A consequence of this is that the VA/GO ratio falls relative to the roundabout case and so the value-added shares γ_j need to increase in order for the model to also fit this dimension of the data. As a result, imposing symmetry in the sequential production of intermediate inputs and final goods imposes so much structure on the estimated model that the average value-added share γ_j with $N = 2$ is high at 0.88 and close to the upper bound of one for many countries (see Appendix Table A.2). Increasing N to three or more stages of production no longer improves the model's fit in the symmetric case because there is no further room for the value-added shares to adjust upwards in order to fit the VA/GO ratios. Given the limitations of this symmetric parameterization, hereafter we focus the empirical exercises on the asymmetric case. Furthermore, Table III shows clearly that, for a given N, the asymmetric parameterizations fit the data much better than the symmetric parameterization.

We summarize our model's fit in Figure 4. Recall that since we are trying to fit the total variation in the data, there are no untargeted moments to compare our estimated model to. Rather, we use Figure 4 to show how the model's fit to the targeted moments improves when including different margins of adjustment. Since our model is fairly flexible, it can fit very accurately (in all parameterizations) both the distribution of country sizes (i.e., GDP shares) as well as the VA/GO ratio in each country. However, the heterogeneity in aggregate domestic expenditure shares Π^F_{jj} and Π^X_{jj} can only be accurately matched with the full flexibility of our parameters. Thus, we study the precision of our model's fit through the lense of these statistics.

Specifically, Figure 4 compares the domestic expenditure share ratio Π^F_{jj}/Π^X_{jj} in the data to the estimated model with the two panels on the left representing the asymmetric parameterization and the two panels in the right representing the symmetric parameterization. The two upper panels present the estimated model fit when $N = 1$ while the two lower panels present the estimated model when $N = 2$. As discussed above, the fit with $N > 2$ in the asymmetric case improves slightly but the effect is quite marginal so we do not include these plots; in the symmetric case the fit does not improve at all and so there is nothing to show. As can be seen, the roundabout model's fit is much better in the asymmetric case than in the symmetric case (the top two panels) indicating that heterogeneity across T^X_j and T^F_j is useful for matching the data. Further, the fit improves when moving from $N = 1$ to $N = 2$ indicating that the sequentiality of production is an additional margin that is useful for matching the heterogeneity in input shares observed in the data. Overall, the best fit is that shown in the southwest panel which presents the $N = 2$ estimated model with heterogeneity in both absolute productivity T^X_j and T^F_j and input expenditure shares α^X_n and α^F_n across intermediate inputs and final goods.

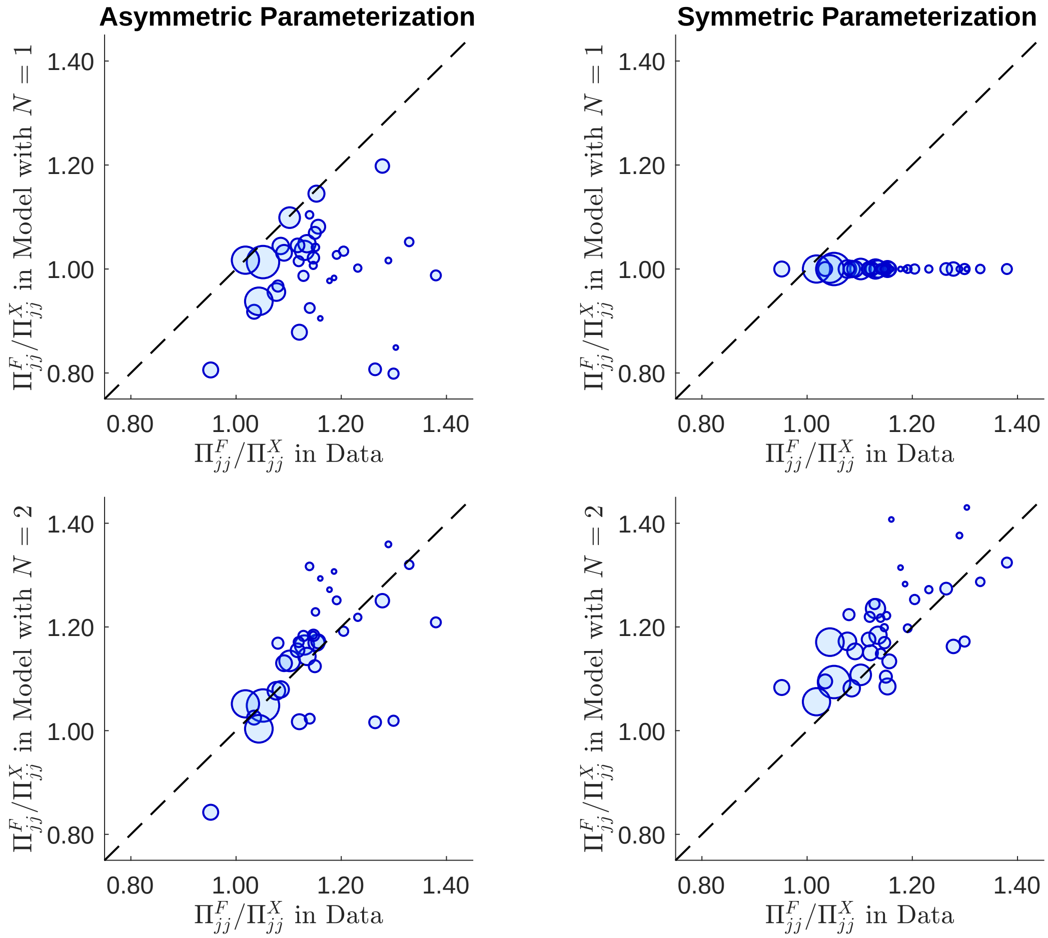

Figure 4: Model Fit.

6.2 Revisiting the Calibration of N

Up to now, we have fixed the number of stages N and estimated the model conditional on this parameter. While increasing N makes our estimation computationally more demanding, note that in terms of the parameters to estimate, this only amounts to estimating a longer vector of input shares α_n^X and α_n^F. Perhaps surprisingly, our structural estimation shuts down the upstream sequential stages in the production of finished final goods in our asymmetric parameterizations, while the number of stages in intermediate input production does not appear to be larger than 2 in the symmetric parameterizations. Why does this happen?

First, note that we are estimating an average chain-length N for the *whole* world economy, including industries in which producing goods might require long or short chains. If we were to estimate our model at the industry level for various industries, we might well recover heterogeneous values of N for these industries, reflecting variation in sequentiality across industries. As discussed before, however, an estimation of a full-fledged multi-industry model poses unsurmountable computational challenges (at present).

Second, while increasing N always improves the fit of the model, as N increases, the value-added shares γ_j also need to adjust upwards to make sure our model is able to match the observed aggregate

VA/GO ratio. As N increases more and more, the required adjustment in γ_j becomes less and less feasible, and the gain in the model's fit is smaller and smaller. This tight relation between the average number of stages of production and the aggregate VA/GO ratio resonates with the theoretical results in the input-output model of Fally (2012). Nevertheless, our estimated model yields a more subtle prediction: while the average number of stages of production is pinned down by the VA/GO ratio, our estimated model fits the data better when the total number of stages of production N increases as long as the volume of gross output traded at each stage adjusts accordingly.

Some readers might still object that observing no improvement in the model's fit for $N > N'$ is not synonymous with correctly identifying $N = N'$. In Appendix B.5, we show through simulations that there is a precise sense in which recovering the same parameters for $N \geq N'$ implies that the true N is indeed equal to N' and that $N > N'$ can be rejected. We focus on simulated economies with $J = 5$ countries and fix N at either 1, 2, or 3. For each economy we simulate a set of primitives of the model and compute the general equilibrium. We then take the resulting simulated WIOT entries and estimate the parameters using our MPEC algorithm. Our results confirm that we are able to recover the "true" value of N.

Before concluding this section on estimation, it is important to stress that our identification of N relies heavily on our assumption that the matrix of trade costs τ_{ij} is common (and symmetric) for inputs and final goods. For example, one can show that an extension of our framework without multi-stage production (i.e., $N = 1$) could be calibrated to *exactly* match a WIOT, provided that one allows for arbitrary and asymmetric trade costs τ_{ij}^X and τ_{ij}^F for intermediate inputs and final goods. Thus, admittedly, our data cannot reject $N = 1$ if one allows enough flexibility in the modeling of trade costs.

7 Counterfactuals

Having estimated the fundamental parameters of the model, we next explore how counterfactual changes in trade costs, holding other parameters constant, alter the entries of WIOTs, thereby affecting the real income and positioning of countries in GVCs.

7.1 Autarky and Zero Gravity

We begin by revisiting two focal counterfactual exercises in quantitative international trade, namely an increase in trade costs large enough to bring back autarky, and a complete elimination of trade barriers. Both of these counterfactuals are extreme in nature, but they are useful for understanding some distinctive features of our framework. From equation (18), the loss in welfare from reverting back to autarky in a single industry world is given by

$$\hat{W}_j^{aut} = \left(\pi_{jj}^F \left(\pi_{jj}^X \right)^{\frac{1-\gamma_j}{\gamma_j}} \right)^{\frac{1}{\theta}}, \tag{25}$$

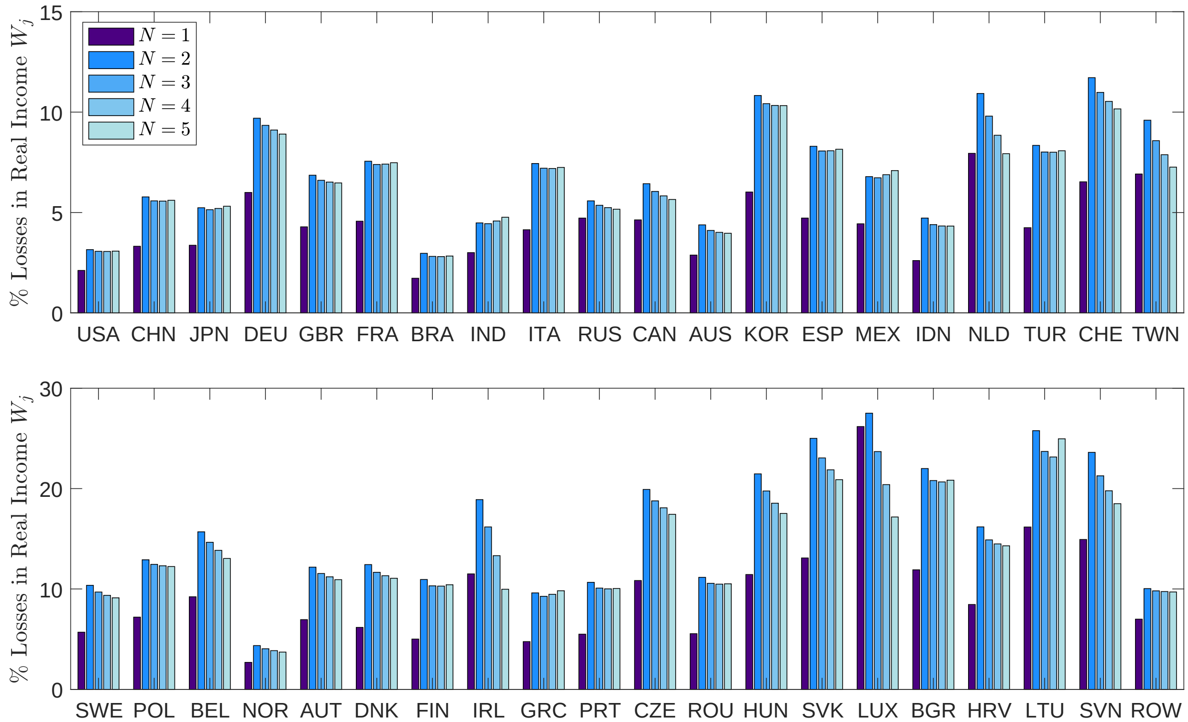

Figure 5: The Welfare Losses from Autarky.

where $\pi^F_{\boldsymbol{jj}}$ and $\pi^X_{\boldsymbol{jj}}$ are the share of final good and intermediate input varieties, respectively, produced entirely through domestic supply chains.[42] In general, these shares are not observable in the data and so have to be estimated through the lens of our model. However, in the special case of roundabout production, this formula depends on the aggregate domestic expenditure shares, i.e., $\hat{W}^{aut}_j = \left(\Pi^F_{jj}\left(\Pi^X_{jj}\right)^{(1-\gamma_j)/\gamma_j}\right)^{1/\theta}$, and so can be implemented directly with the observed domestic shares (as in Alexander 2017). We choose instead to follow the alternative approach of studying $\hat{W}^{aut}_j$ in the roundabout case using our estimated model with $N = 1$ in order to be internally consistent when comparing to the estimated model with $N > 1$ (in practice this is largely immaterial for our results).

Figure 5 presents the losses from reverting to autarky across all countries using the asymmetric parameterizations in Table III (note each panel has different y-axis). Three features are particularly salient. First, the gains from trade (or costs of autarky) are considerably larger with sequential production than in the roundabout model. On average, the losses with $N = 2$ are 60% higher than with $N = 1$. This evidence is consistent with Yi 's (2003) result that deep vertical specialization magnifies the costs of protectionism. A reversal to autarky when supply chains are highly fragmented is costly since countries can no longer exploit the comparative advantage of foreign countries in the various stages of production in value chains.

Second, while increasing N could, in principle, either increase or decrease the losses from going

[42]This formula still measures the real income gains from trade in the presence of trade imbalances. The implications for real spending, however, may be quite different since autarky implies a closing of trade imbalances.

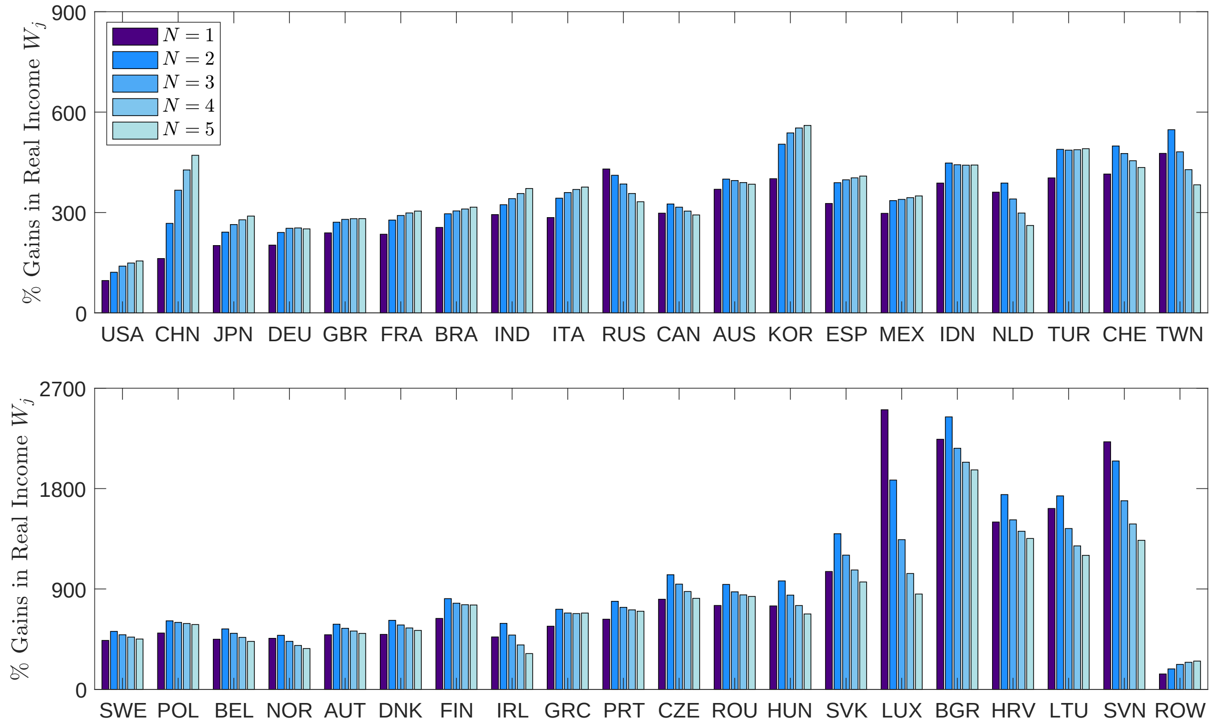

Figure 6: The Welfare Gains from Free Trade.

to autarky, in practice it either does not affect the losses (for countries such as the U.S. and China) or decreases the losses (for countries such as the Netherlands and Taiwan). To see what drives these results, note from (25) that three elements account for the autarky losses. Since final goods are produced in a single stage, the estimated domestic final good share $\pi^F_{\boldsymbol{jj}}$ roughly equals the observed aggregate domestic expenditure share Π^F_{jj} and so is constant across all N. On the other hand, remember from Table III that the value-added parameters γ_j are increasing with N. This force decreases the losses from trade because it implies that finished intermediate input varieties becomes less important for final consumption. In other words, leveraging international trade to deliver cheap inputs is less important when γ_j is high. Finally, as N increases, each country gets more opportunities to leverage other countries' comparative advantage through additional stages of production. Since upstream value added can be sourced from a richer combination of foreign countries, this decreases $\pi^X_{\boldsymbol{jj}}$. Overall, the increase in γ_j either cancels out or dominates the decrease in $\pi^X_{\boldsymbol{jj}}$: while disrupting intermediate input supply chains becomes more costly when N increases, it is also true that these supply chains become less important in terms of the value they contribute to final consumption.

As a corollary of this discussion, a third feature of our model is that the decrease in $\pi^X_{\boldsymbol{jj}}$ rarely dominates the increase in γ_j, and thus the gains from trade do not become unboundedly large as $N \to \infty$. As discussed in Section 4.3, while this might seem at odds with the results of Melitz and Redding (2014), this occurs because our estimation exercise targets the GDP/GO ratios observed in the data. Furthermore, while Proposition 4 generalized their results and showed that it is possible to obtain both gains from trade that become unboundedly large with a constant GDP/GO ratio,

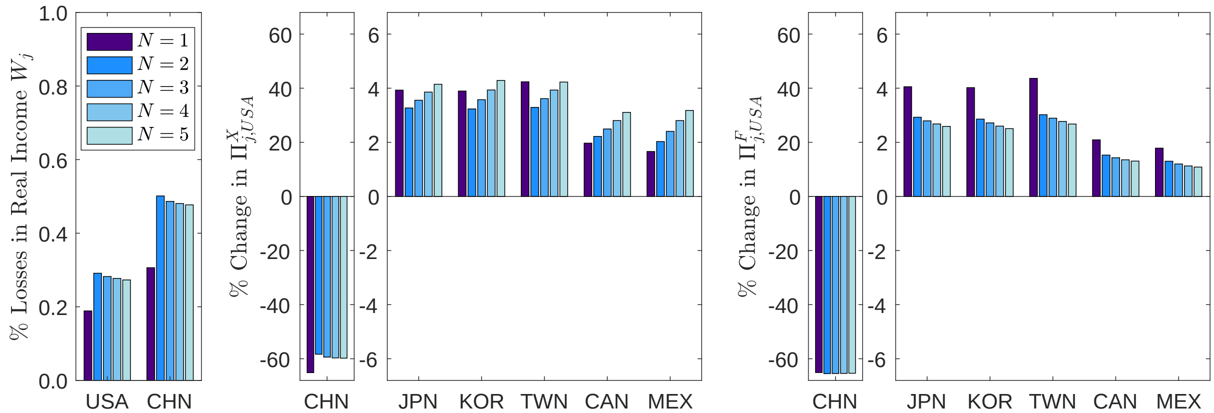

Figure 7: The U.S.-China Trade War.

Table III and Figure 5 show that in practice this is not the case.

We next explore the implications of a (hypothetical) complete elimination of trade barriers. The real income consequences of a move to a world with zero gravity are more pronounced. Figure 6 shows that without sequential production, i.e., with $N = 1$, these gains range from 100% for the United States and 160% for China, to a staggering 2500% for Luxembourg. In contrast to the autarky case, increasing N does raise the potential gains from trade by sizable amounts for some countries: increasing N from 1 through 5 raises the U.S. gains up to 150% and the China gains up to 470%. More generally, since the potential gains increase strongly with N for rich countries, the GDP-weighted potential gains from free trade across countries also increases from 225% with $N = 1$ to 276% and 317% when $N = 2$ and $N = 5$, respectively. Hence, the roundabout model vastly understates the average potential gains from free trade.

7.2 The U.S.-China Trade War

We now study how our model's distinctive elements shape a real-world inspired counterfactual: namely, a trade war between the U.S. and China in which trade barriers increase by 25% on each other's exports. The three panels in Figure 7 present the change in welfare, the change in U.S. intermediate input import shares, and the change in U.S. final good import shares. In the last two cases, we focus on the import shares from China, from three close competitors in Asian supply chains (Japan, Korea, and Taiwan), and from the other two NAFTA countries (Canada and Mexico).

In terms of welfare, Figure 7 reveals three features. First, both the U.S. and China suffer substantial welfare losses. Second, China's losses are about 50% higher than for the U.S. These results echo the patterns in Figure 5 and are driven by China being more exposed than the U.S. to international trade, and also to China being more exposed to the U.S. as a specific trade partner than the U.S. is to China. Third, the losses from a trade war are almost 50% higher in a world with sequential production, once again indicating that GVC linkages amplify the effects of trade-cost shocks.

Unsurprisingly, in terms of trade diversion, the last two panels show that China's U.S. import share falls dramatically while other countries' import share goes up. More interestingly, however, are the differential patterns observed across trading partners and across intermediate-input and final-good imports. Overall, Japan, Korea, and Taiwan are close substitutes for China in Asian supply chains and thus enjoy a strong export boom. The NAFTA countries also increase their U.S. import share, but at a smaller rate since they are competing less directly with Chinese exports.

Finally, note that the effects on intermediate-input and final-good exports differ strikingly in the context of multi-stage production. When $N = 1$, the growth of intermediate-input and final-good import shares is common across countries since there is a common trade elasticity for both types of goods. However, when $N > 1$ the export boom in intermediate inputs is increasing in N for the NAFTA countries, while the export boom of final goods is decreasing in N for all countries. Since the U.S. imports a large share of final goods from China, the fall in Chinese imports is partially offset by an increase in final-good imports from other countries. Remember, though, that the increasing trade elasticity implies that the increase in import costs is disproportionately large for final goods. This drives the U.S. to import an even larger amount of intermediate inputs from its main trade partners, Canada and Mexico, so that it can domestically finish producing final goods. This last force is strongest when sequential production is pervasive (high N) since there is more flexibility when allocating segments of the supply chain across locations. Thus, when N increases we see a substitution from final-good export growth towards intermediate-input export growth.

7.3 Regional vs Global Integration

As a third set of exercises, we contribute to the debate arguing that much of the global fragmentation in GVCs is actually regional in nature (see Johnson and Noguera 2012b). As should be intuitive, while the relative importance of global GVC integration monotonically increases when trade costs fall, the relative importance of regional GVC participation initially rises but eventually falls when trade costs are lowered sufficiently.

