Production Clustering and Offshoring

Vladimir Tyazhelnikov*[†]

July 2019

Abstract

I introduce a quantifiable model of international production that allows for a production chain of any length, any number of sourcing countries, and weak assumptions on the structure of production and trade costs. Furthermore, the production process does not have to be perfectly sequential, and the final goods can be made from any number of independent subchains. I show that in this model allocation decisions on different stages of production are interdependent, which generates a new channel of proximity-concentration trade-off. The presence of trade costs makes firms cluster their production in certain countries, while trade liberalization allows firms to fragment their production more and exploit productivity differences between countries more efficiently. I then present a general equilibrium heterogeneous firms model in which every firm solves the allocation problem described above. In this model, the distribution of firms' productivity is endogenous with respect to trade costs: trade liberalization leads to a distribution that stochastically dominates the old one, thus leading to an increase in welfare. I use the model to decompose the welfare gains from trade liberalization through two channels: cheaper intermediate inputs and a more efficient production structure. I apply the model to the data and study Chinese joining the WTO. Using the simulated maximum likelihood technique to calibrate the model, I find that a more efficient production structure accounts for approximately 12% of gains from trade.

*I am grateful to Robert Feenstra, Katheryn Russ and Deborah Swenson for their suggestions and support. I also thank Kirill Borusyak, Lorenzo Caliendo, Jaerim Choi, Oleg Itskhoki, Réka Juhász, Andreas Moxnes, Dmitry Mukhin, Stephen Redding, John Romalis, Ina Simonovska, Alan Taylor, and Mingzhi Xu for helpful comments, as well as seminar participants at UC Davis, Warwick Economics PhD Conference 2016, Sonoma State University, European Trade Study Group 2016, Iowa State University, University of Oslo, The University of Sydney, Western Social Sciences Association 2017, Australian National University, Conference on Global Production (National University of Singapore), Australasian Trade Workshop, Society for Advanced Economic Theory, and Moscow International Economics Conference.

[†]The University of Sydney. E-mail: vladimir.tyazhelnikov@sydney.edu.au.

1 Introduction

Production processes are becoming more global — about two thirds of overall world trade accounts for trade in parts and components (Johnson and Noguera (2012)). One of the prevalent modes of global production, the global value chains (GVCs), has recently attracted a lot of attention. The key property of GVCs is that they represent sequential production — intermediate goods are used in the production of other intermediate goods. This paper studies the implications of sequentiality assumption from both the theoretical and the empirical angle.

In this paper, I introduce a model of offshoring where a firm solves such problem for an arbitrarily long production chain. The main novelty of the model is a new channel of proximity-concentration trade-off: due to the presence of trade costs, a firm has to organize its production in clusters, even though some parts within these clusters may be cheaper to produce in another location. Decreased trade costs allow a firm to fragment its production more and exploit productivity differences between countries more efficiently. Production clusters of complementary intermediate parts are easily observable in the managerial practices of multinationals. Frigant and Lung (2002) describe the strategy of modular production and its prevalence in the car manufacturing market. Other examples of production clustering are electronics (Baldwin and Clark (2000)) and the bicycle industry (Galvin and Morkel (2001)).

These predictions are consistent with a sequential production model (the "snake") proposed by Baldwin and Venables (2013). They show that assumptions on production structure matter a lot: the snake and simultaneous production model, they call the spider, generate qualitatively different predictions on trade flows. Snakes and spiders are two limiting cases of perfectly sequential and non-sequential technologies,[1] but these two cases can be too restrictive. I am the first to allow firms to have a more general class of technology, which I call "trees", that nests both spider and snake technologies. A tree can have more than one sequential production subchain. These subchains represent the production technology of complex intermediate parts that are assembled together to become a final good.

I show that in the presence of sequentiality, the problem of choosing optimal production locations becomes nontrivial. Most of the previous literature on GVCs introduced particular structures in order to get the closed form solution of this problem: these restrictions include a number of stages and particular structures of production and trade costs. These restrictions made a closed form solution possible by shutting down the interdependence of stages of production, and hence they ignored the clustering effect I am studying in this paper. The allocation problem is complex, because it cannot be broken down by a sequence of independent decisions at every stage: loca-

[1] In this paper, I use terms "snake" and "sequential production" model and "spider" and "simultaneous production" model interchangeably.

tion decisions at every stage affect costs associated with production decisions on the previous and subsequent stages.

I solve the allocation problem by employing optimal control algorithm, based on the Bellman optimality principle. Conditional on the production location on the previous stage, a decision at the current stage becomes a simple problem. This algorithm, however, works only for the case of the snake technology and specific trade costs, so I propose a modification, which I call "forward induction algorithm", that reverses the order of induction. Now the firm makes a decision conditional on its production location at the next stage. It turns out that this solution to the central planner's problem is the same to the allocation in de-centralized equilibrium, where independent firms are looking for the lowest-price supplier. The reason is that the sufficient statistic to make an optimal decision on every stage is the price of the intermediate good at the previous stage.

Next, I describe firm behavior in snake and tree models. Trade liberalization leads to the lower prices and higher welfare in most of the offshoring models, which is not surprising — even if a firm does not change its production decisions, inputs that a firm is sourcing from abroad are now cheaper. I show that in the case of sequential production, changes in optimal allocation necessarily lead to an increase in production efficiency, no matter whether the new optimal path corresponds to higher or lower spending on transportation. I attribute these gains in production efficiency to higher fragmentation.

My model generates a few other interesting outcomes. First, depending on the destination country of the final good, a firm can choose different optimal paths (for example, Ford Europe and Ford USA). The idea is that there is a trade-off between production efficiency and proximity to consumer markets. If the costs of shipping of the final good are high enough, a firm would prefer to perform later stages of production in the destination countries.

Another implication of production clustering is that trade liberalization between two countries has an ambiguous effect on third countries. The mechanism here is similar to "Bridge Production" from Ramondo and Rodrıguez-Clare (2013); if some parts were cheap to produce in a particular country, but its adjacent stages were produced in a remote country, the likelihood that these parts will be produced in a low-cost country are low. At the same time, as bilateral trade liberalization can lead to the reallocation of some production stages to a country that is closer, offshoring production to the third country in question will make more sense.

Finally, my model is consistent with reshoring, a well-documented phenomenon (Sirkin et al. (2011), Wu and Zhang (2011)), in which a previously offshored part is once again produced domestically. A firm facing high trade costs will choose to offshore a large cluster of production activities. With lower costs, a firm can afford to have smaller clusters and then may choose to reshore production of some parts previously produced in the large cluster. It follows that one should be careful interpreting the impact of reshoring on domestic employment and wages; reshoring in my model is driven by a fall in trade costs and, hence, is accompanied by offshoring of parts previously produced domestically. This mechanism is similar to the one in Harms et al. (2012), which my model espouses.

Further, I combine the firm's problem described above with the heterogeneous firms' model. In order to quantify the model, I employ stochastic production costs formulation: the costs of production at each stage and in each country are a random draw from some probability distribution. After solving the allocation problem, each firm is characterized by marginal costs of production of the final good. The distribution of these marginal costs defines the economy and determines welfare in the model. This distribution is endogenous with respect to the costs of offshoring. I show that in the case of trade liberalization, the new distribution of firms' marginal costs is first-order stochastically dominated by the old one. Trade costs can then be interpreted as frictions that prevent firms from efficiently allocating their resources.

Following this interpretation, I perform an empirical exercise inspired by the rich literature on the identification of resource misallocation started by Hsieh and Klenow (2009)[2]. I analyze the evolution of market shares of Chinese firms pre- and post-WTO. I allow trade costs and technology to change over time, and then, using the structure of the model, I isolate welfare gains that are associated with the change in trade costs only. I further decompose gains from trade liberalization by two channels: cheaper inputs and higher fragmentation.