We explore the relative importance of domestic, regional and global value chains across several trade equilibria defined by a value of Δ_τ such that $\tau'_{ij} = 1 + \Delta_\tau(\tau_{ij} - 1)$, with τ_{ij} being our calibrated trade costs for the WIOD sample in 2014. Focusing on our estimated global economy with $N = 2$, we define a domestic GVC as $\boldsymbol{\ell}_d = \{USA, USA\}$ and associate the prevalence of *domestic* value chains in overall U.S. consumption with the share $\pi^X_{\boldsymbol{\ell}_d,USA}$. Similarly, we capture the relative prevalence of *regional* (or NAFTA) value chains in overall U.S. consumption by $\sum_{\boldsymbol{\ell}_r} \pi^X_{\boldsymbol{\ell}_r,USA}$, where $\boldsymbol{\ell}_r$ are all chains that only include the U.S., Canada or Mexico, with the exception of the chain $\boldsymbol{\ell}_d$. Finally, we define the relative prevalence of *global* value chains in U.S. consumption as $\sum_{\boldsymbol{\ell}_g} \pi^X_{\boldsymbol{\ell}_g,USA}$, where $\boldsymbol{\ell}_g$ are all the possible chains that involve at least one country outside of NAFTA. Naturally, the sum of these three relative measures is one. An important caveat is that, due to the use of a bundle of materials at each stage, what we label as domestic and regional value chains actually embody value added from countries outside NAFTA. In fact, for bounded trade costs, our model features no purely domestic value chains. Yet, the above taxonomy is useful for understanding the broad orientation of value chains serving U.S. consumers for different levels of trade costs.

The left panel of Figure 8 plots these three measures for various values of Δ_τ between $1/32 \simeq 0.031$ and 10, and provides evidence supporting our intuition for the evolution of regional and global

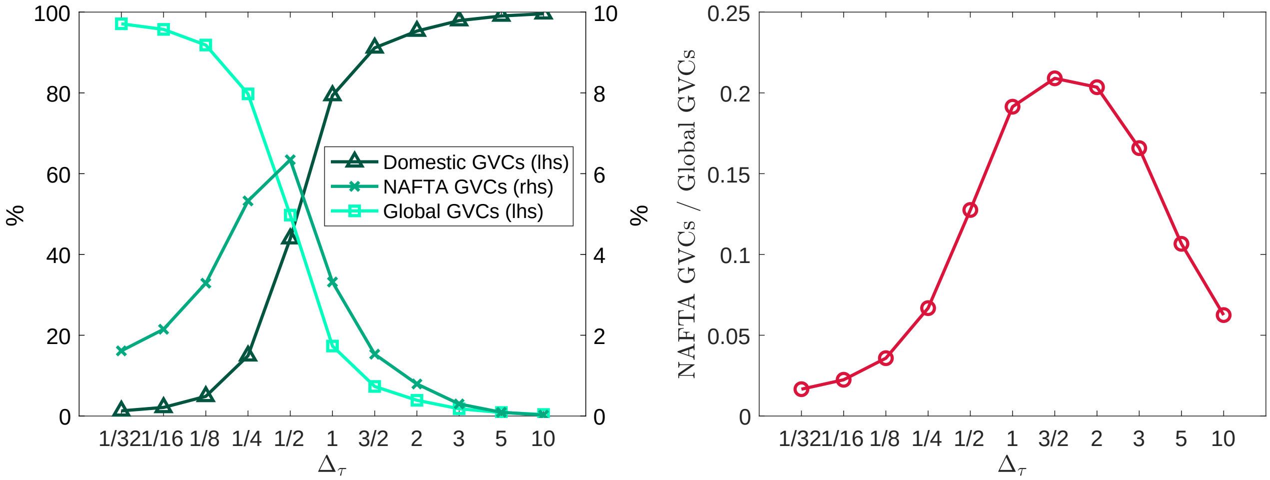

Figure 8: Regional vs Global Integration.

integration. Furthermore, the right panel of Figure 8 plots the ratio of the relative importance of regional (NAFTA) versus global value chains for the same values of Δ_τ. Interestingly, our benchmark equilibrium, $\Delta_\tau = 1$, is very close to the point at which the relative importance of regional value chains is maximized. Thus, further reductions in trade costs would reduce the relative importance of regional value chains in U.S. consumption, and increase that of global value chains.

7.4 A Multi-Industry Application

As a final exercise, we provide a preliminary empirical exploration of the multi-industry, multi-stage production model with input-output linkages described in Section 5. For computational reasons, we focus on the case of two industries and aggregate the WIOD's industries into manufacturing and services blocks. We then reestimate our model using a combination of the multi-industry extension of our MPEC algorithm and a simulated annealing algorithm. Again for the sake of simplicity, we focus on estimating the roundabout model with $N = 1$ and the GVC model with $N = 2$.

Table IV contains our estimation results. The first row corresponds to the benchmark roundabout model, while the last four rows correspond to our GVC model. We study four different GVC estimations for the following two reasons. First, as in the single industry case, we always find that the production of both manufacturing and services final goods requires a single stage of sequential production – i.e., $\alpha_2^{M,F} = \alpha_2^{S,F} = 1$ – with the intuition being analogous to the discussion in Section 6.1. However, we also find that the loss function is relatively flat in the parameter governing the sequential production of intermediate inputs in the services industry and this leads to four multiple local minima that fit the data quite well.[43] This can be seen in the heatmap in Appendix Section B.6 which plots the optimized loss function across $\alpha_2^{M,X}$ and $\alpha_2^{S,X}$. The loss function is quite flat in $\alpha_2^{S,X}$ because services intermediate inputs are relatively rare (54% of world services output is in final goods). On the other hand, the loss function is substantially nonlinear in $\alpha_2^{M,X}$ since manufactured

[43] We estimate the multi-industry model with only the largest $J = 11$ countries in order to tractably compute the loss function across the parameter space. This implies that the values of the optimized loss function in Tables III and IV are not comparable since the bilateral flows of the 'Rest-of-World' aggregate are larger in the latter case.

Table IV: Multi-Industry Counterfactuals.

	Estimation Results					World Weighted Average $\hat{W}_j$					
	Obj. F.	$\alpha_2^{M,X}$	$\alpha_2^{S,X}$	$\alpha_2^{M,F}$	$\alpha_2^{S,F}$	$\hat{\tau}^M = 2$	$\hat{\tau}^S = 2$	$\hat{\tau} = 2$	$\hat{\tau}^M = \infty$	$\hat{\tau}^S = \infty$	$\hat{\tau} = \infty$
Roundabout	49	1.00	1.00	1.00	1.00	3.7	1.0	4.6	4.9	1.1	6.0
GVC 1	34	0.00	0.18	1.00	1.00	3.6	1.6	5.1	4.8	1.9	6.7
GVC 2	40	1.00	0.40	1.00	1.00	3.7	2.4	5.9	4.9	3.0	7.7
GVC 3	37	0.00	0.93	1.00	1.00	3.6	3.0	6.5	4.8	23.1	26.8
GVC 4	39	1.00	0.91	1.00	1.00	3.8	3.4	7.0	5.0	21.7	25.5

intermediate inputs are quite prevalent (only 31% of world manufacturing output is in final goods). Second, we also compare these four GVC estimations in order to shed more light on how the model's main mechanisms interact in a multi-industry setting.

The last six columns of Table IV presents the world's weighted-average change in welfare across six counterfactuals. The first three counterfactuals correspond to a doubling of iceberg trade costs.[44] In the first case we assume that trade costs double only for manufactures, in the second only for services, and in the third for both industries. The last three counterfactuals are analogous but impose autarky-level trade costs.

Our multi-industry counterfactuals reveal four important insights. First, as is well known, the multi-industry gains from trade are higher than the single-industry gains. Second, shutting down trade in manufactures is, in most cases, much more costly than shutting down trade in services even though services accounts for a much larger share of world output. This occurs because the services industry is much less open than manufactures. Third, multi-stage production in services can lead to extremely high gains from trade. While this might sound counter-intuitive, explaining this further is useful for understanding the role of the model's increasing trade elasticity. In the last two estimations, GVC 3 and GVC 4, finished services inputs are built almost exclusively with downstream production (over 90% of stage-2 value-added). This leads to a tiny upstream trade elasticity and, thus, in the trade equilibria the upstream services inputs are sourced largely from abroad. Since the tiny trade elasticity leads to a tiny domestic share, this implies that countries are benefiting highly from other countries' comparative advantage and thus reverting back to autarky is extremely costly. Thus, a reversal to autarky in the estimations of GVC 3 and GVC 4 generates very high welfare losses that are driven mostly by services rather than by manufactures. Fourth, and relatedly, it should be made clear that these former statements are only true in the extreme case of autarky. When trade costs in services only double, the welfare losses across all GVC estimations are quite similar and the losses from manufactures dominate in all cases. Again, this is driven by the tiny upstream trade elasticity in services: While going back to autarky is extremely costly, the tiny trade elasticity also implies that these losses only kick in if trade costs change drastically.

[44] More precisely, we define the counterfactual trade costs as $\tau'_{ij} = 1 + 2\left(\tau_{ij} - 1\right)$.

8 Conclusion

In this paper, we have studied how trade barriers shape the location of production along GVCs. Relative to the case of free trade, trade costs generate interdependencies in the sourcing decisions of firms. Specifically, when deciding on the location of production of a given stage, firms necessarily take into account where the good is coming from and where it will be shipped next. As a result, instead of solving N location decisions (where N is the number of stages), firms need to solve the much more computationally burdensome problem of finding an optimal *path* of production. Despite these complications, we have proposed tools to feasibly solve the model in high-dimensional environments.

After deriving these results in partial equilibrium, we have developed a multi-stage general-equilibrium model in which countries specialize in different segments of GVCs. We have demonstrated that, due to the compounding effect of trade costs along value chains, relatively central countries gain comparative advantage in relatively downstream stages of production. We have also borrowed from the seminal work of Eaton and Kortum (2002) to develop a tractable quantitative model of GVCs in a multi-country environment with costly trade. Relative to previous quantitative models of multi-stage production, our suggested approach maps more directly to world Input-Output tables, and allows the use of a mix of calibration and structural estimation to back out the model's parameters. We finally illustrated some distinctive features of the model by performing counterfactual analyses.

Our framework is admittedly stylized and abstracts from many realistic features that we hope will be explored in future work. For instance, a potentially interesting avenue for future research would be to introduce scale economies (external or internal) into our analysis. In a previous version of the paper, we explored a variant of our model with external economies of scale featuring a proximity-concentration tradeoff. The interaction of trade costs and scale economies substantially enriches – but also complicates – the analysis. Although our dynamic programming approach is no longer feasible in that setting, the integer linear programming approach developed in Appendix A.1.2 is still quite powerful in that environment. We believe that variants of that approach could also prove useful in extending our framework to include internal economies of scale and imperfect competition. It would also be interesting to incorporate contractual frictions into our framework and study the optimal governance of GVCs in a multi-stage, multi-country environment.

Beyond these extensions of our framework, we view our work as a stepping stone for a future analysis of the role and scope of man-made trade barriers in GVCs. Although we have focused on an analysis of the implications of exogenously given trade barriers, our theoretical framework should serve as a useful platform to launch a study of the role of trade policies, and of policies more broadly, in shaping the position of countries in value chains. Should countries actively pursue policies that foster their participation in GVCs? Should they implement policies aimed at moving them to particular stages of those chains? If so, what are the characteristics of these particularly appealing segments of GVCs? These are the type of questions we hope to tackle in future research.

References

Adao, Rodrigo, Arnaud Costinot and Dave Donaldson (2017) "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade," *American Economic Review*, 107(3): 633-89.

Alexander, Patrick (2017), "Vertical Specialization and the Gains from Trade," Bank of Canada Staff Working Paper 2017-17.

Alfaro, Laura, Pol Antràs, Paola Conconi, and Davin Chor (2019), "Internalizing Global Value Chains: A Firm-Level Analysis,", *Journal of Political Economy*, 127(2), pp. 508-559.

Allen, Treb, and Costas Arkolakis (2019), "The Welfare Effects of Transportation Infrastructure Improvements," NBER Working Paper No. 25487.

Anderson, James E., and Eric van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review* 93(1): 170-192.

Antràs, Pol, and Davin Chor (2013), "Organizing the Global Value Chain," *Econometrica* 81(6): 2127-2204.

Antràs, Pol, and Davin Chor (2018), "On the Measurement of Upstreamness and Downstreamness in Global Value Chains," in *World Trade Evolution: Growth, Productivity and Employment*, pp. 126-194. Taylor & Francis Group.

Antràs, Pol, Davin Chor, Thibault Fally, and Russell Hillberry, (2012), "Measuring the Upstreamness of Production and Trade Flows," *American Economic Review Papers & Proceedings* 102(3): 412-416.

Antràs, Pol and Alonso de Gortari (2017), "On the Geography of Global Value Chains," NBER Working Paper No. 23456.

Antràs, Pol, Teresa Fort, and Felix Tintelnot (2017), "The Margin of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, Vol. 107 (9), pp. 2514-64.

Antràs, Pol, and Robert W. Staiger (2012), "Offshoring and the Role of Trade Agreements," *American Economic Review* 102, no. 7: 3140-3183.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare (2012), "New trade models, same old gains?" The American Economic Review 102.1: 94-130.

Baldwin, Richard, and Anthony Venables (2013), "Spiders and Snakes: Offshoring and Agglomeration in the Global Economy," *Journal of International Economics* 90(2): 245-254.

Baldwin, Richard and Daria Taglioni (2014), "Gravity chains: Estimating Bilateral Trade Flows When Parts and Components Trade Is Important," *Journal of Banking and Financial Economics*, vol. 2(2), pages 61-82, November.

Baqaee, David, and Emmanuel Farhi (2019), "Networks, Barriers, and Trade," NBER Working Paper No. 26108.

Basco, Sergi, and Martí Mestieri (2019), "The world income distribution: the effects of international unbundling of production," *Journal of Economic Growth*, 24(2), pp. 189-221.

Bergstrand, Jeffrey H., and Peter Egger (2010), "A General Equilibrium Theory for Estimating Gravity Equations of Bilateral FDI, Final Goods Trade and Intermediate Goods Trade," in *The Gravity Model in International Trade: Advances and Applications*, Cambridge University Press, New York (2010).

Blanchard, Emily, Chad Bown, and Robert Johnson (2018), "Global Supply Chains and Trade Policy", mimeo Dartmouth College.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou (2017), "Geography, Search Frictions and Trade Costs," mimeo Harvard University.

Byrd, Richard H., Jorge Nocedal, and Richard A. Waltz (2006), "KNITRO: An integrated package for nonlinear optimization", in *Large-Scale Nonlinear Optimization*, di Pillo, G, and M. Roma (Eds), Springer, 35–59.

Caliendo, Lorenzo, and Fernando Parro (2015), "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies* 82 (1): 1-44.

Costinot, Arnaud (2009), "An Elementary Theory of Comparative Advantage," *Econometrica* 77:4, pp. 1165-1192.

Costinot, Arnaud, and Andrés Rodríguez-Clare (2015), "Trade Theory with Numbers: Quantifying the Consequences of Globalization" in *Handbook of International Economics*, vol. 4: pp. 197-261.

Costinot, Arnaud, Jonathan Vogel, and Su Wang, (2013), "An Elementary Theory of Global Supply Chains," *Review of Economic Studies* 80(1): 109-144.

de Gortari, Alonso (2019), "Disentangling Global Value Chains," NBER Working Paper 25868.

de Gortari, Alonso (2020), "Global Value Chains and Increasing Returns."

Dekle, Robert, Jonathan Eaton, and Samuel Kortum (2008), "Global rebalancing with gravity: measuring the burden of adjustment." IMF Economic Review 55.3: 511-540.

Dixit, Avinash, and Gene Grossman, (1982), "Trade and Protection with Multistage Production," *Review of Economic Studies* 49(4): 583-594.

Eaton, Jonathan and Samuel Kortum, (2002), "Technology, Geography, and Trade," *Econometrica*, 70:5, 1741-1779.

Ethier, Wilfred J. (1984), "Higher Dimensional Issues in Trade Theory," Ch. 3 in *Handbook of International Economics*, Grossman, Gene M., and Kenneth Rogoff (Eds), Vol. 1.

Fally, Thibault (2012), "Production Staging: Measurement and Facts" mimeo University of Colorado Boulder.

Fally, Thibault, and Russell Hillbery (2018), "A Coasian Model of International Production Chains," *Journal of International Economics*, Volume 114, September 2018, Pages 299-315.

Fally, Thibault, and James Sayre (2018), "Commodity Trade Matters,"NBER Working Paper No. 24965.

Harms, Philipp, Oliver Lorz, and Dieter Urban, (2012), "Offshoring along the Production Chain," *Canadian Journal of Economics* 45(1): 93-106.

Head, Keith, and Philippe Mayer (2014), "Gravity Equations: Workhorse, Toolkit, and Cookbook," Ch. 3 in *Handbook of International Economics*, Gopinath, Gita, Elhanan Helpman and Kenneth Rogoff (Eds), Vol. 4.

Head, Keith, and John Ries (2001), "Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade," *American Economic Review*, 91(4): 858-876.

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

Johnson, Robert (2014), "Five Facts about Value-Added Exports and Implications for Macroeconomics and Trade Research," *Journal of Economic Perspectives*, 28(2), pp. 119-142.

Johnson, Robert, and Andreas Moxnes (2013), "Technology, Trade Costs, and the Pattern of Trade with Multi-Stage Production," mimeo.

Johnson, Robert C. and Guillermo Noguera (2012), "Accounting for Intermediates: Production Sharing and Trade in Value Added," *Journal of International Economics,* 86(2): 224-236.

Johnson, Robert C. and Guillermo Noguera (2012b), "Proximity and Production Fragmentation," *American Economic Review: Papers & Proceedings*, 102(3): 407-411.

Kikuchi, Tomoo, Kazuo Nishimura, and John Stachurski (2014), "Transaction Costs, Span of Control and Competitive Equilibrium," mimeo.

Kohler, Wilhelm, (2004), "International Outsourcing and Factor Prices with Multistage Production," *Economic Journal* 114(494): C166-C185.

Koopman, Robert, William Powers, Zhi Wang, and Shang-Jin Wei (2010), "Give Credit Where Credit Is Due," NBER Working Paper 16426, September.

Koopman, Robert, Zhi Wang, and Shang-Jin Wei (2014), "Tracing Value-Added and Double Counting in Gross Exports", *American Economic Review* 104 (2): 459-94.

Kremer, Michael, (1993), "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics* 108(3): 551-575.

Lee, Eunhee, and Kei-Mu Yi (2018), "Global Value Chains and Inequality with Endogenous Labor Supply," *Journal of International Economics*, 115, pp. 223-241.

Lenzen, M., Kanemoto, K., Moran, D., Geschke, A. (2012), "Mapping the Structure of the World Economy," *Environmental Science and Technology*, 46(15) pp 8374-8381.

Lenzen, M., Moran, D., Kanemoto, K., Geschke, A. (2013), "Building Eora: A Global Multi-regional Input-Output Database at High Country and Sector Resolution," *Economic Systems Research*, 25:1, 20-49.

Mayer, Thierry and Soledad Zignago (2006), "Notes on CEPII's distances measures" mimeo available at http://www.cepii.fr/distance/noticedist_en.pdf,

Melitz, Marc J, and Stephen J Redding (2014), "Missing Gains from Trade?" *American Economic Review* 104 (5): pp. 317-21.

Morales, Eduardo, Gloria Sheu, and Andrés Zahler (2014), "Gravity and Extended Gravity: Using Moment Inequalities to Estimate a Model of Export Entry," NBER Working Paper No. 19916.

Oberfield, Ezra (2018), "A Theory of Input-Output Architecture," *Econometrica*, Vol. 86:2, pp. 559-589.

Ossa, Ralph (2015), "Why Trade Matters After All," *Journal of International Economics* 97(2): pp. 266-277.

Sanyal, Kalyan K., and Ronald W. Jones, (1982), "The Theory of Trade in Middle Products," *American Economic Review* 72(1): 16-31.

Su, Che-Lin, and Kenneth L. Judd, (2012), "Constrained optimization approaches to estimation of structural models." Econometrica 80(5) : 2213-2230.

Timmer, Marcel P., Abdul Azeez Erumban, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries (2014), "Slicing Up Global Value Chains." *Journal of Economic Perspectives* 28, no. 2: 99-118.

Timmer, Marcel P., Eric Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries (2015), "An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production," *Review of International Economics*, 23: 575–605.

Tintelnot, Felix (2017), "Global Production with Export Platforms," *Quarterly Journal of Economics,* 132 (1): 157-209.

Tyazhelnikov, Vladimir (2016), "Production Clustering and Offshoring," mimeo UC Davis.

Wang, Zhi, Shang-Jin Wei, Xinding Yu, Kunfu Zhu (2017), "Measures of Participation in Global Value Chains and Global Business Cycles," NBER Working Paper 23222, March.

Yi, Kei-Mu, (2003), "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy* 111(1): 52-102.

A Supplement to "On the Geography of Global Value Chains" (For Typeset Publication Online)

A.1 Mathematical Proofs

A.1.1 Increasing Trade-Cost Elasticity

We demonstrate that Proposition 1 holds true for arbitrary constant-returns-to-scale production technologies. With that in mind, let the sequential cost function associated with a path of production $\boldsymbol{\ell} = \{\ell(1), \ell(2), ..., \ell(N)\}$ be defined by

$$p^n_{\ell(n)}(\boldsymbol{\ell}) = g^n_{\ell(n)}\left(c_{\ell(n)}, p^{n-1}_{\ell(n-1)}(\boldsymbol{\ell})\,\tau_{\ell(n-1)\ell(n)}\right), \text{ for all } n \in \mathcal{N}, \tag{A.1}$$

where the stage- and country-specific cost functions $g^n_{\ell(n)}$ in equation (A.1) are assumed to feature constant-returns-to-scale and diminishing marginal products. The cost of the first stage depends only on the local composite factor, so constant returns to scale implies $p^1_{\ell(1)}(\boldsymbol{\ell}) = g^1_{\ell(1)}\left(c_{\ell(1)}\right)$ for all paths $\boldsymbol{\ell}$, with the function $g^1_{\ell(1)}$ necessarily being linear in $c_{\ell(1)}$.