The purpose of the empirical exercise is twofold. First, I show how micro-level data can be used to quantify the model. Second, the results on the decomposition of gains from trade and comparison of sequential and non-sequential models depend on the parameters of the model. My calibration assures that parameters I am using are realistic and consistent with the data. I find that compared to the simultaneous production model, gains from trade are higher for the case of sequential production when trade costs are low and higher when trade costs are high (FIX). The reason is that the clustering effect makes production decisions sticky when trade costs are high, because reallocation of the whole cluster is problematic, while offshoring just one part in a spider model is easier. As a result, a firm that faces spider technology offshores cheaper parts when trade costs are high, so when trade costs are low, there is not much space for the increase in productivity for the case of the simultaneous production model.

This paper contributes to the rich literature on global value chains – which includes Antràs and Chor (2013), Costinot et al. (2013), Fally and Hillberry (2015), Johnson and Moxnes (2016), Razhev (2015), and Harms et al. (2012), by introducing a quantifiable model of offshoring with a minimal structure on production and transportation costs, number of stages of production, and number of countries, and that it deviates from the assumption of perfect sequentiality.

My model is similar in spirit to Tintelnot (2017) and Antras et al. (2017). Both papers introduce

[2]Wang (2017) applies the Hsieh and Klenow (2009) approach to offshoring.

quantifiable models with complex problems of firms, that do not have a closed form solution, but they can be solved with the help of a numerical algorithm. In Tintelnot (2017) the combination of fixed costs of opening a new plant and variable shipping costs drives the proximity-concentration trade-off. Antras et al. (2017) assume that intermediate inputs are imperfect substitutes or complements, and a firm incurs fixed costs of sourcing inputs from every country. In my model, there are no fixed costs, and I focus on firms' allocation problem that arises in cases when technology exhibits at least some degree of sequentiality.

The clustering mechanism and forward induction algorithm described in this paper are close in spirit to the contemporaneous work by Antràs and De Gortari (2017). There are, however, three main distinctions. First, Antràs and De Gortari (2017) focus on "centrality-downsreamness nexus," a mechanism that leads to the fact that more centrally located countries specialize on more downstream stages of production under the assumption of iceberg trade costs, while I focus on the clustering effect — lower trade costs are associated with higher fragmentation, whether downstream or upstream in the value chain, and I am agnostic about assumptions on trade costs. Second, in their empirical application, Antràs and De Gortari (2017) assume that at every stage a firm considers its productivity in every country, along with trade costs and productivity of its suppliers in every country, so a firm does not account that some suppliers can have an advantage through access to cheaper inputs, thus degenerating clustering mechanism.[3] In my model, a firm accounts for a price each supplier can offer, not for their productivity, thus also accounting for costs of inputs for the suppliers. Finally, while Antràs and De Gortari (2017) use the Ricardian framework for their general equilibrium section of the model, I embed the allocation problem in question to the heterogeneous firms' model.

From the empirical side, my paper contributes to the literature on the link between access to cheap intermediate inputs and firms' productivity (for example Amiti and Davis (2012) and Goldberg et al. (2010)). The main difference is that in my model the productivity gains are driven not only by cheaper intermediate inputs but also through increased fragmentation of the production process.

The rest of the paper is organized as follows: Part 2 describes a firm's cost minimization problem and algorithms to solve it. Part 3 describes properties of firms' behavior. Part 4 introduces the general equilibrium model. Part 5 calibrates the model, estimates the welfare consequences of China joining the WTO, and compares gains from trade for simultaneous and sequential production models. Part 6 concludes.

[3] A model by Johnson and Moxnes (2016) can also generate clustering effects, the authors, however, limit they analysis by the case of 2 stages, thus, also closing this channel.

2 Firms' Problem

In this Section, I provide the solution to a firm's cost minimization problem. For clarity of exposition, I start with the case of a standard sequential production technology and specific trade costs and introduce a backward induction algorithm. Then I extend the problem for the more general case of tree technology, allow for iceberg trade costs, and then present a forward induction algorithm that solves this new problem.