Define $\tilde{p}^{n-1}_{\ell(n)}(\boldsymbol{\ell}) = p^{n-1}_{\ell(n-1)}(\boldsymbol{\ell})\,\tau_{\ell(n-1)\ell(n)}$ to be the price paid in $\ell(n)$ for the good finished up to stage $n-1$ in country $\ell(n-1)$, so that we can express the sequential unit cost function as

$$p^n_{\ell(n)}(\boldsymbol{\ell}) = g^n_{\ell(n)}\left(c_{\ell(n)}, \tilde{p}^{n-1}_{\ell(n)}(\boldsymbol{\ell})\right).$$

Define the elasticity of $p^F_j(\boldsymbol{\ell})$ with respect to the trade costs that stage n's production faces as

$$\beta^j_n = \frac{\partial \ln p^F_j(\boldsymbol{\ell})}{\partial \ln \tau_{\ell(n)\ell(n+1)}},$$

with the convention that $\ell(N+1) = j$ so that β^j_N is the elasticity of $p^F_j(\boldsymbol{\ell})$ with respect to the trade costs faced when shipping assembled goods to final consumers in j. Because $\tau_{\ell(n)\ell(n+1)}$ increases $\tilde{p}^n_{\ell(n+1)}(\boldsymbol{\ell})$ with a unit elasticity, the following recursion holds for all $n' > n$

$$\frac{\partial \ln p^{n'+1}_{\ell(n'+1)}(\boldsymbol{\ell})}{\partial \ln \tau_{\ell(n)\ell(n+1)}} = \frac{\partial \ln p^{n'+1}_{\ell(n'+1)}(\boldsymbol{\ell})}{\partial \ln \tilde{p}^{n'}_{\ell(n'+1)}(\boldsymbol{\ell})} \frac{\partial \ln p^{n'}_{\ell(n')}(\boldsymbol{\ell})}{\partial \ln \tau_{\ell(n)\ell(n+1)}}.$$

At the same time, the unit cost elasticity at stage $n+1$ satisfies

$$\frac{\partial \ln p^{n+1}_{\ell(n+1)}(\boldsymbol{\ell})}{\partial \ln \tau_{\ell(n)\ell(n+1)}} = \frac{\partial \ln p^{n+1}_{\ell(n+1)}(\boldsymbol{\ell})}{\partial \ln \tilde{p}^{n}_{\ell(n+1)}(\boldsymbol{\ell})}.$$

Hence, the elasticity of finished good prices can be decomposed as

$$\beta^j_n = \prod_{n'=n+1}^{N} \frac{\partial \ln p^{n'}_{\ell(n')}(\boldsymbol{\ell})}{\partial \ln \tilde{p}^{n'-1}_{\ell(n')}(\boldsymbol{\ell})}, \tag{A.2}$$

invoking the convention $\prod_{n'=N+1}^{N} f(n') = 1$ for any function $f(\cdot)$. Constant returns to scale in production implies that the function $g^n_{\ell(n)}$ is homogeneous of degree one. As a result, the elasticity of unit costs with respect to input prices is always less or equal than one, so for all $n > 1$ we have

$$\frac{\partial \ln p^{n}_{\ell(n)}(\boldsymbol{\ell})}{\partial \ln \tilde{p}^{n-1}_{\ell(n)}(\boldsymbol{\ell})} \leq 1,$$

with strict inequality whenever a stage adds value to the product. From equation (A.2), it is then clear that

$$\beta^1_j \leq \beta^2_j \leq \cdots \leq \beta^N_j = 1,$$

with strict inequality when value added is positive at all stages.

A.1.2 Fighting the Curse of Dimensionality: Dynamic and Linear Programming

When discussing the lead-firm problem in section 2.2, we mentioned that there are J^N sequences that deliver distinct finished good prices $p^F_j(\boldsymbol{\ell})$ in country j. Hence, solving for the optimal sequences $\boldsymbol{\ell}^j$ for all j by brute force requires J^{N+1} computations and is infeasible to do when J and N are sufficiently large. However, we show below that use of dynamic programming surmounts this problem by reducing the computation of all sequences to only $J \times N \times J$ computations. Furthermore, in the special case in which production is Cobb-Douglas, the minimization problem can be modeled with zero-one linear programming, for which very efficient algorithms exist.

I. Dynamic Programming

Define $\boldsymbol{\ell}^j_n \in \mathcal{J}^n$ as the optimal sequence for delivering the good completed up to stage n to producers in country j. This term can be found recursively for all $n = 1, \ldots, N$ by simply solving

$$\boldsymbol{\ell}^j_n = \underset{k \in \mathcal{J}}{\arg\min} p^n_k \left(\boldsymbol{\ell}^k_{n-1}\right) \tau_{kj}, \tag{A.3}$$

since the optimal source of the good completed up to stage n is independent of the local factor cost c_j at stage n, of the specifics of the cost function g^n_j, or of the future path of the good. For this same reason, we have written the pricing function p^n_k in terms of the $n-1$ stage sequence $\boldsymbol{\ell}^k_{n-1}$ since it does not depend on future stages of production (though it should be clear that p^n_k will also be a function of the production costs and technology available for producers at that chosen location k). The convention at $n = 1$ is that there is no input sequence so that $\boldsymbol{\ell}^k_0 = \varnothing$ for all $k \in \mathcal{J}$ and the price depends only the composite factor cost: $p^1_k(\varnothing) = g^1_k(c_k)$.

The formulation in (A.3) makes it clear that the optimal path to deliver the assembled good to consumers in each country j, i.e., $\boldsymbol{\ell}^j = \boldsymbol{\ell}^j_N$, can be solved recursively by comparing J numbers for each location $j \in \mathcal{J}$ at each stage $n \in \mathcal{N}$, for a total of only $J \times N \times J$ computations.

To further understand this dynamic programming approach, Figure A.1 illustrates a case with 3 stages and 4 countries. Instead of computing $J^N = 64$ paths for each of the four locations of consumption, it suffices to determine the optimal source of (immediately) upstream inputs (which entails $J \times J = 16$ computations at stages $n = 2$ and $n = 3$, and for consumption). In the example, the optimal production path to serve consumers in A, B, and C is $A \to B \to B$, while the optimal path to serve consumers in D is $C \to D \to D$.

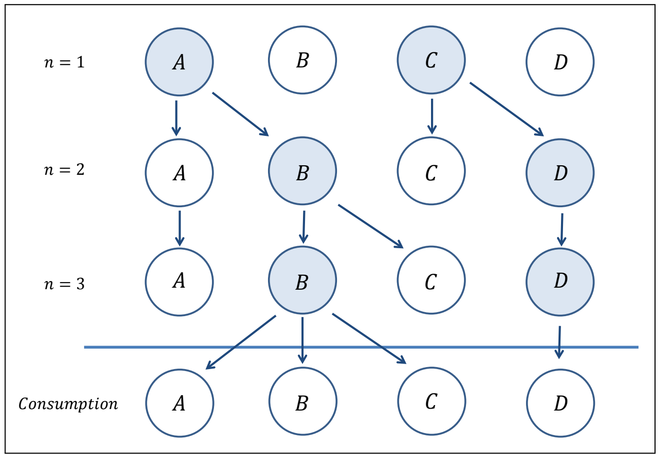

Figure A.1: Dynamic Programming − An Example with Four Countries and Three Stages

II. Linear Programming

In the special case in which production is Cobb-Douglas, the optimal sourcing sequence can be written as a log-linear minimization problem

$$\boldsymbol{\ell}^j = \arg \min_{\boldsymbol{\ell} \in \mathcal{J}^N} \sum_{n=1}^{N-1} \beta_n \ln \tau_{\ell(n)\ell(n+1)} + \ln \tau_{\ell(N)j} + \sum_{n=1}^{N} \alpha_n \beta_n \ln \left(a^n_{\ell(n)} c_{\ell(n)} \right).$$

This can in turn be reformulated as the following zero-one integer linear programming problem

$$\begin{aligned} \boldsymbol{\ell}^j = \arg\min & \sum_{n=1}^{N-1} \beta_n \sum_{k \in \mathcal{J}} \sum_{k' \in \mathcal{J}} \zeta^n_{kk'} \left(\ln \tau_{kk'} + \alpha_n a^n_k c_k \right) + \sum_{k \in \mathcal{J}} \zeta^N_k \left(\ln \tau_{kj} + \alpha_N a^N_k c_k \right) \\ \text{s.t.} & \sum_{k' \in \mathcal{J}} \zeta^n_{k'k} = \sum_{k' \in \mathcal{J}} \zeta^{n+1}_{kk'}, \forall k \in \mathcal{J}, n = 1, \ldots, N-2 \\ & \sum_{k' \in \mathcal{J}} \zeta^{N-1}_{k'k} = \zeta^N_k, \forall k \in \mathcal{J} \\ & \sum_{k \in \mathcal{J}} \zeta^N_k = 1; \; \zeta^n_{kk'}, \zeta^N_k \in \{0, 1\}. \end{aligned}$$

A.1.3 Proof of Proposition 3

If there is free trade or τ is constant across all country pairs (including domestically), then all countries source each variety from the same sequence of countries with $\pi_{\boldsymbol{\ell}j} = \pi_{\boldsymbol{\ell}}$ for all $j \in \mathcal{J}$. Analogously, price indices are the same in all markets so that $P_j = P$ for all $j \in \mathcal{J}$. The probability of sourcing a variety through a given sequence is thus

$$\pi_{\boldsymbol{\ell}} = \frac{\prod_{n \in \mathcal{N}} \left(T^n_{\ell(n)} w^{-\gamma\theta}_{\ell(n)} \right)^{1/N}}{\sum_{\boldsymbol{\ell}' \in \mathcal{J}^N} \prod_{n \in \mathcal{N}} \left(T^n_{\ell'(n)} w^{-\gamma\theta}_{\ell'(n)} \right)^{1/N}}.$$

We will now prove that wages are equalized across countries. Note that the total probability of any country being in a given stage n is the same regardless of the destination country and equals

$$\sum_{i\in\mathcal{J}} \Pr\left(\Lambda_i^n\right) = \sum_{i\in\mathcal{J}}\sum_{\boldsymbol{\ell}\in\Lambda_i^n} \frac{\prod_{n'\in\mathcal{N}}\left(T_{\ell(n')}^{n'} w_{\ell(n')}^{-\gamma\theta}\right)^{1/N}}{\Theta} = \sum_{i\in\mathcal{J}}\left(T_i^n w_i^{-\gamma\theta}\right)^{1/N} \times \frac{\prod_{n'\in\mathcal{N}\backslash n}\left(T_{\ell(n')}^{n'} w_{\ell(n')}^{-\gamma\theta}\right)^{1/N}}{\Theta}.$$

Now, suppose that wages are common across countries with $w_j = w$ for all $j \in \mathcal{J}$. Since the probability of any country being at a given stage n needs to equal 1, this implies that

$$\sum_{i\in\mathcal{J}} \left(T_i^n\right)^{1/N} \times \frac{\prod_{n'\in\mathcal{N}\backslash n}\left(T_{\ell(n')}^{n'}\right)^{1/N}}{w^{\gamma\theta}\Theta} = 1 \Rightarrow \frac{\prod_{n'\in\mathcal{N}\backslash n}\left(T_{\ell(n')}^{n'}\right)^{1/N}}{w^{\gamma\theta}\Theta} = \frac{1}{J\overline{T}},$$

where the second line uses our assumption that the geometric mean of T_i^n across countries is constant across stages of production. Let us now plug this into the right-hand side of the general equilibrium equation together with our guess that wages are equalized across countries

$$\begin{aligned} w_i &= \sum_{j\in\mathcal{J}}\sum_{n\in\mathcal{N}} \frac{1}{N} \times \Pr\left(\Lambda_i^n, j\right) \times w = \sum_{n\in\mathcal{N}} \frac{1}{N} \times \frac{\left(T_i^n\right)^{1/N} \times \prod_{n'\in\mathcal{N}\backslash n}\left(T_{\ell(n')}^{n'}\right)^{1/N}}{w^{\gamma\theta}\Theta} \times Jw \\ &= \sum_{n\in\mathcal{N}} \frac{1}{N} \times \frac{\left(T_i^n\right)^{1/N}}{J\overline{T}} \times Jw = \frac{1}{N} \times \frac{N\overline{T}}{J\overline{T}} \times Jw = w. \end{aligned}$$

Where the third line uses the previous result and where the fourth line uses our assumption that the geometric mean of T_i^n across stages of production is constant across countries. Hence, guessing that wages are equalized across countries delivers a fixed point in those wages. Since the equilibrium is unique, this is the only set of wages satisfying the general equilibrium equation.

To derive the share of goods produced in a domestic supply chain under free trade, rewrite Θ as

$$\Theta = \prod_{n\in\mathcal{N}}\sum_{i\in\mathcal{J}} \left(T_i^n\right)^{1/N} = \left(J \times \overline{T}\right)^N = \left(J \times \frac{1}{N}\sum_{n\in\mathcal{N}}\left(T_j^n\right)^{1/N}\right)^N,$$

for any $j \in \mathcal{J}$. Inserting this into the domestic expenditure share finalizes the proof

$$\pi_{\boldsymbol{j}} = \left(\frac{\text{Geometric Mean}_n\left[\left(T_j^n\right)^{1/N}\right]}{J \times \text{Arithmetic Mean}_n\left[\left(T_j^n\right)^{1/N}\right]}\right)^N.$$

A.1.4 Proof of Proposition 4

If all countries are symmetric, wages are equalized and the domestic expenditure share is

$$\pi_{\boldsymbol{j}} = \frac{1}{\sum_{\boldsymbol{\ell}'\in\mathcal{J}^N}\prod_{n\in\mathcal{N}}\left(\tau_{\ell(n)\ell(n+1)}\right)^{-\theta\beta_n}}.$$

The denominator can be rewritten as

$$
\begin{aligned}
&\sum_{\ell(1)\in\mathcal{J}} \cdots \sum_{\ell(N)\in\mathcal{J}} \prod_{n\in\mathcal{N}} \left(\tau_{\ell(n)\ell(n+1)}\right)^{-\theta\beta_n}, \\
&= \sum_{\ell(1)\in\mathcal{J}} \sum_{\ell(2)\in\mathcal{J}} \left(\tau_{\ell(1)\ell(2)}\right)^{-\beta_1\theta} \times \cdots \times \sum_{\ell(N)\in\mathcal{J}} \left(\tau_{\ell(N-1)\ell(N)}\right)^{-\theta\beta_{N-1}} \times \left(\tau_{\ell(N)j}\right)^{-\theta\beta_N}, \\
&= \sum_{\ell(N)\in\mathcal{J}} \left(1+(J-1)\,\tau^{-\beta_1\theta}\right) \times \left(1+(J-1)\,\tau^{-\beta_2\theta}\right) \times \cdots \times \left(1+(J-1)\,\tau^{-\beta_{N-1}\theta}\right) \times \left(\tau_{\ell(N)j}\right)^{-\theta\beta_N}, \\
&= \prod_{n=1}^{N} \left(1+(J-1)\,\tau^{-\beta_n\theta}\right).
\end{aligned}
$$

Substituting this in the domestic share finishes the proof.

A.1.5 Proof of Proposition 5

Let $(\tau_{ij})^{-\theta} = \rho_i\rho_j$. In such a case, the probability of country j sourcing through $\boldsymbol{\ell}$ reduces to

$$
\pi_{\boldsymbol{\ell}j} = \frac{\prod_{m=1}^{N} \left(T_{\ell(m)} \left(c_{\ell(m)}\right)^{-\theta}\right)^{\alpha_m\beta_m} \left(\rho_{\ell(m)}\right)^{\beta_{m-1}+\beta_m}}{\sum_{\boldsymbol{\ell}\in\mathcal{J}} \prod_{m=1}^{N} \left(T_{\ell(m)} \left(c_{\ell(m)}\right)^{-\theta}\right)^{\alpha_m\beta_m} \left(\rho_{\ell(m)}\right)^{\beta_{m-1}+\beta_m}}
$$

and is thus independent of the destination country j. The aggregate probability of observing country i in location n can thus be expressed as

$$
\Pr\left(\Lambda_i^n\right) = \sum_{\boldsymbol{\ell}\in\Lambda_i^n} \pi_{\boldsymbol{\ell}j} = \frac{\sum_{\boldsymbol{\ell}\in\Lambda_i^n} \prod_{m=1}^{N} \left(T_{\ell(m)} \left(c_{\ell(m)}\right)^{-\theta}\right)^{\alpha_m\beta_m} \left(\rho_{\ell(m)}\right)^{\beta_{m-1}+\beta_m}}{\sum_{k\in J} \sum_{\boldsymbol{\ell}\in\Lambda_k^n} \prod_{m=1}^{N} \left(T_{\ell(m)} \left(c_{\ell(m)}\right)^{-\theta}\right)^{\alpha_m\beta_m} \left(\rho_{\ell(m)}\right)^{\beta_{m-1}+\beta_m}}. \tag{A.4}
$$

But note that we can decompose this as

$$
\Pr\left(\Lambda_i^n\right) = \frac{\left(T_i\left(c_i\right)^{-\theta}\right)^{\alpha_n\beta_n} \left(\rho_i\right)^{\beta_{n-1}+\beta_n} \times \sum_{\boldsymbol{\ell}\in\Lambda_i^n} \prod_{m\neq n} \left(T_{\ell(m)} \left(c_{\ell(m)}\right)^{-\theta}\right)^{\alpha_m\beta_m} \left(\rho_{\ell(m)}\right)^{\beta_{m-1}+\beta_m}}{\sum_{k\in\mathcal{J}} \left(T_k\left(c_k\right)^{-\theta}\right)^{\alpha_n\beta_n} \left(\rho_k\right)^{\beta_{n-1}+\beta_n} \times \sum_{\boldsymbol{\ell}\in\Lambda_k^n} \prod_{m\neq n} \left(T_{\ell(m)} \left(c_{\ell(m)}\right)^{-\theta}\right)^{\alpha_m\beta_m} \left(\rho_{\ell(m)}\right)^{\beta_{m-1}+\beta_m}} \tag{A.5}
$$

$$
= \frac{\left(T_i\left(c_i\right)^{-\theta}\right)^{\alpha_n\beta_n} \left(\rho_i\right)^{\beta_{n-1}+\beta_n}}{\sum_{k\in\mathcal{J}} \left(T_k\left(c_k\right)^{-\theta}\right)^{\alpha_n\beta_n} \left(\rho_k\right)^{\beta_{n-1}+\beta_n}} \tag{A.6}
$$

where the second line follows from the fact that, for GVCs in the sets Λ_i^n and Λ_k^n, the set of all possible paths excluding the location of stage n are necessarily identical (and independent of the country where n takes place), and thus the second terms in the numerator and denominator of the first line cancel out.

For the special symmetric case with $\alpha_n\beta_n = 1/N$ and $\alpha_n = 1/n$ we obtain that

$$
\Pr\left(\Lambda_i^n\right) = \frac{\left(T_i\left(c_i\right)^{-\theta}\right)^{\frac{1}{N}} \left(\rho_i\right)^{\frac{2n-1}{N}}}{\sum_{k\in\mathcal{J}} \left(T_k\left(c_k\right)^{-\theta}\right)^{\frac{1}{N}} \left(\rho_k\right)^{\frac{2n-1}{N}}}
$$

Now consider our definition of upstreamness

$$U(i)=\sum_{n=1}^{N}(N-n+1)\times\frac{\Pr\left(\Lambda_i^n\right)}{\sum\limits_{n'=1}^{N}\Pr\left(\Lambda_i^{n'}\right)}. \tag{A.7}$$

This is equivalent to the expect distance from final-good demand at which a country will contribute to global value chains. The expectation is defined over a country-specific probability distribution over stages, $f_i(n)=\Pr\left(\Lambda_i^n\right)/\sum_{n'=1}^{N}\Pr\left(\Lambda_i^{n'}\right)$.

Finally, note that for two countries with $\rho_{i'}>\rho_i$ and two inputs with $n'>n$ we necessarily have

$$\frac{f_{i'}(n')/f_{i'}(n)}{f_i(n')/f_i(n)}=\left(\frac{\rho_{i'}}{\rho_i}\right)^{2\left(n'-n\right)/N}>1.$$

As a result, the probability functions $f_{i'}(n)$ and $f_i(n)$ satisfy the monotone likelihood ratio property in n. As is well known, this is a sufficient condition for $f_{i'}(n)$ to first-order stochastically dominate $f_i(n)$ when $\rho_{i'}>\rho_i$. But then it is immediate that $\mathbb{E}_{f_{i'}}[n]>\mathbb{E}_{f_i}[n]$, and thus the expected value in (A.7), which is simply $N+1-\mathbb{E}_{f_i}[n]$, will be lower for country i' than for country i when $\rho_{i'}>\rho_i$. This completes the proof of Proposition 5.

We can finally consider the case with a general path of α_n, but common technology $T_i=T$ across countries. From equation A.6, we have

$$\Pr\left(\Lambda_i^n,j\right)=\frac{(c_i)^{-\theta\alpha_n\beta_n}\left(\rho_i\right)^{\beta_{n-1}+\beta_n}}{\sum\limits_{k\in\mathcal{J}}(c_k)^{-\theta\alpha_n\beta_n}\left(\rho_k\right)^{\beta_{n-1}+\beta_n}}.$$

We then have

$$\frac{\Pr\left(\Lambda_i^{n'}\right)/\Pr\left(\Lambda_j^{n'}\right)}{\Pr\left(\Lambda_i^n\right)/\Pr\left(\Lambda_j^n\right)}=\left(\frac{c_i}{c_j}\right)^{-\left(\alpha_{n'}\beta_{n'}-\alpha_n\beta_n\right)}\left(\frac{\rho_i}{\rho_j}\right)^{\beta_{n'-1}+\beta_{n'}-\beta_{n-1}-\beta_n}$$

Take $n'=n+1$. Then

$$\frac{\Pr\left(\Lambda_i^{n'}\right)/\Pr\left(\Lambda_j^{n'}\right)}{\Pr\left(\Lambda_i^n\right)/\Pr\left(\Lambda_j^n\right)}=\left(\frac{c_i}{c_j}\right)^{-\theta\left(\alpha_{n+1}\beta_{n+1}-\alpha_n\beta_n\right)}\left(\frac{\rho_i}{\rho_j}\right)^{\beta_{n+1}-\beta_{n-1}}$$

Let's inspect the exponents more closely. Note $\beta_{n-1}=(1-\alpha_n)\beta_n$, so $\alpha_n\beta_n=\beta_n-\beta_{n-1}$ and

$$\frac{\Pr\left(\Lambda_i^{n'}\right)/\Pr\left(\Lambda_j^{n'}\right)}{\Pr\left(\Lambda_i^n\right)/\Pr\left(\Lambda_j^n\right)}=\left(\left(\frac{c_i}{c_j}\right)^{-\theta}\right)^{\beta_{n+1}-2\beta_n+\beta_{n-1}}\left(\frac{\rho_i}{\rho_j}\right)^{\beta_{n+1}-\beta_{n-1}}.$$

But

$$\beta_{n+1}-2\beta_n+\beta_{n-1}<\beta_{n+1}-\beta_{n-1}$$

because $\beta_{n-1}<\beta_n$. This can be iterated starting for $n''=n'+1$. This result implies that a sufficient condition for

$$\frac{\Pr\left(\Lambda_i^{n'}\right)/\Pr\left(\Lambda_j^{n'}\right)}{\Pr\left(\Lambda_i^n\right)/\Pr\left(\Lambda_j^n\right)}>1$$

for $n'>n$ and $\rho_i>\rho_j$ is that $(c_i)^{-\theta}\rho_i$ is larger for more central countries. Unfortunately, the general equilibrium conditions of the model are too complex for us to be able to formally establish that this is indeed the case for all possible parameters values. But, as stated in the main text, we have run millions of simulations and have not found a single case contradicting the claim.

A.2 General Equilibrium under Decentralized Approaches

This Appendix demonstrates the isomorphism between the general equilibrium conditions derived under the lead firm (chain-productivity) formulation in the main text, and the two alternative decentralized approaches outlined in 3.2.

A.2.1 Incomplete Information Approach

We begin with the first approach with stage-specific Fréchet distributions and incomplete information. On the technology side, we now assume that $1/a_i^n(z)$ is drawn independently (across goods and stages) from a Fréchet distribution satisfying

$$\Pr\left(a_i^n(z)^{\alpha_n\beta_n} \geq a\right) = \exp\left\{-a^{\theta}(T_i)^{\alpha_n\beta_n}\right\}. \tag{A.8}$$

To build intuition, we begin by sketching why and how the approach works for the simple case with only two stages, input production (stage 1) and assembly (stage 2). Later, we will show how the approach naturally generalizes to the case $N > 2$.

With $N = 2$, input producers of a given good z in a given country $\ell(1) \in \mathcal{J}$ observe their productivity $1/a_{\ell(1)}^1(z)$, and simply hire labor and buy materials to minimize unit production costs, which results in $p_{\ell(1)}^1(z) = a_{\ell(1)}^1(z)\, c_{\ell(1)}$. Assemblers of good z in any country $\ell(2) \in \mathcal{J}$ observe their own productivity $1/a_{\ell(2)}^2(z)$, as well as that of all potential input producers worldwide, and solve

$$p_{\ell(2)}^2(z) = \min_{\ell(1)\in\mathcal{J}}\left\{\left(a_{\ell(2)}^2(z)\, c_{\ell(2)}\right)^{\alpha_2}\left(a_{\ell(1)}^1(z)\, c_{\ell(1)}\tau_{\ell(1)\ell(2)}\right)^{1-\alpha_2}\right\}.$$

Independently of the values of $a_{\ell(2)}^2(z)$, $c_{\ell(2)}$, and α_2, the solution of this problem simply entails procuring the input from the location $\ell^*(1)$ satisfying $\ell^*(1) = \arg\min\left\{\left(a_{\ell(1)}^1(z)\, c_{\ell(1)}\tau_{\ell(1)\ell(2)}\right)^{1-\alpha_2}\right\}$. As is well-known, the Fréchet assumption in (A.8) will make characterizing this problem fairly straightforward. Consider finally the problem of retailers in each country j seeking to procure a final good z to local consumers at a minimum cost. These retailers observe the productivity $1/a_{\ell(2)}^2(z)$ of all assemblers worldwide, but *not* the productivity of input producers, and thus seek to solve

$$p_j^F(z) = \min_{\ell(2)\in\mathcal{J}}\left\{\left(a_{\ell(2)}^2(z)\, c_{\ell(2)}\right)^{\alpha_2}\mathbb{E}\left[a_{\ell^*(1)}^1(z)\, c_{\ell^*(1)}\tau_{\ell^*(1)\ell(2)}\right]^{1-\alpha_2}\tau_{\ell(2)j}\right\}. \tag{A.9}$$

If retailers could observe the particular realizations of input producers, the expectation in (A.9) would be replaced by the realization of $a_{\ell(1)}^1(z)\, c_{\ell(1)}\tau_{\ell(1)\ell(2)}$ in all $\ell(1) \in \mathcal{J}$, and characterizing the optimal choice would be complicated because it would depend on the product of the distributions $a_{\ell(2)}^2(z)$ and $a_{\ell(1)}^1(z)$,which is not Fréchet under (A.8). Given our incomplete information assumption, however, the expectation in (A.9) does not depend on the particular realizations of upstream productivity draws, and this allows us to apply the well-know properties of the *univariate* Fréchet distribution in (A.8) to characterize the problem of retailers.

To see this, take two countries $\ell(1)$ and $\ell(2)$ and consider the probability $\pi_{\boldsymbol{\ell}j}$ of a GVC flowing through $\ell(1)$ and $\ell(2)$ before reaching consumers in j. This probability is simply the product of (i) the probability of $\ell(1)$ being the cost-minimizing location of input production conditional on assembly happening in $\ell(2)$, and (ii) the probability of $\ell(2)$ being the cost-minimizing location of assembly for GVC serving consumers in j.

Denoting $\mathcal{E}_{\ell(2)} = \mathbb{E}\left[\tau_{\ell^*(1)\ell(2)} a^1_{\ell^*(1)}(z)\, c_{\ell^*(1)}\right]^{1-\alpha_2}$, and using the properties of the Fréchet distribution, it is easy to verify that we can write $\pi_{\boldsymbol{\ell} j}$ as

$$\pi_{\boldsymbol{\ell} j} = \underbrace{\frac{\left(T_{\ell(1)}\right)^{1-\alpha_2}\left(c_{\ell(1)}\tau_{\ell(1)\ell(2)}\right)^{-\theta(1-\alpha_2)}}{\sum\limits_{k\in\mathcal{J}}\left(T_k\right)^{1-\alpha_2}\left(c_k\tau_{k\ell(2)}\right)^{-\theta(1-\alpha_2)}}}_{\Pr(\ell(1)|\ell(2))} \times \underbrace{\frac{\left(T_{\ell(2)}\right)^{\alpha_2}\left(\left(c_{\ell(2)}\right)^{\alpha_2}\tau_{\ell(2)j}\right)^{-\theta}\left(\mathcal{E}_{\ell(2)}\right)^{-\theta}}{\sum\limits_{i\in\mathcal{J}}\left(T_i\right)^{\alpha_2}\left(\left(c_i\right)^{\alpha_2}\left(\tau_{ij}\right)\right)^{-\theta}\left(\mathcal{E}_i\right)^{-\theta}}}_{\Pr(\ell(2))}. \tag{A.10}$$

A bit less trivially, but also exploiting well-known properties of the Fréchet distribution, it can be shown that

$$\mathcal{E}_{\ell(2)} = \mathbb{E}\left[\tau_{\ell^*(1)\ell(2)} a^1_{\ell^*(1)}(z)\, c_{\ell^*(1)}\right]^{1-\alpha_2} = \varsigma\left(\sum_{k\in\mathcal{J}}\left(T_k\right)^{1-\alpha_2}\left(c_k\tau_{k\ell(2)}\right)^{-\theta(1-\alpha_2)}\right)^{-1/\theta},$$

for some scalar $\varsigma > 0$. This allows us to reduce (A.10) to

$$\pi_{\boldsymbol{\ell} j} = \frac{\left(T_{\ell(1)}\right)^{1-\alpha_2}\left(c_{\ell(1)}\tau_{\ell(1)\ell(2)}\right)^{-\theta(1-\alpha_2)}\left(T_{\ell(2)}\right)^{\alpha_2}\left(\left(c_{\ell(2)}\right)^{\alpha_2}\tau_{\ell(2)j}\right)^{-\theta}}{\sum\limits_{k\in\mathcal{J}}\sum\limits_{i\in\mathcal{J}}\left(T_k\right)^{1-\alpha_2}\left(c_k\tau_{ki}\right)^{-\theta(1-\alpha_2)}\left(T_i\right)^{\alpha_2}\left(\left(c_i\right)^{\alpha_2}\left(\tau_{ij}\right)\right)^{-\theta}}. \tag{A.11}$$

It should be clear that this expression is identical to (8) – plugging in (9) – for the special case $N = 2$. It is also straightforward to verify that the distribution of final-good prices $p^F_j(\boldsymbol{\ell}, z)$ paid by consumers in j is independent of the actual path of production $\boldsymbol{\ell}$ and is again characterized, as in equation (7), by $\Pr\left(p^F_j(\boldsymbol{\ell}, z) \le p\right) = 1 - \exp\left\{-\tilde{\Theta}_j p^\theta\right\}$, where $\tilde{\Theta}_j$ is the denominator in (A.11), and is the analog of Θ_j in (9) when $N = 2$.

In sum, this alternative specification of the stochastic nature of technology delivers the exact same distribution of GVCs and of consumer prices as the one in which the overall GVC unit cost is distributed Fréchet.

We next generalize this result to an environment with more than two stages. It should be clear that the input sourcing decisions for the two most upstream stages work in the same way as outlined above. Let $\ell^j_z(n)$ be the tier-one sourcing decision of a firm producing good z at stage $n+1$ in j. Generalizing the notation above, define for any $s > 0$ the expectation

$$\mathcal{E}^n_j[s] = \mathbb{E}_n\left[\left(p^n_{\ell^j_z(n)}(z)\,\tau_{\ell^j_z(n)j}\right)^s\right],,$$

where we have written the expectation with an n subscript indicating that the expectation takes that unit costs (and prices) from stages $1,\ldots,n$ as unobserved. To be fully clear, a firm at $n+2$ observes the productivity draws from stage $n+1$ but does not know previous sourcing decisions. Hence it must form an expectation over the location from which its stage n suppliers source, $\ell^j_z(n)$, and use this to calculate the expected input prices $\mathcal{E}^n_j[s]$. As will become clear in the next paragraph, denoting the expectations for a general $s > 0$ is useful since downstream firms between $n+2,\ldots,N$ and final consumers will all use the information on expected input prices at n but in different ways depending on the objective function they seek to minimize.

Substituting in the Cobb-Douglas production process in (1), we can write

$$\mathcal{E}^n_j[s] = \mathbb{E}_n\left[\left(a^n_{\ell^j_z(n)}(z)\, c_{\ell^j_z(n)}\right)^{\alpha_n s} \times \mathcal{E}^{n-1}_{\ell^j_z(n)}\left[(1-\alpha_n)s\right] \times \left(\tau_{\ell^j_z(n)j}\right)^s\right].$$

The crucial observation is that to determine expected input prices from stage n a firm must also incorporate expected input prices from stage $n-1$, and so on until input prices from all upstream stages have been incorporated. Note that productivity draws across stages of production are independent, but even more importantly, sourcing decisions across stages of production are also independent. Hence, one can use the law of iterated expectations to compute expected input prices from $n-1$, $\mathcal{E}^{n-1}_{\ell^j_z(n)}[\cdot]$, in the computation of expected prices at n in $\mathcal{E}_j^n[\cdot]$. The latter expectation is over $\ell_z^j(n)$ but once we condition on a specific value for $\ell_z^j(n)$, the expectation $\mathcal{E}_{\ell_z^j(n)}^{n-1}[\cdot]$ is a constant. Finally, note also that this recursion starts at $n=1$ with $\mathcal{E}_j^0[s]=1$ since only labor and materials are used in that initial stage.

Let us next illustrate why these definitions are useful. Consider the optimal sourcing strategies related to procuring the good finished up to stage $n<N$. Given the sequential cost function in (1), the problem faced by a stage $n+1$ producer in j can be written as

$$\ell_z^j(n)=\arg\min_{\ell(n)\in\mathcal{J}}\left\{\left(a_{\ell(n)}^n(z)\,c_{\ell(n)}\right)^{\alpha_n(1-\alpha_{n+1})}\times\mathcal{E}_{\ell(n)}^{n-1}\left[(1-\alpha_n)(1-\alpha_{n+1})\right]\times\left(\tau_{\ell(n)j}\right)^{1-\alpha_{n+1}}\right\}.$$

where the $1-\alpha_{n+1}$ superscript comes from the stage $n+1$ producer wanting to minimize its own expected input price and in which the stage n input price enters its own unit cost to this power. Meanwhile, final consumers (or local retailers on their behalf) source their goods by solving

$$\ell_z^j(N)=\arg\min_{\ell(N)\in\mathcal{J}}\left\{\left(a_{\ell(N)}^N(z)\,c_{\ell(N)}\right)^{\alpha_N}\times\mathcal{E}_{\ell(N)}^{N-1}\left[1-\alpha_N\right]\times\tau_{\ell(N)j}\right\}.$$

The probability of sourcing inputs from a specific location i at any stage n can be determined by invoking the properties of the Fréchet distribution, given that $1/a_i^n(z)$ is drawn independently (across goods and stages) from a Fréchet distribution satisfying

$$\Pr\left(a_j^n(z)^{\alpha_n\beta_n}\geq a\right)=\exp\left\{-a^\theta\left(T_j\right)^{\alpha_n\beta_n}\right\}.$$

In particular, we obtain

$$\Pr\left(\ell_z^j(n)=i\right)=\frac{\left((T_i)^{\alpha_n}\left((c_i)^{\alpha_n}\tau_{ij}\right)^{-\theta}\right)^{\beta_n}\mathcal{E}_i^{n-1}\left[(1-\alpha_n)(1-\alpha_{n+1})\right]^{-\beta_{n+1}\theta}}{\sum_{l\in J}\left((T_l)^{\alpha_n}\left((c_l)^{\alpha_n}\tau_{lj}\right)^{-\theta}\right)^{\beta_n}\mathcal{E}_l^{n-1}\left[(1-\alpha_n)(1-\alpha_{n+1})\right]^{-\beta_{n+1}\theta}}.$$

These probabilities can now be leveraged in order to compute expected input prices. Define $\tilde{a}_{ij}=(c_i)^{\alpha_n s}\,\mathcal{E}_i^{n-1}\left[(1-\alpha_n)s\right](\tau_{ij})^s$ so that $1/\left(a_i^{\alpha_n s}\tilde{a}_{ij}\right)\sim$ Fréchet $\left(T_i^{\alpha_n\beta_n}\tilde{a}_{ij}^{-\frac{\beta_n}{s}\theta},\frac{\beta_n}{s}\theta\right)$ (note that the above distribution is the special case in which $s=1-\alpha_{n+1}$). Then using the moment generating formula for the Fréchet distribution, it can be verified that

$$\mathcal{E}_j^n[s]=q\left[\sum_{l\in J}T_l^{\alpha_n\beta_n}\tilde{a}_{lj}^{-\frac{\beta_n}{s}\theta}\right]^{-\frac{s}{\beta_n\theta}}\Gamma\left(1+\frac{\beta_n}{s}\theta\right),$$

where Γ is the gamma function. From this equation it should also be clear why we are denoting $E_j^n[s]$ as a function of s, since as we move down the value chain we need to compute the upstream expectations at different 'moments'.

We are now ready to determine the equilibrium variables: (1) material prices P_j and (2) the distribution of GVCs. Material prices can be derived recursively using our expectations:

$$\begin{aligned}P_j=\left(\mathcal{E}_j^N[1-\sigma]\right)^{\frac{1}{1-\sigma}}&=\left[\sum_{l\in\mathcal{J}}(T_l)^{\alpha_N}\left((c_l)^{\alpha_N}\tau_{lj}\right)^{-\theta}\mathcal{E}_l^{N-1}\left[(1-\alpha_N)(1-\sigma)\right]^{-\frac{\theta}{1-\sigma}}\right]^{-\frac{1}{\theta}}\Gamma\left(1+\frac{1-\sigma}{\theta}\right)\\&=\varsigma\left[\sum_{\boldsymbol{\ell}\in\mathcal{J}}\prod_{n=1}^{N}\left(\left(T_{\ell(n)}\right)^{\alpha_n}\left(\left(c_{\ell(n)}\right)^{\alpha_n}\tau_{\ell(n)\ell(n+1)}\right)^{-\theta}\right)^{\beta_n}\right]^{-\frac{1}{\theta}},\end{aligned}$$

where $\varsigma = \prod_{n=1}^{N} \Gamma\left(1+\frac{1-\sigma}{\beta_n \theta}\right)^{\frac{1}{1-\sigma}}$. This expression is identical to (10) up to a scalar (which is irrelevant for all equilibrium conditions and that could be 'neutralized' by an appropriate rescaling of the stage-specific Fréchet distributions).

Finally, since input decisions from n are independent from the decisions that firms at $n-1$ made then

$$
\begin{aligned}
\pi_{\boldsymbol{\ell} j} &= \Pr\left(\ell_z^j(N) = \ell(N) \,\middle|\, \ell_z^{\ell(N)}(N-1) = \ell(N-1)\right) \times \\
&\quad \times \prod_{n=2}^{N-1} \Pr\left(\ell_z^{\ell(n+1)}(n) = \ell(n) \,\middle|\, \ell_z^{\ell(n)}(n-1) = \ell(n-1)\right) \times \Pr\left(\ell_z^{\ell(2)}(1) = \ell(1)\right) \\
&= \Pr\left(\ell_z^j(N) = \ell(N)\right) \times \prod_{n=1}^{N} \Pr\left(\ell_z^{\ell(n+1)}(n) = \ell(n)\right) \\
&= \frac{\prod_{n=1}^{N-1}\left(\left(T_{\ell(n)}\right)^{\alpha_n}\left(\left(c_{\ell(n)}\right)^{\alpha_n}\tau_{\ell(n)\ell(n+1)}\right)^{-\theta}\right)^{\beta_n} \times \left(T_{\ell(N)}\right)^{\alpha_N}\left(\left(c_{\ell(N)}\right)^{\alpha_N}\tau_{\ell(N)j}\right)^{-\theta}}{\sum_{\boldsymbol{\ell}' \in \mathcal{J}} \prod_{n=1}^{N-1}\left(\left(T_{\ell'(n)}\right)^{\alpha_n}\left(\left(c_{\ell'(n)}\right)^{\alpha_n}\tau_{\ell'(n)\ell'(n+1)}\right)^{-\theta}\right)^{\beta_n} \times \left(T_{\ell'(N)}\right)^{\alpha_N}\left(\left(c_{\ell'(N)}\right)^{\alpha_N}\tau_{\ell'(N)j}\right)^{-\theta}}, \quad \text{(A.12)}
\end{aligned}
$$

which is identical to equation (8) in the main text obtained in the 'randomness-in-the-chain' formulation of technology.

A.2.2 Oberfield Approach

We next turn to the second decentralized approach inspired by work of Oberfield (2018). To ease the notation, let us define

$$
Z_{\ell(n)}^n = \left(a_{\ell(n)}^n\right)^{-\alpha_n},
$$

so that we can write equation (1) as

$$
p_{\ell(n)}^n = \frac{1}{Z_{\ell(n)}^n}\left(c_{\ell(n)}\right)^{\alpha_n}\left(p_{\ell(n-1)}^{n-1}\tau_{\ell(n-1)\ell(n)}\right)^{1-\alpha_n}.
$$

A key conceptual difference with this approach is that the efficiency level $Z_{\ell(n)}^n$ is now assumed to be buyer-seller specific (or *match* specific). In particular, a firm producing stage n in location $\ell(n)$ meets a certain number of potential sellers of stage $n-1$ in each location $\ell(n-1)$, with each of these potential sellers being associated with a distinct 'match' productivity of combining the good completed up to stage $n-1$ with the labor and materials at stage n. This buyer-seller specific productivity is drawn from a Pareto distribution with shape parameter θ and lower bound $\underline{Z}_{\ell(n)}^n$. Below, we will focus on the limiting case in which $\underline{Z}_{\ell(n)}^n \to 0$. Given all the available match-specific productivities and production costs, each stage-n producer (or buyer) chooses the supplier offering the lowest cost for the good produced at stage $n-1$. The number of available potential suppliers in each sourcing country $\ell(n-1)$ varies across producers, and the precise number $m_{\ell(n-1)\ell(n)}^n$ of potential suppliers based in country $\ell(n-1)$ available to a given firm producing stage n in country $\ell(n)$ is assumed to follow a Poisson distribution with arrival rate $\left(T_{\ell(n)}\right)^{\alpha_n}\left(\underline{Z}_{\ell(n)}^n\right)^{-\theta}$. For $n=1$, and for the time being, we assume that productivity in location $\ell(1)$ is fixed at $Z_{\ell(1)}^1 = \left(T_{\ell(1)}\right)^{1/\theta}$, though we will relax this assumption below.

We now derive the distribution of final good prices in country j when sourcing goods through an arbitrary supply chain $\boldsymbol{\ell}$. To build intuition, let us first study the case with two stages ($N=2$). Consider the distribution of prices that a stage 2 producer in country $\ell(2)$ can offer to consumers in country j if stage 1 output is bought from country $\ell(1)$ and the highest matched-pair productivity with suppliers in that country is $\hat{Z}_{\ell(2)}^2$. This distribution is given by

$$\begin{aligned} G_j^2\left(p\,|\ell(1),\ell(2)\right) &= \Pr\left(p \leq \frac{1}{\hat{Z}^2_{\ell(2)}}\left(c_{\ell(2)}\right)^{\alpha_2}\left(p^1_{\ell(1)}\tau_{\ell(1)\ell(2)}\right)^{1-\alpha_2}\tau_{\ell(2)j}\right), \\ &= \Pr\left(\hat{Z}^2_{\ell(2)} \leq \tilde{Z}(p)\right), \end{aligned}$$

where $\tilde{Z}(p) = \left(c_{\ell(2)}\right)^{\alpha_2}\left(c_{\ell(1)}\left(T_{\ell(1)}\right)^{-1/\theta}\tau_{\ell(1)\ell(2)}\right)^{1-\alpha_2}\tau_{\ell(2)j}/p$.

Now remember that the stage-2 producer has various potential suppliers in each country $\ell(1)$, so for the price to be higher than p, or for $\max_{\mu=1,\ldots m_{\ell(1)\ell(2)}}\left\{Z^{2,\mu}_{\ell(2)}\right\} = \hat{Z}^2_{\ell(2)} < \tilde{Z}(p)$, we need $Z^{2,m}_{\ell(2)} < \tilde{Z}(p)$ for all the draws μ associated with all the potential suppliers $m_{\ell(1)\ell(2)}$ that a specific firm has. Since both the number of suppliers and productivity of each set of suppliers is stochastic, we can obtain the overall distribution of prices invoking the formula for the Poisson probability density function and also plugging in the cumulative density function for the Pareto distribution:

$$\begin{aligned} &G_j\left(p\,|\ell(1),\ell(2)\right) \\ &= \sum_{m=0}^{\infty}\prod_{\mu=1}^{m}\Pr\left(Z^{2,\mu}_{\ell(2)} \leq \tilde{Z}(p)\right)\times\Pr\left(m_{\ell(1)\ell(2)} = m\right), \\ &= \sum_{m=0}^{\infty}\prod_{\mu=1}^{m}\left(1-\left(\frac{\underline{Z}^2_{\ell(2)}}{\tilde{Z}(p)}\right)^{\theta}\right)\times\frac{\left(\left(T_{\ell(2)}\right)^{\alpha_2}\left(\underline{Z}^2_{\ell(2)}\right)^{-\theta}\right)^m\exp\left\{-\left(T_{\ell(2)}\right)^{\alpha_2}\left(\underline{Z}^2_{\ell(2)}\right)^{-\theta}\right\}}{m!}, \\ &= \exp\left\{-p^{\theta}\left(T_{\ell(1)}c^{-\theta}_{\ell(1)}\tau^{-\theta}_{\ell(1)\ell(2)}\right)^{1-\alpha_2}\left(T_{\ell(2)}c^{-\theta}_{\ell(2)}\right)^{\alpha_2}\tau^{-\theta}_{\ell(2)j}\right\}. \end{aligned} \tag{A.13}$$

This is the same expression we obtain in the "Fréchet-in-the-chain" formulation in the main text.

Now let us extend these results to the case with $N = 3$, and consider the problem of producers of the final assembly stage $n = 3$ in country $\ell(3)$. For such a producer, the distribution of prices it can offer to consumers in j when stage-2 inputs are bought from country $\ell(2)$ is more involved than before because it now depends on the product of the distribution of buyer-seller productivity draws $Z^3_{\ell(3)}$ and the upstream input prices $p^2_{\ell(2)}$ that each input seller itself sells at (that is, influenced by the buyer-seller productivity that the stage-2 seller has with its own input suppliers). However, note that the buyer-seller productivity draws at stage-3 are independent of the upstream productivity draws. Instead, what is crucial to take into account is the fact that stage-3 producers that get more stage-2 matches will get, on average, both a better buyer-seller productivity draw but also a better stage-2 input price. Thus, we can split the problem into two parts. We first obtain the expected price distribution conditional on a buyer-seller relationship and then we obtain the price distribution by characterizing the distribution of optimal matches.

Define the distribution of stage-3 prices in j from a given supply chain $\boldsymbol{\ell} = \{\ell(1),\ell(2),\ell(3)\}$ conditional on a specific buyer-seller relationship characterized by $Z^3_{\ell(3)}$ as

$$\begin{aligned} F_j\left(p\left|\boldsymbol{\ell}, Z^3_{\ell(3)}\right.\right) &= \Pr\left(p \leq \frac{1}{Z^3_{\ell(3)}}\left(c_{\ell(3)}\right)^{\alpha_3}\left(p^2_{\ell(2)}\left(\boldsymbol{\ell}\right)\tau_{\ell(2)\ell(3)}\right)^{1-\alpha_3}\tau_{\ell(3)j}\right), \\ &= \exp\left\{-\Theta_{\ell(1)\ell(2)}\left(\frac{pZ^3_{\ell(3)}}{\tau^{1-\alpha_3}_{\ell(2)\ell(3)}c^{\alpha_3}_{\ell(3)}\tau_{\ell(3)j}}\right)^{\theta/(1-\alpha_3)}\right\}, \end{aligned}$$

where $\Theta_{\ell(1)\ell(2)} = \left(T_{\ell(1)}c^{-\theta}_{\ell(1)}\tau^{-\theta}_{\ell(1)\ell(2)}\right)^{1-\alpha_2}\left(T_{\ell(2)}c^{-\theta}_{\ell(2)}\right)^{\alpha_2}$ and where we have invoked our above distribution

(A.13). As in the $N = 2$ stage case, with $N = 3$ the distribution of prices along chain $\boldsymbol{\ell}$ will be determined by the fact that each producer at the assembly stage chooses the upstream supplier that offers the best combination of buyer-seller productivity and input prices. That is

$$\begin{aligned}
&G_j\left(p\left|\ell\left(1\right),\ell\left(2\right),\ell\left(3\right)\right.\right)\\
&=\sum_{m=0}^{\infty}\prod_{\mu=1}^{m}\int_{\underline{Z}^3_{\ell(3)}}^{\infty}F_j\left(p\left|\boldsymbol{\ell},Z^{3,\mu}_{\ell(3)}\right.\right)\Pr\left(Z^{3,\mu}_{\ell(3)}=Z\right)dZ\times\Pr\left(m_{\ell(2)\ell(3)}=m\right),\\
&=\exp\left\{-\left(T_{\ell(3)}\right)^{\alpha_3}\left(\underline{Z}^3_{\ell(3)}-\int_{\underline{Z}^3_{\ell(3)}}^{\infty}F_j\left(p\left|\boldsymbol{\ell},Z\right.\right)\frac{\theta}{Z^{\theta+1}}dZ\right)\right\},
\end{aligned}$$

where we used the fact that $Z^{3,\mu}_{\ell(3)}$ is a Pareto random variable with lower bound $\underline{Z}^3_{\ell(3)}$ and shape parameter θ. Now, define $\chi\left(p\right)$ such that $F_j\left(p\left|\boldsymbol{\ell},Z^3_{\ell(3)}\right.\right)=\exp\left\{-\chi\left(p\right)\left(Z^3_{\ell(3)}\right)^{\theta/(1-\alpha_3)}\right\}$, and solve the above integral by taking the limit when $\underline{Z}^3_{\ell(3)}\to 0$ and using a change of variable $\zeta\left(p\right)=\chi\left(p\right)Z^{\theta/(1-\alpha_3)}$ to obtain

$$\begin{aligned}
\int_0^{\infty}F_j\left(p\left|\boldsymbol{\ell},Z\right.\right)\frac{\theta}{Z^{\theta+1}}dZ&=\frac{\chi\left(p\right)^{1-\alpha_3}}{1-\alpha_3}\int_0^{\infty}\exp\left\{-\zeta\left(p\right)\right\}\zeta\left(p\right)^{\alpha_3-1}d\zeta\left(p\right),\\
&=\frac{\chi\left(p\right)^{1-\alpha_3}}{1-\alpha_3}\Gamma\left(\alpha_3\right),
\end{aligned}$$

where $\Gamma\left(\cdot\right)$ is the gamma function. Plugging this back in (and remember that we took the limit $\underline{Z}^3_{\ell(3)}\to 0$) we obtain that

$$\begin{aligned}
&G_j\left(p\left|\ell\left(1\right),\ell\left(2\right),\ell\left(3\right)\right.\right)\\
&=\exp\left\{-\left(T_{\ell(3)}\right)^{\alpha_3}\frac{\chi\left(p\right)^{1-\alpha_3}}{1-\alpha_3}\Gamma\left(\alpha_3\right)\right\},\\
&=\exp\left\{-p^{\theta}\times\prod_{n=1}^{3}\left(c^{-\theta}_{\ell(n)}T_{\ell(n)}\right)^{\alpha_n\beta_n}\times\prod_{n=1}^{2}\left(\tau_{\ell(n)\ell(n+1)}\right)^{-\theta\beta_n}\times\left(\tau_{\ell(3)j}\right)^{-\theta}\times\frac{\Gamma\left(\alpha_3\right)}{1-\alpha_3}\right\}.
\end{aligned}$$

where notation is such that $\beta_n\equiv\prod\limits_{m=n+1}^{N}\left(1-\alpha_m\right)$ and $\alpha_1=1$. This last expression is the exact same expression we obtain in the main text for $\Pr\left(p^F_j\left(\boldsymbol{\ell},z\right)\geq p\right)$ in the $N=3$ case except for the last scalar term involving the gamma function term. Nevertheless, this scalar term is irrelevant for the main equilibrium conditions in the model.

We have derived this result for stages $n=3$ and $n=2$, but it should be clear that the above derivations would work for any two stages n and $n-1$, as long as the distribution of production costs in upstream stage $n-1$ is Fréchet distributed. This has two implications. First, our assumption above that, for $n=1$, productivity in location $\ell\left(1\right)$ is fixed at $Z^1_{\ell(1)}=\left(T_{\ell(1)}\right)^{1/\theta}$ can be relaxed and we can instead assume that $Z^1_{\ell(1)}$ is Fréchet distributed with shape parameter θ and scale parameter $T_{\ell(1)}$. Second, one can use induction to conclude from our results above that, for a general N, we obtain

$$\Pr\left(p^F_j\left(\boldsymbol{\ell},z\right)\geq p\right)=\exp\left\{-p^{\theta}\times\prod_{n=1}^{N}\left(c^{-\theta}_{\ell(n)}T_{\ell(n)}\right)^{\alpha_n\beta_n}\times\prod_{n=1}^{N-1}\left(\tau_{\ell(n)\ell(n+1)}\right)^{-\theta\beta_n}\times\left(\tau_{\ell(N)j}\right)^{-\theta}\times\tilde{\varsigma}\right\},$$

where $\tilde{\varsigma}$ is a positive scalar that is irrelevant for all equilibrium conditions and that can be 'neutralized' by an appropriate rescaling of the stage-specific Poisson distributions. It should be apparent that this expression coincides with equation (7) in the main text, up to this immaterial scalar $\tilde{\zeta}$.

Finally, it remains to be show that this decentralized solution not only delivers the same distribution of final-good prices, but also the same GVC trade shares as in expression (8) in the main text. But this is implied by our previous derivations related to the decentralized approach with incomplete information. In particular, fixing a downstream stage n, the distribution of upstream costs at $n-1$ is again Fréchet distributed, so applying the law of total probability in the same manner as in (A.12) above, it is straightforward to re-derive equation (8) in the main text. And, to reiterate, the scalar $\tilde{\zeta}$ is irrelevant for these equilibrium conditions.

A.3 Introducing Trade Deficits

Let D_j be country j's aggregate deficit in dollars, where $\sum_j D_j = 0$ holds since global trade is balanced. The only difference in the model's equations is that the general equilibrium equation is given by

$$\frac{1}{\gamma_i} w_i L_i = \sum_{j \in \mathcal{J}} \sum_{n \in \mathcal{N}} \alpha_n \beta_n \times \Pr\left(\Lambda_i^n, j\right) \times \left(\frac{1-\gamma_j}{\gamma_j} w_j L_j + w_j L_j - D_j \right).$$

where $w_j L_j - D_j$ is aggregate final good consumption in country j.

A.4 Graphical Description of Multi-Stage Production

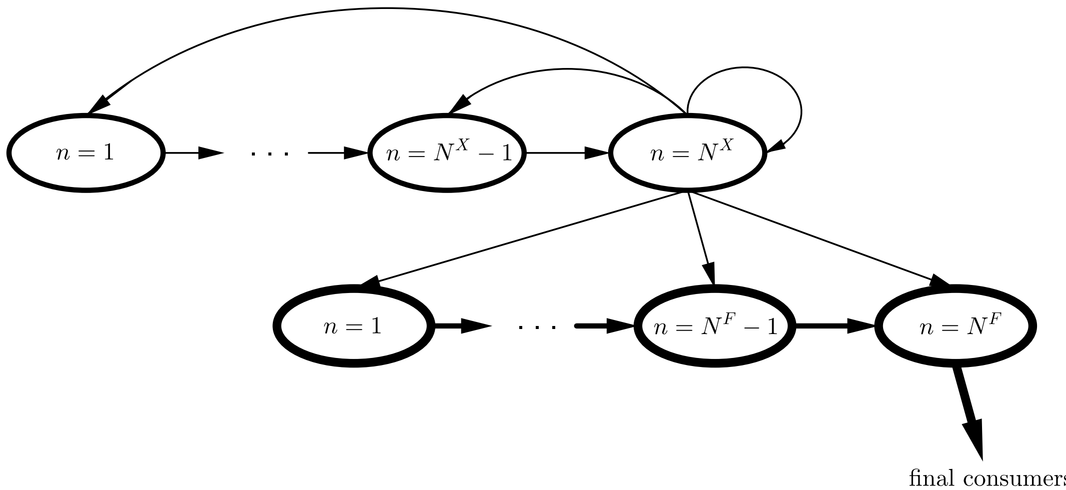

Figure A.2: Multi-stage production with separate intermediate input and final good supply chains.

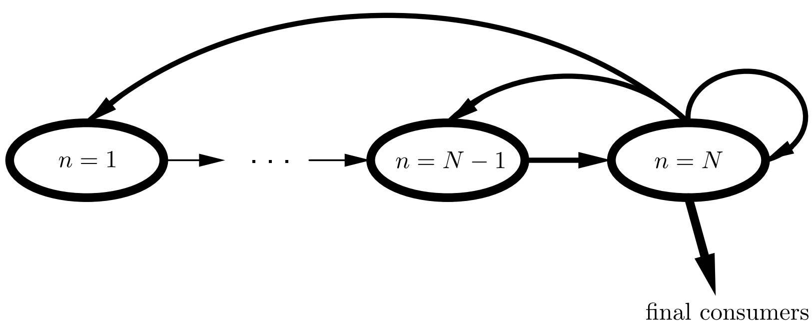

Figure A.3: Multi-stage production with common intermediate input and final good supply chains.

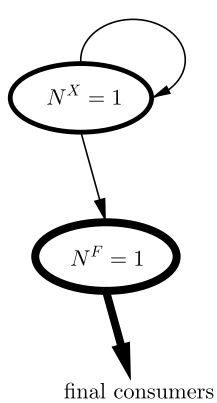

Figure A.4: Single-stage production with separate intermediate input and final good technology (Alexander 2017).

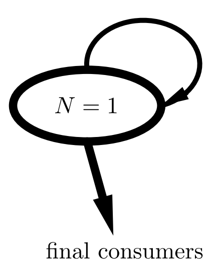

Figure A.5: Single-stage production with common intermediate input and final good technology (Eaton and Kortum 2002).

A.5 Estimation Results

Table A.1: Estimation Results - Asymmetric Parameterizations.

	γ_j					T_j^X					T_j^F				
N	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
AUS	0.54	0.65	0.74	0.79	0.83	7.96	7.97	8.03	8.13	8.29	3.52	4.88	5.01	4.96	4.86
AUT	0.53	0.63	0.73	0.78	0.83	1.42	1.04	0.72	0.55	0.43	2.64	2.07	1.74	1.53	1.40
BEL	0.53	0.62	0.72	0.79	0.84	1.67	0.95	0.65	0.48	0.36	2.01	1.64	1.45	1.33	1.26
BGR	0.36	0.48	0.58	0.65	0.70	0.02	0.01	0.00	0.00	0.00	0.03	0.02	0.01	0.00	0.00
BRA	0.56	0.68	0.76	0.81	0.84	0.10	0.08	0.04	0.03	0.02	0.37	0.12	0.06	0.04	0.03
CAN	0.58	0.68	0.77	0.83	0.87	6.28	5.35	4.59	4.23	4.03	3.44	3.40	3.25	3.08	2.94
CHE	0.53	0.62	0.72	0.78	0.82	9.21	8.00	7.85	7.46	6.92	10.9	12.4	14.5	16.5	17.5
CHN	0.33	0.45	0.55	0.62	0.67	0.16	0.13	0.07	0.05	0.04	0.35	0.12	0.08	0.05	0.04
CZE	0.44	0.53	0.64	0.71	0.76	0.15	0.06	0.03	0.01	0.01	0.22	0.13	0.08	0.05	0.04
DEU	0.54	0.65	0.74	0.80	0.83	3.10	3.16	2.40	1.98	1.66	5.57	4.95	4.69	4.47	4.29
DNK	0.57	0.64	0.73	0.79	0.83	3.01	1.55	1.14	0.86	0.66	5.39	5.08	4.42	4.03	3.76
ESP	0.52	0.63	0.72	0.77	0.80	0.57	0.44	0.27	0.20	0.16	1.25	0.78	0.56	0.44	0.38
FIN	0.51	0.60	0.69	0.74	0.78	1.29	0.70	0.51	0.39	0.32	1.99	2.08	1.61	1.35	1.19
FRA	0.55	0.66	0.74	0.79	0.82	1.87	1.93	1.39	1.12	0.94	4.07	3.07	2.59	2.28	2.07
GBR	0.56	0.67	0.75	0.80	0.84	3.49	3.29	2.59	2.23	2.01	3.32	3.13	2.81	2.57	2.39
GRC	0.58	0.66	0.74	0.78	0.81	0.08	0.03	0.01	0.01	0.01	0.24	0.14	0.07	0.05	0.03
HRV	0.46	0.57	0.68	0.74	0.78	0.02	0.01	0.00	0.00	0.00	0.03	0.03	0.01	0.01	0.00
HUN	0.52	0.59	0.70	0.78	0.83	0.05	0.01	0.00	0.00	0.00	0.09	0.05	0.03	0.01	0.01
IDN	0.53	0.65	0.73	0.79	0.82	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
IND	0.53	0.65	0.73	0.77	0.80	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
IRL	0.62	0.67	0.79	0.88	0.95	3.03	0.93	0.56	0.33	0.19	2.97	2.29	2.20	2.22	2.31
ITA	0.51	0.62	0.71	0.76	0.80	0.91	0.82	0.56	0.45	0.38	1.82	1.28	1.01	0.85	0.75
JPN	0.52	0.64	0.72	0.77	0.80	1.32	1.88	1.39	1.16	1.02	6.56	3.66	3.11	2.73	2.47
KOR	0.42	0.53	0.63	0.69	0.74	0.55	0.50	0.34	0.26	0.21	1.72	0.93	0.76	0.64	0.56
LTU	0.46	0.57	0.69	0.76	0.81	0.03	0.01	0.00	0.00	0.00	0.02	0.04	0.02	0.01	0.01
LUX	0.32	0.51	0.67	0.79	0.88	0.96	2.65	3.89	4.76	5.19	0.18	1.61	2.80	4.28	5.90
MEX	0.59	0.70	0.77	0.81	0.84	0.05	0.03	0.01	0.01	0.01	0.38	0.11	0.06	0.04	0.03
NLD	0.60	0.69	0.80	0.88	0.92	5.95	3.74	3.18	2.85	2.62	2.89	2.83	2.83	2.81	2.80
NOR	0.62	0.73	0.83	0.88	0.91	31.8	32.5	40.1	43.7	46.1	14.3	21.2	22.5	23.2	23.2
POL	0.48	0.59	0.68	0.74	0.78	0.20	0.10	0.05	0.03	0.02	0.32	0.18	0.11	0.08	0.06
PRT	0.54	0.63	0.71	0.77	0.80	0.13	0.06	0.03	0.02	0.01	0.27	0.15	0.08	0.05	0.04
ROU	0.49	0.59	0.68	0.74	0.78	0.04	0.02	0.01	0.00	0.00	0.07	0.03	0.02	0.01	0.01
ROW	0.44	0.57	0.67	0.73	0.77	0.06	0.03	0.01	0.01	0.00	0.06	0.02	0.01	0.01	0.00
RUS	0.55	0.69	0.79	0.85	0.89	0.63	0.45	0.29	0.23	0.20	0.06	0.04	0.03	0.02	0.02
SVK	0.44	0.52	0.64	0.72	0.77	0.12	0.04	0.02	0.01	0.00	0.20	0.15	0.08	0.06	0.04
SVN	0.38	0.53	0.65	0.73	0.79	0.05	0.03	0.01	0.00	0.00	0.06	0.08	0.04	0.02	0.02
SWE	0.55	0.65	0.74	0.80	0.83	3.64	2.43	1.98	1.66	1.41	4.30	4.39	3.98	3.71	3.52
TUR	0.53	0.63	0.71	0.76	0.80	0.09	0.04	0.02	0.01	0.01	0.15	0.08	0.04	0.03	0.02
TWN	0.55	0.64	0.75	0.83	0.88	0.79	0.35	0.22	0.17	0.14	0.31	0.24	0.17	0.13	0.11
USA	0.57	0.69	0.77	0.82	0.85	9.15	18.7	17.0	16.6	16.6	17.8	16.5	16.7	16.4	15.9

Table A.2: Estimation Results - Symmetric Parameterizations.

	γ_j		T_j	
N	1	2	1	2
AUS	0.52	0.88	4.79	3.48
AUT	0.55	0.87	2.14	0.57
BEL	0.54	0.83	1.92	0.45
BGR	0.61	0.95	0.10	0.00
BRA	0.57	0.99	0.15	0.01
CAN	0.55	0.92	3.43	1.65
CHE	0.52	0.81	9.44	6.35
CHN	0.33	0.57	0.18	0.03
CZE	0.48	0.73	0.23	0.02
DEU	0.55	0.87	3.90	1.65
DNK	0.59	0.92	5.04	1.83
ESP	0.54	0.88	0.77	0.14
FIN	0.54	0.87	2.01	0.59
FRA	0.56	0.93	2.58	0.93
GBR	0.55	0.91	2.95	1.24
GRC	0.63	1.00	0.16	0.01
HRV	0.70	1.00	0.22	0.01
HUN	0.61	0.91	0.14	0.01
IDN	0.55	0.93	0.01	0.00
IND	0.56	0.97	0.00	0.00
IRL	0.63	0.92	3.89	0.94
ITA	0.51	0.85	1.15	0.29
JPN	0.54	0.93	2.43	1.08
KOR	0.44	0.71	0.79	0.21
LTU	0.71	1.00	0.58	0.10
LUX	0.45	0.81	8.02	42.5
MEX	0.64	1.00	0.11	0.00
NLD	0.56	0.86	2.83	0.81
NOR	0.61	0.98	22.3	17.8
POL	0.50	0.79	0.24	0.02
PRT	0.57	0.93	0.22	0.01
ROU	0.53	0.84	0.06	0.00
ROW	0.43	0.72	0.05	0.00
RUS	0.50	0.84	0.22	0.02
SVK	0.56	0.85	0.47	0.05
SVN	0.62	1.00	0.68	0.25
SWE	0.56	0.88	4.04	1.50
TUR	0.54	0.88	0.11	0.01
TWN	0.51	0.80	0.43	0.06
USA	0.58	1.00	11.2	15.3

B Online Appendix for "On the Geography of Global Value Chains" (Not for Typeset Publication Online)

B.1 Partial Equilibrium Model: An Example

In this Appendix, we illustrate some of the salient and distinctive features of our partial model of sequential production in section 2 via a simple example. We consider a world with four countries ($J = 4$) and four stages ($N = 4$). Technology is given by a symmetric Cobb-Douglas specification with $\alpha_n \beta_n = 1/4$ for all n. The four countries are divided into two regions, the West (comprising countries A and B) and the East (comprising countries C and D). The 'geography' of this example is illustrated in Figure B.1. Note that we impose a great deal of symmetry: intra-regional trade costs are common in both regions, and inter-regional costs between A and C are identical to those between B and D. On the other hand, trade costs between B and C are lower than between A and D. For simplicity, all domestic trade costs are set to 0, so $\tau_{ii} = 1$ for $i = A, B, C, D$. We are interested in solving for the optimal path of a four-stage production process leading to consumption in country D (in green in the figure). Note that shipping to D directly is least costly when shipping from D itself, followed by C (the other country in the East), then by A and finally by B, which is the most remote country relative to D.

We compute the optimal path leading to D for different levels of trade costs starting with a benchmark with $\tau_{AB} = \tau_{CD} = 1.3$, $\tau_{BC} = 1.5$, $\tau_{AD} = 1.75$, $\tau_{AC} = \tau_{BD} = 1.8$, and then scale these international trade costs up or down by a shifter s (so starting from τ_{ij}, we instead use $\tilde{\tau}_{ij}(s) = 1 + s \times (\tau_{ij} - 1)$).[45] For each matrix of trade costs, we run one million simulations with production costs $a_j^n c_j$ being drawn independently for each stage n and each country j from a lognormal distribution with mean 0 and variance 1. By choosing a common distribution across countries and stages, we seek to isolate the role of trade costs in shaping the optimal path of sequential value chains.

The results of these simulations are depicted in Figure B.2 for various levels of s ranging from 0 (free trade) to 50 (which results in close to prohibitive trade costs). The upper left panel shows the average propensity of each country to appear in GVCs leading to consumption in D. The upper right panel depicts the average position (or downstreamness) of countries in these GVCs. Finally, the lower panel decomposes GVCs into purely domestic ones (with all production stages in D), purely regional ones (with some stages in C and D, but not in A or B) and global ones (involving at least one stage in A or B).

Several aspects of Figure B.2 are worth highlighting. First, focusing on the upper left panel, notice that country B, which is farthest away from country D, appears slightly more often in value chains leading to D than its Western neighbor A does. The reason for this surprising fact is tightly related to the sequential nature of production. Even though, A is closer to D than B is, B is relatively close to D's Eastern neighbor C, and this makes this 'remote' country B a particularly appealing location from which to set off value chains that will flow to D through C.[46] A second noteworthy aspect, apparent from the upper right panel of Figure B.2, is that remoteness appears to shape the average position of a country in GVCs, a fact we anticipated above. More specifically, country B, which is farthest away from D, is on average the most upstream of all countries, followed by its Western neighbor A, and then by C, with D being naturally the country positioned most downstream in value chains leading to consumption in D. Finally, the lower panel of Figure B.2 illustrates

[45] These parameters are chosen such that for all values of s considered, the triangle inequality holds for any three given countries.

[46] As we show in Online Appendix B.1 of the working paper version of our paper (Antràs and de Gortari, 2017), in an analogous world without sequentiality, the above pattern would not hold and the relative prevalence of countries would be strictly monotonic in the level trade costs incurred when shipping to the assembly location.

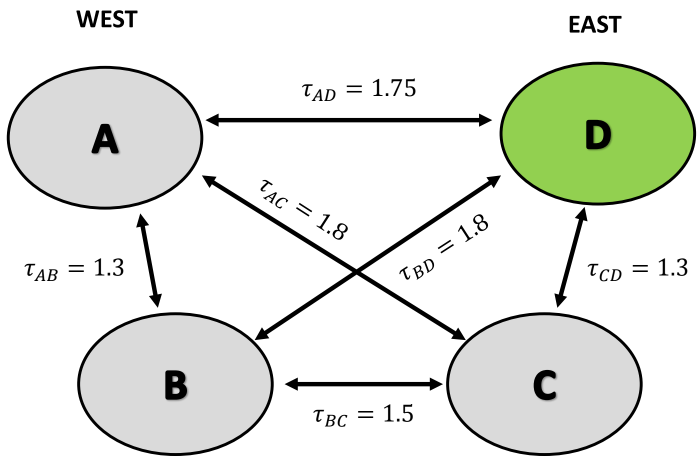

Figure B.1: An Example with Four Countries

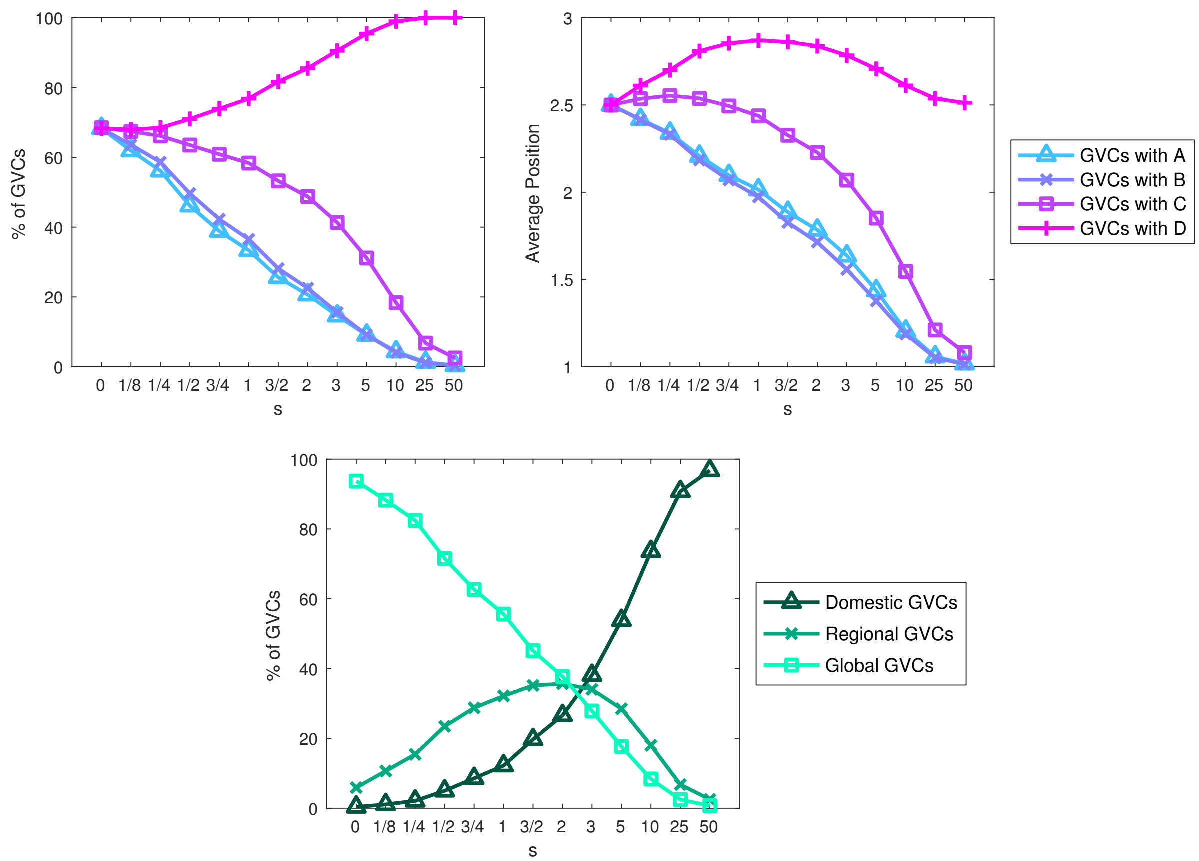

Figure B.2: Some Features of Optimal Production Paths how the progressive reduction of international trade costs first gives rise to GVCs that are largely regional in nature, and then later to truly global value chains involving inter-regional trade. It is also worth highlighting that even for fairly low trade costs, purely domestic GVCs remain quite prevalent, much more so than would be predicted by an analogous model without sequentiality (see the Online Appendix B.1 in Antràs and de Gortari, 2017). The reason for this is the compounding effect of trade costs, which other things equal makes it costly to offshore intermediate production stages in chains in which D has comparative advantage in the most upstream *and* downstream stages.

B.2 Proof of Existence and Uniqueness

The aim of this Appendix is to study the existence and uniqueness of the general equilibrium of our model. Because the core of our estimation focuses on a one-sector version of our model with variation in value added shares γ_i across countries, we develop our proofs for that more general case. Let us begin with some assumptions and definitions.

We shall assume throughout the following:

1. $\forall i \in J$: $\gamma_i \in (0,1]$.
2. $\sum_{n \in N} \alpha_n \beta_n = 1$.
3. There exist lower ($T_{\min}$, $\tau_{\min}$) and upper ($T_{\max}$, $\tau_{\max}$) bounds on τ_{ij} $\forall \{i,j\} \in \mathcal{J}^2$ and T_j $\forall j \in \mathcal{J}$.

Definition 6 ***(M-matrix)*** *An $n \times n$ matrix A is an M-matrix if the following* ***equivalent*** *statements hold:*

(i) A can be represented as $sI - B$, where I is $n \times n$ identity matrix, $s \in R_{++}$ is a constant and B is the matrix with positive elements and the moduli of B's eigenvalues are all $\leq s$.

(ii) A has a non-negative inverse.

Definition 7 ***(Excess demand)*** *The excess demand function $\mathbf{Z}(\mathbf{w})$ is defined as*

$$Z_i(\mathbf{w}) = \frac{1}{w_i}\left(\sum_{j \in \mathcal{J}} \sum_{n \in \mathcal{N}} \alpha_n \beta_n \times \Pr(\Lambda_i^n, j) \times \frac{1}{\gamma_j} w_j L_j\right) - \frac{1}{\gamma_i} L_i, \tag{B.1}$$

with $\Pr(\Lambda_i^n, j) = \sum_{\boldsymbol{\ell} \in \Lambda_i^n} \pi_{\boldsymbol{\ell} j}$, and where remember that $\Lambda_i^n = \left\{\boldsymbol{\ell} \in \mathcal{J}^N \mid \ell(n) = i\right\}$.

Definition 8 ***(Gross Substitutes)*** *The function $\mathbf{F}(\mathbf{w}) : \mathbb{R}^J \to \mathbb{R}^J$ has the gross substitutes property in $\mathbf{w}$ if*

$$\forall \{i,j\} \in \mathcal{J}^2, i \neq j : \qquad \frac{\partial F_i}{\partial w_j} > 0.$$

We next use these assumptions and definitions to develop proofs of existence and uniqueness that parallel those of Theorems 1-3 in Alvarez and Lucas (2007).

Theorem 1 *For any $\mathbf{w} \in \mathbb{R}_{++}^J$ there is a unique $\mathbf{p}^*(\mathbf{w})$ that solves, for all $j \in J$*

$$P_j = \kappa \left(\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \prod_{n=1}^{N} \left(\left(\left(w_{\ell(n)}\right)^{\gamma_{\ell(n)}} \left(P_{\ell(n)}\right)^{1-\gamma_{\ell(n)}} \right)^{-\theta} T_{\ell(n)} \right)^{\alpha_n \beta_n} \times \prod_{n=1}^{N-1} \left(\tau_{\ell(n)\ell(n+1)}\right)^{-\theta \beta_n} \times \left(\tau_{\ell(N)j}\right)^{-\theta} \right)^{-1/\theta}. \tag{B.2}$$

The function $\mathbf{p}^*(\mathbf{w})$ has the following properties

(i) continuous in $\mathbf{w}$.

(ii) each component of $\mathbf{p}^*(\mathbf{w})$ *is homogeneous of degree one in* $\mathbf{w}$*;*

(iii) strictly increasing in $\mathbf{w}$*;*

(iv) strictly decreasing in τ_{ij} *for all* $\{i,j\} \in \mathcal{J}^2$ *and strictly increasing in* T_j *for all* $j \in \mathcal{J}$*.*

(v) $\forall \mathbf{w} \in R^J_{++}$*, bounded between* $\underline{\mathbf{p}^*}(\mathbf{w})$ *and* $\overline{\mathbf{p}^*}(\mathbf{w})$*:*

Proof. Let us set $\tilde{p}_j = \log(P_j)$ and $\tilde{w}_j = \log(w_j)$. For each supply chain $\boldsymbol{\ell} \in \mathcal{J}^N$, let

$$d_{p,i}(\boldsymbol{\ell}) = (1-\gamma_i) \sum_{n:\boldsymbol{\ell}(n)=i} \alpha_n \beta_n < 1 \qquad d_{w,i}(\boldsymbol{\ell}) = \gamma_i \sum_{n:\boldsymbol{\ell}(n)=i} \alpha_n \beta_n < 1$$

Note that for all $i \in \mathcal{J}$, $d_{p,i} \leq 1$ and $d_{w,i} \leq 1$. Now, for all $j \in \mathcal{J}$, define $f_j(\tilde{p}, \tilde{w})$

$$f_j(\tilde{p}, \tilde{w}) = \log(\kappa) - \frac{1}{\theta} \log \left(\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \prod_{n=1}^{N} \exp\left\{ -\theta \alpha_n \beta_n \left[\gamma_{\ell(n)} \tilde{w}_{\ell(n)} + \left(1-\gamma_{\ell(n)}\right) \tilde{p}_{\ell(n)} \right] \right\} T_{\ell(n)}^{\alpha_n \beta_n} \times \Upsilon_{\boldsymbol{\ell}} \right)$$

where $\Upsilon_{\boldsymbol{\ell}} = \prod_{n=1}^{N-1} \left(\tau_{\ell(n)\ell(n+1)}\right)^{-\theta\beta_n} \times \left(\tau_{\ell(N)j}\right)^{-\theta}$.

To establish uniqueness of $\mathbf{p}^*(\mathbf{w})$, we need to show that the Blackwell's sufficiency conditions for the contraction mapping theorem hold. Note that we also need to show that $f(p) = f(p, \tilde{w})$ is a bounded function for all values of $\tilde{w}$. This corresponds to property (v) of $\mathbf{p}^*(\mathbf{w})$, which will be proven below. For the time being, we proceed to prove the other parts of the theorem assuming a unique solution to the system exists.

If there indeed exists a unique solution to $\tilde{p} - f(\tilde{p}, \tilde{w}) = 0$, then homogeneity of degree one in wages (property (ii)) is simple to verify by noting that, given that $\sum_n \alpha_n \beta_n = 1$, if all wages and prices in the right-hand-side of (B.2) are multiplied by a common factor, the price level in the left-hand-side of that equation () is also scaled up or down by the same factor.

To prove differentiability and monotonicity with respect to $\mathbf{w}$, we need to determine the comparative static $\frac{\partial \mathbf{p}}{\partial \mathbf{w}}$. First, note that

$$\frac{\partial f_j(\tilde{p}, \tilde{w})}{\partial p_k} = \sum_{\boldsymbol{\ell} \in \mathcal{J}^N} d_{p,k}(\boldsymbol{\ell}) \pi_{\boldsymbol{\ell} j}, \tag{B.3}$$

where $\pi_{\boldsymbol{\ell} j}$ is given in (8) in the main text. Then, the Jacobian of the system $\tilde{p} - f(\tilde{p}, \tilde{w})$ is given by

$$\frac{\partial\left(\tilde{p} - f(\tilde{p}, \tilde{w})\right)}{\partial \tilde{p}} = I - A^P,$$

where $\left[A^P\right]_{ij} = \frac{\partial f_i(\tilde{p}, \tilde{w})}{\partial p_j}$. Note that matrix A^P is totally positive (this follows from the equation (B.3)), and therefore, by the Perron-Frobenius Theorem, we can bound above the largest eigenvalue of A^P, denoted by $\lambda_{\max}$, by the largest row sum of A^P. More precisely,

$$\begin{aligned} \lambda_{\max} &\leq \max_k \sum_i \frac{\partial f_k(\tilde{p}, \tilde{w})}{\partial \tilde{p}_i} = \max_k \sum_i \left(\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} d_{p,i}(\boldsymbol{\ell}) \pi_{\boldsymbol{\ell} k} \right) \\ &= \max_k \left(\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \left(\sum_{n \in \mathcal{N}} (1-\gamma_{\boldsymbol{\ell}(n)}) \alpha_n \beta_n \right) \pi_{\boldsymbol{\ell} k} \right) \end{aligned}$$

But consider the country with the lowest $\gamma_j = \underline{\gamma}$. And note that

$$\lambda_{\max} \leq (1-\underline{\gamma}) \max_k \left(\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \left(\sum_{n \in \mathcal{N}} \alpha_n \beta_n \right) \pi_{\boldsymbol{\ell} j} \right) = 1 - \underline{\gamma}.$$

Because $\lambda_{\max} < 1$, it follows that $I - A^P$ is an M-matrix, and, by properties of M-matrices, the inverse $\left(I - A^P\right)^{-1}$ is totally (weakly) positive. By the implicit function theorem, the Jacobian $\frac{\partial \tilde{p}}{\partial \tilde{w}}$ is given by

$$\frac{\partial \tilde{p}}{\partial \tilde{w}} = \left[I - A^P\right]^{-1} A^W,$$

where A^W is defined as

$$\left[A^W\right]_{ij} = \frac{\partial f_i(\tilde{p}, \tilde{w})}{\partial \tilde{w}_j} = \sum_{\boldsymbol{\ell} \in \mathcal{J}^N} d_{w,j}(\boldsymbol{\ell}) \pi_{\boldsymbol{\ell} i}.$$

Both A^W and $\left[I - A^P\right]^{-1}$ are totally positive, so $\tilde{p}$ is continuous (property (i)) and monotonically increasing (property (iii)) in $\tilde{w}$.

By analogy, we can show that property (iv) of the theorem also holds by defining $\forall \{i,j\} \in \mathcal{J}^2$, $\tilde{\tau}_{ij} = \log \tau_{ij}$ and $\forall j \in \mathcal{J}$, $\tilde{T}_j = \log T_j$, and also

$$d_{\tau,i}(\boldsymbol{\ell}) = \sum_{n: \boldsymbol{\ell}(n) = i} \beta_n, \qquad d_{T,i}(\boldsymbol{\ell}) = -\frac{1}{\theta} \sum_{n: \boldsymbol{\ell}(n) = i} \alpha_n \beta_n.$$

Applying the implicit function theorem to $f(p) = f(p, \tilde{w})$, we get:

$$\forall \{k,j\} \in \mathcal{J}^2: \qquad \frac{\partial \mathbf{p}}{\partial \tilde{\tau}_{kj}} = \left[I - A^P\right]^{-1} A^{\tau_{kj}},$$

where $A^{\tau_{kj}}$ is $J \times 1$ vector with

$$\left[A^{\tau_{kj}}\right]_i = \frac{\partial f_i(p)}{\partial \tilde{\tau}_{kj}} = \sum_{\boldsymbol{\ell} \in \mathcal{J}} d_{\tau_{kj},i}(\boldsymbol{\ell}) \pi_{\boldsymbol{\ell} i}.$$

Also,

$$\forall j \in \mathcal{J}: \qquad \frac{\partial \mathbf{p}}{\partial \tilde{T}} = \left[I - A^P\right]^{-1} A^T,$$

where A^T is $J \times J$ matrix with elements

$$\left[A^T\right]_{ij} = \frac{\partial f_i(p)}{\partial T_j} = \sum_{\boldsymbol{\ell} \in \mathcal{J}} d_{T,i}(\boldsymbol{\ell}) \pi_{\boldsymbol{\ell} i}.$$

Note that, as was shown above, $\left[I - A^P\right]^{-1}$ is totally positive. Then, since for all $i \in \mathcal{J}$ and for all supply chains $d_{T,i}(\boldsymbol{\ell}) \geq 0$, $f(p)$ is decreasing in T. By analogy, since for all $\{k,j,i\} \in \mathcal{J}^3$, $d_{\tau_{kj},i}\left(\boldsymbol{\ell}^{\boldsymbol{i}}\right)$ is totally positive, $f(p)$ is increasing in τ_{jk}.

As for property (v) on bounds, we can define $\underline{\mathbf{p}^*}(\mathbf{w})$ and $\overline{\mathbf{p}^*}(\mathbf{w})$ in the following way:

$$\overline{\mathbf{p}^*}(\mathbf{w}) = \exp\left(f\left(\log(\mathbf{p}), \tilde{\mathbf{w}}, \mathbf{T}_{\min}, \boldsymbol{\tau}_{\max}\right)\right) \qquad \underline{\mathbf{p}^*}(\mathbf{w}) = \exp\left(f\left(\log(\mathbf{p}), \tilde{\mathbf{w}}, \mathbf{T}_{\max}, \boldsymbol{\tau}_{\min}\right)\right),$$

where $\mathbf{T}_{\max}(\tau_{\max})$ and $\mathbf{T}_{\min}(\tau_{\min})$ are $J \times 1$ $(J \times J)$ vectors (matrices) with all elements equal to the upper bound on labor productivity (trade costs)$T_{\max}$ $(\tau_{\max})$ and the lower bound $T_{\min}(\tau_{\min})$, respectively. Then, we can note that the set $\mathbf{C}$, defined as

$$\mathbf{C} = \left\{ z \in R^J : \quad \log\left(\underline{p^*}_i(\mathbf{w})\right) \leq z_i \leq \log\left(\overline{p^*}_i(\mathbf{w})\right) \right\}$$

is compact and, by analogy with Alvarez and Lucas (2007), $f(\cdot, \tilde{\mathbf{w}}) : \mathbf{C} \to \mathbf{C}$.

Let us finally tackle the existence and unique of the solution by verifying Blackwell's sufficient conditions for $f(\cdot, \tilde{\mathbf{w}})$ to be a contraction on $\mathbf{C}$. We have already shown that $f(\cdot, \tilde{\mathbf{w}})$ is monotone. We next show that the discounting property also holds. Set $f_i(p) = f_i(p, \tilde{w})$ for any fixed $\tilde{w}$. Then, for $a > 0$ and some $\nu \in (0, 1)$, using a Taylor approximation and the mean-value theorem, we get:

$$\forall i \in \mathcal{J} : \quad f_i(p + a) = f_i(p) + \sum_{k \in \mathcal{J}} a \cdot \frac{\partial f_i(p + (1 - \nu)a)}{\partial p_k} \leq f_i(p) + a\left(1 - \underline{\gamma}\right)$$

The last inequality follows from the fact that every row sum of A^P can be bounded above by

$$(1 - \underline{\gamma}) \max_k \left(\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \left(\sum_{n \in \mathcal{N}} \alpha_n \beta_n \right) \pi_{\boldsymbol{\ell} j} \right) = 1 - \underline{\gamma}.$$

Thus, both the monotonicity and discounting properties hold for $f(p) = f(p, \tilde{w})$. Therefore, we can apply the Contraction Mapping Theorem to $f(p, \tilde{w})$, and conclude that there is a unique solution $\mathbf{p}^*(\mathbf{w})$ to the system $\tilde{p} - f(\tilde{p}, \tilde{w})$, and that it satisfies properties (i) through (v). ∎

Theorem 2 *There exists $\mathbf{w}^* \in R^{\mathcal{J}}_{++}$ which solves the system of equations*

$$Z(\mathbf{w}^*) = 0.$$

Proof. To show the existence of the equilibrium, we need to verify that the excess demand satisfies the following properties (see Propositions 17.C.1 in Mas-Colell et al., 1995, p. 585):

(i) $Z(\mathbf{w})$ is continuous on $\mathbb{R}^{\mathcal{J}}_{++}$;

(ii) $Z(\mathbf{w})$ is homogeneous of degree 0 in w

(iii) Walras Law: $\mathbf{w} \cdot Z(\mathbf{w}) = 0 \; \forall \mathbf{w} \in \mathbb{R}^J_{++}$;

(iv) for $k = \max_j L_j > 0$, $Z_i(\mathbf{w}) > -k$ for all $i = 1, ..., n$ and $\mathbf{w} \in \mathbb{R}^n_{++}$;

(v) if $w^m \to w^0$, where $w^0 \neq 0$ and $w^0_i \neq 0$ for some i, then

$$\lim_{w^m \to w^0} \left(\max_j \{Z_j(w^m)\} \right) = \infty$$

Let us discuss each of these properties in turn.

(i) **Continuity** of $Z(\mathbf{w})$ on $\mathbb{R}^{\mathcal{J}}_{++}$ follows since $\Pr(\Lambda^n_i, j)$ is a continuous function of $\mathbf{w}$ – for strictly positive wages, each supply chain $\boldsymbol{\ell}$ in $\mathcal{J}^N$ is realized with non-zero probability.

(ii) **Homogeneity of degree zero** follows since $\Pr(\Lambda_i^n, j)$ is homogeneous of degree 0 in $\mathbf{w}$. To show this, note that, from the proof of Theorem 1, the equilibrium price level $\mathbf{p}^*(\mathbf{w})$ is homogeneous of degree 1 in $\mathbf{w}$. Then, both nominator and denominator (i.e., the destination specific term Θ_j) of $\Pr(\Lambda_i^n, j)$ are homogeneous of degree $-\theta$ in $\mathbf{w}$ (remember that $\sum_{n\in\mathcal{N}} \alpha_n\beta_n = 1$). It follows that $\Pr(\Lambda_i^n, j)$ is homogeneous of degree 0 in $\mathbf{w}$, and thus $Z(\mathbf{w})$ is homogeneous of degree 0 in $\mathbf{w}$ as well.

(iii) **Walras Law** follows since the system, $\mathbf{w}\cdot Z(\mathbf{w}) = 0$ is just the set of the general equilibrium conditions. Moreover, by summing up $Z(\mathbf{w})$, we get:

$$\begin{aligned}
\sum_{i\in\mathcal{J}} w_i \cdot Z_i(\mathbf{w}) &= \sum_{i\in\mathcal{J}} \gamma_i \left(\sum_{j\in\mathcal{J}} \sum_{n\in\mathcal{N}} \alpha_n\beta_n \times \Pr(\Lambda_i^n, j) \times \frac{1}{\gamma_j} w_j L_j \right) - \sum_{i\in\mathcal{J}} \frac{1}{\gamma_i} w_i L_i \\
&= \left(\sum_{n\in\mathcal{N}} \alpha_n\beta_n \times \sum_{j\in\mathcal{J}} \underbrace{\sum_{i\in\mathcal{J}} \Pr(\Lambda_i^n, j)}_{=1} \times \frac{1}{\gamma_j} w_j L_j \right) - \sum_{i\in\mathcal{J}} \frac{1}{\gamma_i} w_i L_i \\
&= \left(\underbrace{\sum_{n\in\mathcal{N}} \alpha_n\beta_n}_{=1} \times \sum_{j\in\mathcal{J}} \frac{1}{\gamma_j} w_j L_j \right) - \sum_{i\in\mathcal{J}} \frac{1}{\gamma_i} w_i L_i = 0.
\end{aligned}$$

Hence, $\mathbf{w}\cdot Z(\mathbf{w}) = 0$.

(iv) **The lower bound on** $Z(\mathbf{w})$: Since the first term in equation (B.1) is always positive, it follows that $Z(\mathbf{w})$ can be bounded from below by $Z_i(\mathbf{w}) \geq -\frac{1}{\gamma_i} L_i$.

(v) **The limit case**: Suppose $\{w^m\}$ is a sequence such that $w^m \to w^0 \neq 0$, and $w_i^0 = 0$ for some $i \in \mathcal{J}$. In this case, and given that all trade costs parameters are bounded, the probability of the supply chain that is located entirely in country i converges to 1, and the probabilities of realization of all other supply chains converge to 0 (keeping the destination fixed). Let $\Pr(i^N, j)$ denote the probability of realization of the supply chain for which all stages are located in country i with destination j. Then,

$$\lim_{w^m\to w^0} \left(\max_k \{Z_k(\mathbf{w})\} \right) = \lim_{w^m\to w^0} (Z_i(\mathbf{w}))$$

and

$$\begin{aligned}
\lim_{w^m\to w^0} \left(\max_k \{Z_k(\mathbf{w})\} \right) &= \lim_{w^m\to w^0} \left(\frac{1}{w_i} \sum_{j\in\mathcal{J}} \left(\sum_{n\in N} \alpha_n\beta_n \right) \Pr(i^N, j) \frac{1}{\gamma_j} w_j L_j \right) - \frac{1}{\gamma_i} L_i \\
&= \lim_{w^m\to w^0} \left(\frac{1}{w_i} \sum_{j\in\mathcal{J}} \Pr(i^N, j) \frac{1}{\gamma_j} w_j L_j \right) - \frac{1}{\gamma_i} L_i \\
&= \lim_{w^m\to w^0} \left(\frac{1}{w_i} \sum_{j\neq i} \Pr(i^N, j) \frac{1}{\gamma_j} w_j L_j \right) = +\infty.
\end{aligned}$$

In sum, conditions (i) through (v) hold and thus a general equilibrium exists.

■

Theorem 3 *The solution* $\mathbf{w}^* \in R_{++}^{\mathcal{J}}$ *to the system of equations* $Z(\mathbf{w}^*) = 0$ *is unique if the following condition holds:*

$$\frac{2(1-\overline{\gamma})}{\xi^{\theta}(1-\underline{\gamma})} - (1-\underline{\gamma}) - \xi^{2\theta} \geq 0, \quad \textit{where} \quad \xi = \max_{i,j\in\mathcal{J}} \frac{\max_{k\in\mathcal{J}} \tau_{kj}/\tau_{ki}}{\min_{k\in\mathcal{J}} \tau_{kj}/\tau_{ki}} = 1,$$

and where $\overline{\gamma}$ *and* $\underline{\gamma}$ *are the largest and smallest values of* γ_j.

Proof. The proof boils down to verifying that $Z(\mathbf{w})$ has the gross substitutes property in $\mathbf{w}$ under the condition stated in the Theorem (see Proposition 17.F.3 in Mas-Colell et al., 1995, p. 613). More specifically, we need to show that

$$\forall\{i,k\} \in \mathcal{J}^2, i \neq k: \qquad \frac{\partial Z_i}{\partial w_k} > 0.$$

Totally differentiating the equation (B.1) wrt w_k, $k \neq i$, we get:

$$\frac{\partial Z_i(\mathbf{w})}{\partial w_k} = \frac{1}{w_i}\left(\sum_{n\in\mathcal{N}} \alpha_n\beta_n \times \left(\frac{1}{\gamma_k} L_k \Pr(\Lambda_i^n, k) + \sum_{j\in\mathcal{J}} \frac{1}{\gamma_j} w_j L_j \frac{d\Pr(\Lambda_i^n, j)}{dw_k}\right)\right),$$

where

$$\frac{d\Pr(\Lambda_i^n, j)}{dw_k} = \frac{\partial \Pr(\Lambda_i^n, j)}{\partial w_k} + \sum_{l\in\mathcal{J}} \frac{\partial \Pr(\Lambda_i^n, j)}{\partial P_l}\frac{\partial P_l}{\partial w_k}$$

From here, we proceed in three steps:

Step 1:.

Remember that $\Pr(\Lambda_i^n, j) = \sum_{\boldsymbol{\ell}\in\Lambda_i^n} \pi_{\boldsymbol{\ell}j}$, where $\Lambda_i^n = \left\{\boldsymbol{\ell} \in \mathcal{J}^N \mid \ell(n) = i\right\}$. Thus,

$$\frac{\partial \Pr(\Lambda_i^n, j)}{\partial w_k} = \frac{\Pr(\Lambda_i^n, j)}{w_k}\left(\frac{\partial \log(\Pr(\Lambda_i^n, j)\cdot\Theta_j)}{\partial \log(w_k)} - \frac{\partial \log(\Theta_j)}{\partial \log(w_k)}\right). \tag{B.4}$$

Since in equilibrium $\Theta_j = (p_j(\mathbf{w}))^{-\theta}$, we can use the envelope theorem to get

$$\frac{\partial \Pr(\Lambda_i^n, j)}{\partial w_k} = \frac{\theta}{w_k}\left(-\sum_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,k}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j} + \Pr(\Lambda_i^n, j)\frac{\partial \tilde{p}_j}{\partial \tilde{w}_j}\right).$$

Step 2: Bounds on $\frac{\partial \tilde{p}}{\partial \tilde{w}}$.

Note that we can bound the row sums of A^P and $[I - A^P]^{-1}$:

$$(1-\overline{\gamma})\,\mathbf{1} \leq A^P\mathbf{1} \leq \left(1-\underline{\gamma}\right)\mathbf{1},$$

$$\left(1-\underline{\gamma}\right)^{-1}\mathbf{1} \leq \left[I - A^P\right]^{-1}\mathbf{1} \leq (1-\overline{\gamma})^{-1}\,\mathbf{1}, \tag{B.5}$$

where $\overline{\gamma}$ and $\underline{\gamma}$ are the largest and smallest values of γ_j.

For two identical supply chains with different destinations i and j, $\boldsymbol{\ell}^{\boldsymbol{i}}$ and $\boldsymbol{\ell}^{\boldsymbol{j}}$ it holds that

$$\forall\{i,j\} \in \mathcal{J}^2: \qquad d_{p,k}(\boldsymbol{\ell}^{\boldsymbol{j}}) = d_{p,k}(\boldsymbol{\ell}^{\boldsymbol{i}}), \qquad d_{w,k}(\boldsymbol{\ell}^{\boldsymbol{j}}) = d_{w,k}(\boldsymbol{\ell}^{\boldsymbol{i}})$$

$$\forall\{i,j\} \in \mathcal{J}^2: \qquad \pi_{\boldsymbol{\ell}j} = \frac{\left(\tau_{\boldsymbol{\ell}(N)j}/\tau_{\boldsymbol{\ell}(N)i}\right)^{-\theta}\pi_{\boldsymbol{\ell}i}}{\sum_{\tilde{\boldsymbol{\ell}}\in\Lambda}\left(\tau_{\tilde{\boldsymbol{\ell}}(N)j}/\tau_{\tilde{\boldsymbol{\ell}}(N)i}\right)^{-\theta}\pi_{\tilde{\boldsymbol{\ell}}i}}$$

Let's set $\xi = \max_{i,j\in\mathcal{J}} \dfrac{\max_{k\in\mathcal{J}} \tau_{kj}/\tau_{ki}}{\min_{k\in\mathcal{J}} \tau_{kj}/\tau_{ki}} \geq 1$.

$$\forall\{i,j,k\} \in \mathcal{J}^2: \qquad \frac{1}{\xi^\theta} \leq [A^W]_{ij} \cdot \left([A^W]_{kj}\right)^{-1} \leq \xi^\theta$$

Since $\dfrac{\partial \mathbf{p}}{\partial w_j} = \left[I - A^P\right]^{-1} A^W_{[j]}$, where $A^W_{[j]}$ is the jth column of A^W, we can bound the ratio $\dfrac{\partial \tilde{p}_j}{\partial \tilde{w}_k} \Big/ \dfrac{\partial \tilde{p}_i}{\partial \tilde{w}_k}$:

$$\forall\{i,j\} \in \mathcal{J}^2: \qquad \frac{(1-\overline{\gamma})}{\xi(1-\underline{\gamma})} \leq \frac{\partial \tilde{p}_j}{\partial \tilde{w}_k} \Big/ \frac{\partial \tilde{p}_i}{\partial \tilde{w}_k} \leq \frac{\xi(1-\underline{\gamma})}{(1-\overline{\gamma})}.$$

Since all elements of A^W and A^P are less than one,

$$\left[A^W\right]_{jk} \leq \frac{\partial \tilde{p}_j}{\partial \tilde{w}_k} \leq \frac{1}{(1-\overline{\gamma})}. \tag{B.6}$$

Finally we show that for all n and i,

$$\frac{\sum\limits_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,m}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j}}{\left[A^W\right]_{jk}} \leq \Pr\left(\Lambda_i^n, j\right) \xi^{2\theta} \tag{B.7}$$

Let $\lambda_{\boldsymbol{\ell}}^n$ denote the set of supply chains, identical to $\boldsymbol{\ell} \in J^N$ in all stages except for n (note that there are J chains in $\lambda_{\boldsymbol{\ell}}^n$). With this definition we have

$$\left[A^W\right]_{jk} \geq \sum_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,m}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j} \left(\frac{\sum_{\tilde{\boldsymbol{\ell}}\in\lambda_{\boldsymbol{\ell}}^n} \pi_{\boldsymbol{\ell}j}}{\pi_{\boldsymbol{\ell}j}}\right)$$

and

$$\frac{\sum\limits_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,m}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j}}{\left[A^W\right]_{jk}} \leq \frac{\sum\limits_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,m}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j}}{\sum\limits_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,m}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j}} \left(\min_{\boldsymbol{\ell}\in\Lambda_i^n} \left(\frac{\sum_{\tilde{\boldsymbol{\ell}}\in\lambda_{\boldsymbol{\ell}}^n} \pi_{\boldsymbol{\ell}j}}{\pi_{\boldsymbol{\ell}j}}\right)\right)^{-1} \tag{B.8}$$

Then, let us bound $\Pr\left(\Lambda_i^n, j\right)$:

$$\Pr\left(\Lambda_i^n, j\right) \geq \left(\max_{\boldsymbol{\ell}\in\Lambda_i^n} \left(\frac{\sum_{\tilde{\boldsymbol{\ell}}\in\lambda_{\boldsymbol{\ell}}^n} \pi_{\boldsymbol{\ell}j}}{\pi_{\boldsymbol{\ell}j}}\right)\right)^{-1} \tag{B.9}$$

Therefore, combining (B.8) and (B.9) we get:

$$\frac{\sum\limits_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,m}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j}}{\left[A^W\right]_{jk}} \leq \left(\max_{\boldsymbol{\ell}\in\Lambda_i^n} \left(\frac{\sum_{\tilde{\boldsymbol{\ell}}\in\lambda_{\boldsymbol{\ell}}^n} \pi_{\boldsymbol{\ell}j}}{\pi_{\boldsymbol{\ell}j}}\right)\right) \cdot \left(\min_{\boldsymbol{\ell}\in\Lambda_i^n} \left(\frac{\sum_{\tilde{\boldsymbol{\ell}}\in\lambda_{\boldsymbol{\ell}}^n} \pi_{\boldsymbol{\ell}j}}{\pi_{\boldsymbol{\ell}j}}\right)\right)^{-1} \Pr\left(\Lambda_i^n, j\right)$$

Note that by definition of $\lambda_{\boldsymbol{\ell}}^n$,

$$\left(\frac{\sum_{\tilde{\boldsymbol{\ell}}\in\lambda_{\boldsymbol{\ell}}^n} \pi_{\boldsymbol{\ell}j}}{\pi_{\boldsymbol{\ell}j}}\right) \in \left[\frac{\sum_{k\in\mathcal{J}} \left((c_k)^{-\theta} T_k\right)^{\alpha_n\beta_n}}{\xi^\theta \left((c_i)^{-\theta} T_i\right)^{\alpha_n\beta_n}}, \frac{\xi^\theta \sum_{k\in\mathcal{J}} \left((c_k)^{-\theta} T_k\right)^{\alpha_n\beta_n}}{\left((c_i)^{-\theta} T_i\right)^{\alpha_n\beta_n}}\right],$$

so

$$\frac{\sum\limits_{\boldsymbol{\ell}\in\Lambda_i^n} d_{w,m}(\boldsymbol{\ell})\pi_{\boldsymbol{\ell}j}}{\left[A^W\right]_{jk}} \leq \xi^{2\theta} \Pr\left(\Lambda_i^n, j\right).$$

Step 3: To prove the GS property, we need to show that for a fixed destination j, fixed stage n and $m \neq i$

$$\frac{\partial \Pr(\Lambda_i^n, j)}{\partial w_m} + \sum_{k\in\mathcal{J}} \frac{\partial \Pr(\Lambda_i^n, j)}{\partial \tilde{p}_k}\frac{\partial \tilde{p}_k}{\partial w_m} \geq 0.$$

By analogy with Step 1,

$$\sum_{k\in\mathcal{J}} \frac{\partial \Pr(\Lambda_i^n, j)}{\partial \tilde{p}_k}\frac{\partial \tilde{p}_k}{\partial \tilde{w}_m} = \Pr(\Lambda_i^n, j) \sum_{k\in\mathcal{J}} \frac{\partial \tilde{p}_k}{\partial \tilde{w}_m}\left(\frac{\partial \log(\Pr(\Lambda_i^n, j)\cdot\Theta_j)}{\partial \log(p_k)} - \frac{\partial \log(\Theta_j)}{\partial \log(p_k)}\right)$$

$$\sum_{k\in\mathcal{J}} \frac{\partial \pi_{\boldsymbol{\ell} j}}{\partial \tilde{p}_k}\frac{\partial \tilde{p}_k}{\partial \tilde{w}_m} = \theta\pi_{\boldsymbol{\ell} j}\left(-\left(\sum_{k\in\mathcal{J}} d_{p,k}(\boldsymbol{\ell})\frac{\partial \tilde{p}_k}{\partial \tilde{w}_m}\right) + \frac{\partial \tilde{p}_j}{\partial \tilde{w}_m}\right). \tag{B.10}$$

Combining equations (B.4) and (B.10),

$$\frac{d\Pr(\Lambda_i^n, j)}{d\tilde{w}_k} = \theta\left(2\Pr(\Lambda_i^n, j)\frac{\partial \tilde{p}_j}{\partial \tilde{w}_m} - \sum_{\boldsymbol{\ell}\in\Lambda_i^n} \pi_{\boldsymbol{\ell} j}\left(\left(\sum_{k\in\mathcal{J}} d_{p,k}(\boldsymbol{\ell})\frac{\partial \tilde{p}_k}{\partial \tilde{w}_m}\right) + d_{w,m}(\boldsymbol{\ell})\right)\right).$$

Let us use the bounds derived in Step 2: from equation (B.5),

$$\frac{d\Pr(\Lambda_i^n, j)}{dw_k} \geq \theta\left(\frac{\partial \tilde{p}_j}{\partial \tilde{w}_m}\left(\frac{2(1-\overline{\gamma})}{\xi^\theta(1-\underline{\gamma})}\Pr(\Lambda_i^n, j) - \sum_{\boldsymbol{\ell}\in\Lambda_i^n} \pi_{\boldsymbol{\ell} j}\left(\sum_{k\in\mathcal{J}} d_{p,k}(\boldsymbol{\ell})\right)\right) - \sum_{\boldsymbol{\ell}\in\Lambda_i^n} \pi_{\boldsymbol{\ell} j} d_{w,m}(\boldsymbol{\ell})\right).$$

Finally, invoking equations (B.6) and (B.6), we have:

$$\frac{d\Pr(\Lambda_i^n, j)}{dw_k} \geq \theta[A^W]_{kj}\Pr(\Lambda_i^n, j)\left(\frac{2(1-\overline{\gamma})}{\xi^\theta(1-\underline{\gamma})} - \frac{1}{\Pr(\Lambda_i^n, j)}\sum_{\boldsymbol{\ell}\in\Lambda_i^n} \pi_{\boldsymbol{\ell} j}\left(\sum_{k\in\mathcal{J}} d_{p,k}(\boldsymbol{\ell})\right) - \xi^{2\theta}\right)$$

and thus

$$\frac{d\Pr(\Lambda_i^n, j)}{dw_k} \geq \theta[A^W]_{kj}\Pr(\Lambda_i^n, j)\left(\frac{2(1-\overline{\gamma})}{\xi^\theta(1-\underline{\gamma})} - (1-\underline{\gamma}) - \xi^{2\theta}\right). \tag{B.11}$$

■

Corollary 1 *Suppose the trade costs have the following form:*

$$(\tau_{ij})^{-\theta} = \rho_i\rho_j.$$

Then the equilibrium is unique if

$$\underline{\gamma}(3-\underline{\gamma}) \geq 2\overline{\gamma} \tag{B.12}$$

Proof. Note that for this specification of trade costs $\xi = 1$, and the RHS of equation (B.11) is positive whenever (B.12) holds. ■

B.3 Further Details on Suggestive Evidence

In this Appendix, we provide additional details on the suggestive empirical results in section 4.5. We begin by exploring the robustness of our results in Table I. For that table, we used 2011 data for 180 countries from the Eora dataset. In Table B.1, we re-run the same specifications but focusing on manufacturing flows (rather than overall flows). Interestingly, the differential home bias in final-good production in Table I of the draft is no longer observed when focusing on manufacturing flows. Nevertheless, the response of trade flows to distance continues to be significantly higher for final goods than for intermediate inputs.

In Table B.2, we revert to the full Eora sample, but instead of focusing on 2011 data, we pool data from the 19 years for which the Eora dataset is available, namely 1995-2013, while including exporter-year and importer-year fixed effects (rather than the simpler exporter and importer fixed effects in Table I). As is apparent from comparing Tables I and B.2, the results are remarkably similar, both qualitatively as well as quantitatively. The reason for this is that the estimated elasticities are quite actually quite stable over time, as we have verified by replicating Table I year by year (details available upon request).

Tables B.3 and B.4 run the same specifications with the WIOT database using its 2013 and 2016 releases, respectively. The former covers the period 1995-2011 for 40 countries, while the latter covers 2000-2014 for 43 countries. As mentioned in the main text, the results with the 2013 release of the WIOD are generally qualitatively in line with those obtained with the Eora database, and indicate a significantly lower distance elasticity and lower 'home bias' in intermediate-input relative to final-good trade. Nevertheless, the results with the 2016 release of the same dataset are much weaker, and only indicate a lower 'home bias' in intermediate-input relative to final-good trade.

We finally incorporate the scatter plots mentioned in section 4.5, when describing the results in Table II. More precisely, the left panel corresponds to the partial correlation underlying column (5) of Table II (i.e., partialling out GDP per capita). The right panel is the analogous scatter plot after dropping the Netherlands ('NLD').

Table B.1. Trade Cost Elasticities for Final Goods and Input Flows (Manufacturing)

	Total Flows		Final-Good Flows		Input Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-1.162***	-0.793***	-1.275***	-0.893***	-1.119***	-0.755***
	(0.018)	(0.015)	(0.019)	(0.016)	(0.018)	(0.014)
Contiguity		1.243***		1.283***		1.246***
		(0.099)		(0.108)		(0.096)
Language		0.473***		0.580***		0.422***
		(0.027)		(0.031)		(0.026)
Domestic		4.299***		4.188***		4.336***
		(0.183)		(0.206)		(0.175)
Observations	32,041	32,041	32,041	32,041	32,041	32,041
R^2	0.973	0.976	0.959	0.963	0.977	0.979

Notes: Standard errors are clustered at the country-pair level. ***, **, and * denote 1, 5, and 10 percent significance levels. All regressions include exporter and importer fixed effects.

Table B.2. Trade Cost Elasticities for Final Goods and Input Flows (Eora all years)

	Total Flows		Final-Good Flows		Input Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-1.118***	-0.716***	-1.242***	-0.812***	-1.065***	-0.680***
	(0.020)	(0.013)	(0.021)	(0.015)	(0.019)	(0.013)
Contiguity		1.170***		1.189***		1.173***
		(0.088)		(0.096)		(0.086)
Language		0.401***		0.504***		0.358***
		(0.024)		(0.028)		(0.023)
Domestic		5.480***		5.800***		5.197***
		(0.159)		(0.180)		(0.151)
Observations	615,600	615,600	615,600	615,600	615,600	615,600
R^2	0.977	0.980	0.958	0.963	0.976	0.980

Notes: Standard errors are clustered at the country-pair level. ***, **, and * denote 1, 5, and 10 percent significance levels. All regressions include exporter-year and importer-year fixed effects.

Table B.3. Trade Cost Elasticities for Final Goods and Input Flows (2013 WIOD sample)

	Total Flows		Final-Good Flows		Input Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-1.550***	-1.072***	-1.579***	-1.021***	-1.541***	-1.110***
	(0.056)	(0.041)	(0.064)	(0.044)	(0.053)	(0.042)
Contiguity		0.370***		0.394***		0.375***
		(0.117)		(0.128)		(0.118)
Language		0.212		0.270*		0.181
		(0.145)		(0.139)		(0.156)
Domestic		3.141***		3.710***		2.771***
		(0.268)		(0.287)		(0.271)
Observations	27,194	27,194	27,186	27,186	27,194	27,194
R^2	0.981	0.986	0.969	0.978	0.978	0.982

Notes: Standard errors are clustered at the country-pair level. ***, **, and * denote 1, 5, and 10 percent significance levels. All regressions include exporter-year and importer-year fixed effects.

Table B.4. Trade Cost Elasticities for Final Goods and Input Flows (2016 WIOD sample)

	Total Flows		Final-Good Flows		Input Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
Distance	-1.638***	-1.222***	-1.641***	-1.142***	-1.654***	-1.289***
	(0.053)	(0.042)	(0.059)	(0.042)	(0.050)	(0.045)
Contiguity		0.266**		0.292**		0.251**
		(0.108)		(0.116)		(0.111)
Language		0.129		0.197		0.091
		(0.126)		(0.125)		(0.134)
Domestic		2.950***		3.537***		2.584***
		(0.249)		(0.263)		(0.257)
Observations	26,460	26,460	26,460	26,460	26,460	26,460
R^2	0.982	0.986	0.973	0.980	0.978	0.981

Notes: Standard errors are clustered at the country-pair level. ***, **, and * denote 1, 5, and 10 percent significance levels. All regressions include exporter-year and importer-year fixed effects.

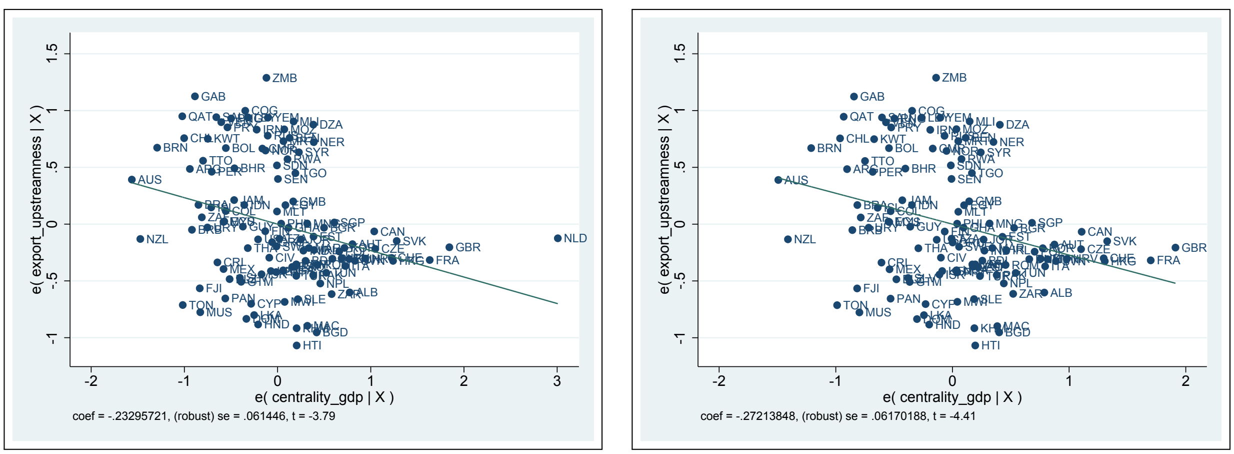

Figure B.3: Partial Correlation between Export Upstreamness and Centrality

B.4 Estimation Algorithm

We estimate our model's parameters using the mathematical programming with equilibrium constraints (MPEC) approach of Su and Judd (2012). We implement our model numerically in Matlab using the KNITRO optimizer developed by Byrd et al. (2006). This approach requires providing the optimizer with the analytical gradient and Jacobian matrix of partial first derivatives in order to deliver precise parameter estimates. We now provide the formulas for these partial derivatives – a tool that will prove useful in future implementations of our model. For the sake of simplicity, we provide the algebra for the model with separate supply chains for intermediate inputs and final goods in the single-industry case. The multi-industry formulas are analogous.

B.4.1 MPEC Estimation Algorithm

The MPEC algorithm requires defining the estimation problem as a minimization problem with two parts. First, the objective function is given by a loss function capturing the distance between a set of targeted and estimated moments. Second, the constraints include both relationships between the estimated parameters of interest and the equilibrium equations defining the model's endogenous variables. In contrast to nested fixed point algorithms, the MPEC approach does not solve for an equilibrium in each parameter iteration but yields much faster performance by only ensuring that the equilibrium constraints are satisfied in the very last iteration.

Endogenous Variables We have three sets of endogenous variables yielding a total of $2\left(N^X-1\right)+2\left(N^X-1\right)+5J+2J^2$ variables. First, the parameters $\left\{\alpha_n^X,\beta_n^X,T_j^X,\alpha_n^F,\beta_n^F,T_j^F,\gamma_j\right\}$. Second, the general equilibrium variables $\left\{w_j,P_j^X\right\}$ with P_j^X the unit price of a CES bundle of finished intermediate input varieties in country j. Third, the estimated input-output moments $\{X_{ij},F_{ij}\}$.

Objective Function We minimize the following loss function

$$Q=\sum_{i\in\mathcal{J}}\sum_{j\in\mathcal{J}}\omega_{ij}^X\left(X_{ij}-\hat{X}_{ij}\right)^2+\sum_{i\in\mathcal{J}}\sum_{j\in\mathcal{J}}\omega_{ij}^F\left(F_{ij}-\hat{F}_{ij}\right)^2,$$

with ω_{ij}^X and ω_{ij}^F the weights on the targeted moments and $\hat{X}_{ij}$ and $\hat{F}_{ij}$ the targeted data.

Linear Constraints We impose general equilibrium and three normalizations as linear constraints

$$\begin{aligned}
w_jL_j &= \sum_{i\in\mathcal{J}}(X_{ji}+F_{ji})-\sum_{i\in\mathcal{J}}X_{ij},\ \forall j,\\
\sum_{j\in\mathcal{J}}w_jL_j &= 100,\\
\sum_{j\in\mathcal{J}}T_j^X &= 100,\\
\sum_{j\in\mathcal{J}}T_j^F &= 100.
\end{aligned}$$

Nonlinear Constraints We impose three sets of nonlinear constraints: i) the relation between the sequential production parameters, ii) the price index definition, iii) the input-output variables definition:

$$\begin{aligned}
\beta_n^X &= \left(1-\alpha_{n+1}^X\right)\beta_{n+1}^X,\ \forall n<N^X,\\
\beta_n^F &= \left(1-\alpha_{n+1}^F\right)\beta_{n+1}^F,\ \forall n<N^F,\\
P_j^X &= \Gamma\left(1+\frac{1-\sigma}{\theta}\right)^{\frac{1}{1-\sigma}}\left[\Theta_j^X\right]^{-\frac{1}{\theta}},\ \forall j,\\
F_{ij} &= \sum_{\boldsymbol{\ell}\in\Lambda_i^N}\pi_{\boldsymbol{\ell}j}^F\times\left(w_jL_j-D_j\right),\ \forall i,j,\\
X_{ij} &= \sum_{n=1}^{N^F-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\beta_n^F\times\sum_{k\in\mathcal{J}}\pi_{\boldsymbol{\ell}k}^F\times\left(w_kL_k-D_k\right),\\
&+\sum_{n=1}^{N^X-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\beta_n^X\times\sum_{k\in\mathcal{J}}\pi_{\boldsymbol{\ell}k}^X\times\frac{1-\gamma_k}{\gamma_k}w_kL_k,\\
&+\sum_{\boldsymbol{\ell}\in\Lambda_i^N}\pi_{\boldsymbol{\ell}j}^X\times\frac{1-\gamma_j}{\gamma_j}w_jL_j,\ \forall i,j,
\end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function, D_j is country j's (exogenous) gross deficit. Note that Θ_j^X, $\pi_{\boldsymbol{\ell}j}^X$, and $\pi_{\boldsymbol{\ell}j}^X$ are auxiliary variables depending only on problem's endogenous variables.

B.4.2 Auxiliary Partial Derivatives

Remember from (17) the marginal costs $c_j = (w_j)^{\gamma_j}\left(P_j^X\right)^{1-\gamma_j}$. The marginal cost's partial derivatives equal

$$\frac{\partial c_j}{\partial w_j} = \gamma_j \frac{c_j}{w_j}, \qquad \frac{\partial c_j}{\partial P_j^X} = \left(1-\gamma_j\right)\frac{c_j}{P_j^X}, \qquad \frac{\partial c_j}{\partial \gamma_j} = c_j \ln\left(\frac{w_j}{P_j^X}\right).$$

Throughout the rest of this subsection we drop the X and F superscripts for notational simplicity but all derivations apply for both cases. Define the auxiliary variable

$$\Theta_j(\boldsymbol{\ell}) = \prod_{n=1}^{N-1}\left(T_{\ell(n)}\right)^{\alpha_n\beta_n}\left(c_{\ell(n)}^{\alpha_n}\tau_{\ell(n)\ell(n+1)}\right)^{-\theta\beta_n} \times \left(T_{\ell(N)}\right)^{\alpha_N\beta_N}\left(c_{\ell(N)}^{\alpha_N}\tau_{\ell(N)j}\right)^{-\theta},$$

so that $\Theta_j = \sum_{\boldsymbol{\ell}\in\mathcal{J}^N}\Theta_j(\boldsymbol{\ell})$. The partial derivatives of Θ_j equal

$$\frac{\partial \Theta_j}{\partial \alpha_m} = \beta_m \sum_{\boldsymbol{\ell}\in\mathcal{J}^N} \ln\left(T_{\ell(m)}c_{\ell(m)}^{-\theta}\right)\Theta_j(\boldsymbol{\ell}),$$

$$\frac{\partial \Theta_j}{\partial \beta_m} = \sum_{\boldsymbol{\ell}\in\mathcal{J}^N}\left(\alpha_m \ln\left(T_{\ell(m)}c_{\ell(m)}^{-\theta}\right) + \ln\left(\tau_{\ell(m)\ell(m+1)}^{-\theta}\right)\right)\Theta_j(\boldsymbol{\ell}),$$

$$\frac{\partial \Theta_j}{\partial T_i} = \frac{1}{T_i}\sum_{n=1}^{N}\alpha_n\beta_n \times \sum_{\boldsymbol{\ell}\in\Lambda_i^n}\Theta_j(\boldsymbol{\ell}),$$

$$\frac{\partial \Theta_j}{\partial \gamma_i} = -\frac{\theta}{c_i}\frac{\partial c_i}{\partial \gamma_i}\sum_{n=1}^{N}\alpha_n\beta_n \times \sum_{\boldsymbol{\ell}\in\Lambda_i^n}\Theta_j(\boldsymbol{\ell}),$$

$$\frac{\partial \Theta_j}{\partial w_i} = -\frac{\theta}{c_i}\frac{\partial c_i}{\partial w_i}\sum_{n=1}^{N}\alpha_n\beta_n \times \sum_{\boldsymbol{\ell}\in\Lambda_i^n}\Theta_j(\boldsymbol{\ell}),$$

$$\frac{\partial \Theta_j}{\partial P_i^X} = -\frac{\theta}{c_i}\frac{\partial c_i}{\partial P_i^X}\sum_{n=1}^{N}\alpha_n\beta_n \times \sum_{\boldsymbol{\ell}\in\Lambda_i^n}\Theta_j(\boldsymbol{\ell}).$$

Similarly, the partial derivatives of $\pi_{\boldsymbol{\ell} j}$ can be written in terms of these partial derivatives as

$$
\begin{aligned}
\frac{\partial \pi_{\boldsymbol{\ell} j}}{\partial \alpha_m} &= \beta_m \ln\left(T_{\ell(n)} c_{\ell(m)}^{-\theta}\right) \times \pi_{\boldsymbol{\ell} j} - \frac{\pi_{\boldsymbol{\ell} j}}{\Theta_j}\frac{\partial \Theta_j}{\partial \alpha_m}, \\
\frac{\partial \pi_{\boldsymbol{\ell} j}}{\partial \beta_m} &= \left(\alpha_m \ln\left(T_{\ell(m)} c_{\ell(m)}^{-\theta}\right) + \ln\left(\tau_{\ell(m)\ell(m+1)}^{-\theta}\right)\right) \times \pi_{\boldsymbol{\ell} j} - \frac{\pi_{\boldsymbol{\ell} j}}{\Theta_j}\frac{\partial \Theta_j}{\partial \beta_m}, \\
\frac{\partial \pi_{\boldsymbol{\ell} j}}{\partial T_i} &= \frac{1}{T_i}\sum_{n=1}^{N} \alpha_n \beta_n \times 1\left[\ell(n) = i\right] \pi_{\boldsymbol{\ell} j} - \frac{\pi_{\boldsymbol{\ell} j}}{\Theta_j}\frac{\partial \Theta_j}{\partial T_i}, \\
\frac{\partial \pi_{\boldsymbol{\ell} j}}{\partial \gamma_i} &= -\frac{\theta}{c_i}\frac{\partial c_i}{\partial \gamma_i}\sum_{n=1}^{N} \alpha_n \beta_n \times 1\left[\ell(n) = i\right] \pi_{\boldsymbol{\ell} j} - \frac{\pi_{\boldsymbol{\ell} j}}{\Theta_j}\frac{\partial \Theta_j}{\partial \gamma_i}, \\
\frac{\partial \pi_{\boldsymbol{\ell} j}}{\partial w_i} &= -\frac{\theta}{c_i}\frac{\partial c_i}{\partial w_i}\sum_{n=1}^{N} \alpha_n \beta_n \times 1\left[\ell(n) = i\right] \pi_{\boldsymbol{\ell} j} - \frac{\pi_{\boldsymbol{\ell} j}}{\Theta_j}\frac{\partial \Theta_j}{\partial w_i}, \\
\frac{\partial \pi_{\boldsymbol{\ell} j}}{\partial P_i^X} &= -\frac{\theta}{c_i}\frac{\partial c_i}{\partial P_i^X}\sum_{n=1}^{N} \alpha_n \beta_n \times 1\left[\ell(n) = i\right] \pi_{\boldsymbol{\ell} j} - \frac{\pi_{\boldsymbol{\ell} j}}{\Theta_j}\frac{\partial \Theta_j}{\partial P_i^X}.
\end{aligned}
$$

Writing these formulas in the computer can be challenging and prone to typos, so it is useful to note that the following relation should hold for any parameter x: $\sum_{\boldsymbol{\ell} \in \mathcal{J}^N} \partial \pi_{\boldsymbol{\ell} j}/\partial x = 0$.

B.4.3 Gradient and Jacobian Matrix

Objective Function Gradient The derivatives for the objective function equal

$$
\frac{\partial Q}{\partial X_{ij}} = 2\omega_{ij}^X \left(X_{ij} - \hat{X}_{ij}\right), \qquad \frac{\partial Q}{\partial F_{ij}} = 2\omega_{ij}^F \left(F_{ij} - \hat{F}_{ij}\right).
$$

Jacobian Matrix of Nonlinear Constraints First-Order Partial Derivatives Call $C(x)$ the nonlinear constraint defining variable x such that $C(x^*) = 0$ at the solution. For example, let $C\left(\beta_n^X\right) = \beta_n^X - \left(1 - \alpha_{n+1}^X\right)\beta_{n+1}^X$. The Jacobian elements for the nonlinear constraints equal the following.

- Sequential production parameters:

$$
\frac{\partial C(\beta_n)}{\partial \beta_n} = 1, \qquad \frac{\partial C(\beta_n)}{\partial \beta_{n+1}} = -\left(1 - \alpha_{n+1}\right), \qquad \frac{\partial C(\beta_n)}{\partial \alpha_{n+1}} = \beta_{n+1}.
$$

  These hold for both intermediate inputs and final goods.

- Price indices:

$$
\frac{\partial C\left(P_j^X\right)}{\partial x} = 1\left[x = P_j^X\right] + \frac{1}{\theta}\frac{P_j^X}{\Theta_j^X}\frac{\partial \Theta_j^X}{\partial x},
$$

  where x is any of the endogenous variables.

- Final good flows:

$$
\frac{\partial C(F_{ij})}{\partial x} = 1\left[x = F_{ij}\right] - \sum_{\boldsymbol{\ell} \in \Lambda_i^N} \frac{\partial \pi_{\boldsymbol{\ell} j}^F}{\partial x} \times \left(w_j L_j - D_j\right) - 1\left[p = w_j\right] \sum_{\boldsymbol{\ell} \in \Lambda_i^N} \pi_{\boldsymbol{\ell} j}^F \times L_j,
$$

  where x is any of the endogenous variables.

- Intermediate input flows:

$$\frac{\partial C\left(X_{ij}\right)}{\partial X_{ij}} = 1,$$

$$\frac{\partial C\left(X_{ij}\right)}{\partial x} = -\sum_{n=1}^{N^X-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\sum_{k\in\mathcal{J}}\left(\beta_n^X\frac{\partial\pi_{\boldsymbol{\ell}k}^X}{\partial x}+1\left[x=\beta_n^X\right]\pi_{\boldsymbol{\ell}k}^X\right)\times\frac{1-\gamma_k}{\gamma_k}w_kL_k-\sum_{\boldsymbol{\ell}\in\Lambda_i^N}\frac{\partial\pi_{\boldsymbol{\ell}j}^X}{\partial x}\times\frac{1-\gamma_j}{\gamma_j}w_jL_j$$

$$-1\left[x=P_h^X\right]\sum_{n=1}^{N^F-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\beta_n^F\times\sum_{k\in\mathcal{J}}\frac{\partial\pi_{\boldsymbol{\ell}k}^F}{\partial x}\times\left(w_kL_k-D_k\right),\ x=\alpha_m^X,\beta_m^X,T_h^X,P_h^X,$$

$$\frac{\partial C\left(X_{ij}\right)}{\partial x} = -\sum_{n=1}^{N^F-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\sum_{k\in\mathcal{J}}\left(\beta_n^F\frac{\partial\pi_{\boldsymbol{\ell}k}^F}{\partial x}+1\left[x=\beta_n^F\right]\pi_{\boldsymbol{\ell}k}^F\right)\times\left(w_kL_k-D_k\right),\ x=\alpha_m^F,\beta_m^X,T_h^F,$$

$$\frac{\partial C\left(X_{ij}\right)}{\partial w_h} = -\sum_{n=1}^{N^X-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\beta_n^X\times\left(\sum_{k\in\mathcal{J}}\frac{\partial\pi_{\boldsymbol{\ell}k}^X}{\partial w_h}\times\frac{1-\gamma_k}{\gamma_k}w_kL_k+\pi_{\boldsymbol{\ell}h}^X\times\frac{1-\gamma_h}{\gamma_h}L_h\right),$$

$$-\sum_{\boldsymbol{\ell}\in\Lambda_i^N}\left(\frac{\partial\pi_{\boldsymbol{\ell}j}^X}{\partial w_h}\times w_j+1\left[h=j\right]\pi_{\boldsymbol{\ell}j}^X\right)\frac{1-\gamma_j}{\gamma_j}L_j,$$

$$-\sum_{n=1}^{N^F-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\beta_n^F\times\left(\sum_{k\in\mathcal{J}}\frac{\partial\pi_{\boldsymbol{\ell}k}^F}{\partial w_h}\times\left(w_kL_k-D_k\right)+\pi_{\boldsymbol{\ell}h}^F\times L_h\right),$$

$$\frac{\partial C\left(X_{ij}\right)}{\partial\gamma_h} = -\sum_{n=1}^{N^X-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\beta_n^X\times\left(\sum_{k\in\mathcal{J}}\frac{\partial\pi_{\boldsymbol{\ell}k}^X}{\partial\gamma_h}\times\frac{1-\gamma_k}{\gamma_k}w_kL_k-\pi_{\boldsymbol{\ell}h}^X\times\left(\frac{1}{\gamma_h}\right)^2w_hL_h\right),$$

$$-\sum_{\boldsymbol{\ell}\in\Lambda_i^N}\left(\frac{\partial\pi_{\boldsymbol{\ell}j}^X}{\partial\gamma_h}\times\frac{1-\gamma_j}{\gamma_j}-1\left[h=j\right]\pi_{\boldsymbol{\ell}j}^X\left(\frac{1}{\gamma_j}\right)^2\right)w_jL_j,$$

$$-\sum_{n=1}^{N^F-1}\sum_{\boldsymbol{\ell}\in\Lambda_i^n\cap\Lambda_j^{n+1}}\beta_n^F\times\sum_{k\in\mathcal{J}}\frac{\partial\pi_{\boldsymbol{\ell}k}^F}{\partial\gamma_h}\times\left(w_kL_k-D_k\right).$$

B.5 Revisiting the Calibration of N

As mentioned in the main text, in this Appendix we show through simulations that there is a precise sense in which recovering the same parameters for $N \geq 2$ implies that the true N is indeed equal to 1 and that $N > 1$ can be rejected. We focus on simulated economies with $J = 5$ countries and fix N at either 1, 2, or 3. For each economy we simulate a set of primitives of the model and compute the general equilibrium. We then take the resulting simulated WIOT entries and estimate the parameters using our MPEC algorithm. Furthermore, for each simulation we run our estimation procedure for various possible values for the number of stages, i.e., $\tilde{N} = 1, 2, 3, 4$. The spirit of the exercise is thus to examine whether our estimation method can successfully recover the true value of N. We simulate 100 economies for each N, and then estimate every economy using each $\tilde{N}$, for a total of 1200 estimations (100 economies for each N, with four estimations each).

Each panel in Figure B.4 summarizes our estimation results for the 100 economies with N stages of sequential production. The x-axis plots the value of the minimized objective function while the y-axis plots the norm between the true parameter values underlying the simulated data and the estimated parameters (note the log-scale). In a nutshell, a lower value in the x-axis implies a better model fit while a lower value in the y-axis implies the estimated parameters are closer to the true parameters. Hence, estimations in the

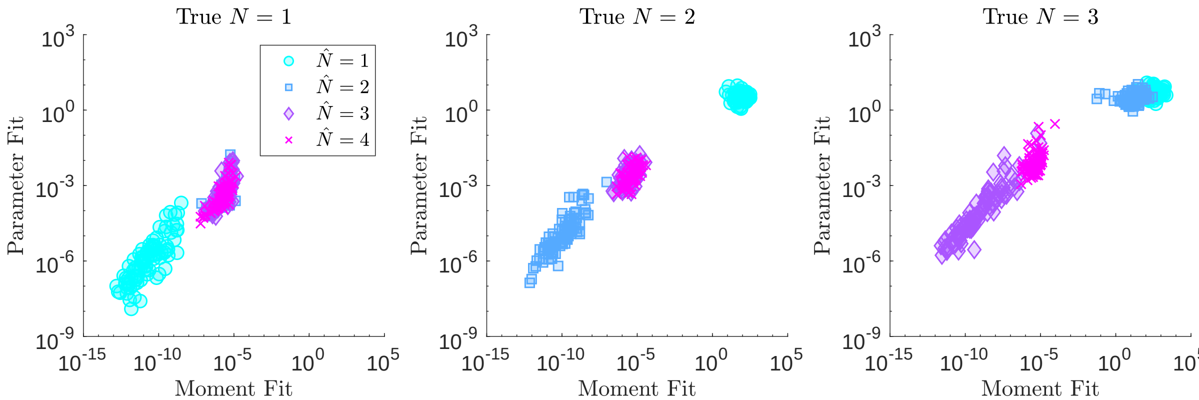

Figure B.4: Calibration of N through simulations.

southwest corner are particularly accurate.

Focus first on the left scatter plot for $N = 1$. The true parameters are recovered in all cases and this is not surprising since the model with $\tilde{N} \geq N$ is more flexible and thus nests model's with lower $\tilde{N}$. In practice, estimations with $\tilde{N} > N$ perform slightly worse than with $\tilde{N} = N$ because of the numerical error involved in shutting down the additional upstream stages of production (i.e., estimating $\alpha_n = 1$ for all $n \leq \tilde{N} - N + 1$). Focus now on the right scatter plot with $N = 3$. In this case, only the estimations with $\tilde{N} = 3, 4$ recover the true parameters. The case with $\tilde{N} = 2$ does better than $\tilde{N} = 1$, but neither recovers the true parameters. The key takeaway from Figure B.4 is that recovering the true N only requires setting $\tilde{N} \geq N$ and the estimation will recover the correct parameters regardless of the particular value of $\tilde{N}$. This is analogous to what occurs in our estimation above, and thus, to the extent that the data generating process underlying the WIOD is consistent with our model, we are able to reject $N > 1$ in the sequential production of final goods.

B.6 Multi-Industry Loss Function

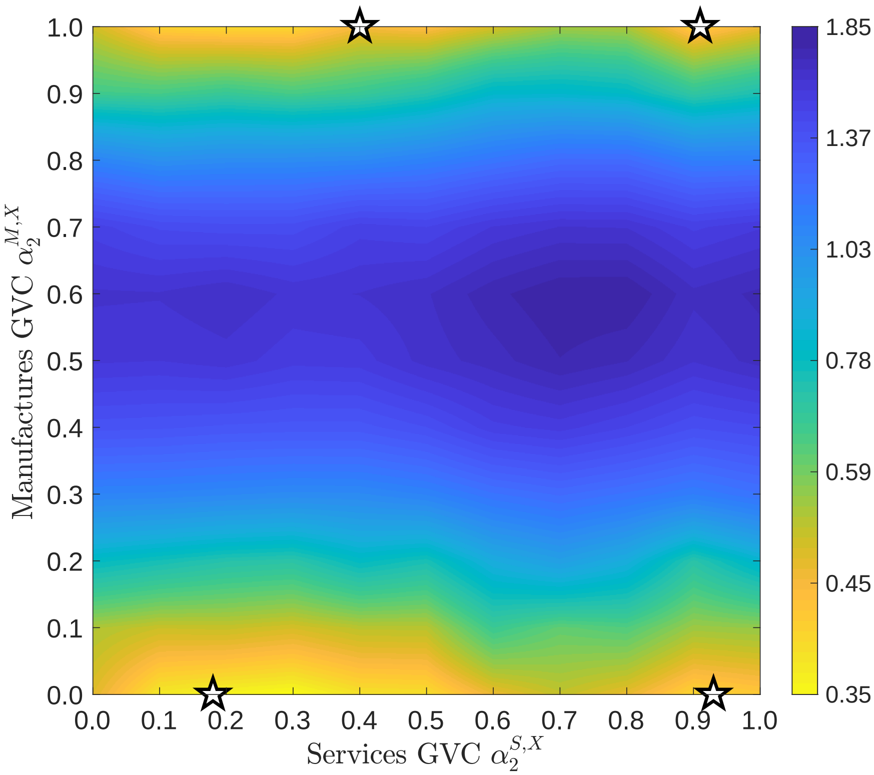

Figure B.5: Multi-Industry Loss Function and Local Minima for $N = 2$.