2.1 Sequential Production Model

2.1.1 Setup

There is one firm that produces a final good from N intermediate parts. Each part can be produced in one of M countries. I assume that parts are perfect substitutes, it is not critical for the functioning of the model, but clustering has more intuitive representation for the case of perfect substitutes.[4] Production costs are country- and stage-specific and are equal to a_{ij} where i is a stage of production and j is the country of production. The parts have to be produced in a given order that is determined by the numeration of the stages. Every time the firm chooses to produce a next part in a different country, it pays trade costs $\tau T(j,k)$, where T is the matrix of trade costs with $T(j,j)$=0, $T(j,k) > 0 | j \neq k$, and τ is trade cost scale parameter.[5] τ then can be interpreted as a special tariff. After all the parts are produced, a final good is delivered to a third country, where production is not possible, and shipment costs are the same for all M countries.[6]

The firm then minimizes its per unit costs, which I call marginal costs MC

$$\min_{\{c_i\}_{i=1}^{N}} \sum_{i=1}^{N} \left(\sum_{k=1}^{M} \mathbb{1}\{c_i = k\} a_{ik} + \tau T(c_{i-1}, c_i) \right), \tag{1}$$

where $c_i = j$ if the firm chooses to produce part i in country j, and $c_0 = c_1$.

Notice that the firm cannot break this problem by N independent subproblems for every stage, as the decision at the current stage affects all subsequent decisions. The main idea of this model is a trade-off between clusterization of production and exploring productivity differences between

[4]Notice that parts produced in different countries are not necessarily identical; cost of production in each country can be interpreted as quality-adjusted. I find this deviation from Armington assumption quite realistic for such industries as auto manufacturing and electronics. When a car producer sources input, for example, wheels, it cares about the price and quality of the wheels but does not benefit from a larger variety of wheels sourced from different locations.

[5]In this paper, by trade costs I mean the costs of offshoring. In a narrow sense, it is the costs associated with shipping intermediate goods between countries.

[6]Here I concentrate on production decisions of the firm. In Section 3.3, I allow the firm to sell the final good in one of M countries.

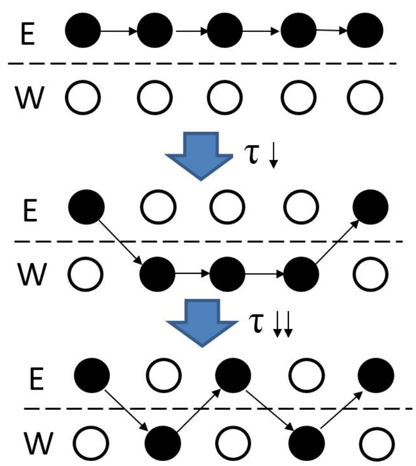

Figure 1: Nonmonotonic Impact of Trade Liberalization

countries.

Figure 1 illustrates the simplest example, with two countries and five stages. Black dots represent firm's choice to produce a part in a given country. The dotted line represents the border between two countries. Arrows represent the firm's optimal path. In this example, firms' optimal choice is a function of trade costs: for high values of τ, the firm chooses to produce the whole good in the East. With lower trade costs it might make sense to offshore a large cluster to the West. Finally, when τ is even lower, the firm breaks its large production cluster in the West and reshores the production of part 3 to the East. A numerical example consistent with this story is provided in Proposition 4.

2.1.2 Algorithm

The firm's choice set includes M^N paths, hence solving (1) with brute force is not feasible, even for moderate values of N and M. As this problem does not have a closed form solution and cannot be solved by brute force, the literature has constrained itself to particular cases of the model that allowed for closed form solution. On the other hand, I propose an optimal control algorithm that can efficiently solve this problem in its general formulation.

Problem (1) can be rewritten in the form of a Bellman equation:

$$V_i(c_i) = \min_{c_i \in M} \left\{ \sum_{k=1}^{M} \mathbb{1}\{c_i = k\} a_{ik} + \tau T(c_i, c_{i+1}) + V_{i+1}(c_{i+1}) \right\}, \tag{2}$$

where $M = \{1, ..., M\}$. This problem can be solved recursively. The algorithm determines for each of M countries at the stage $N-1$ the optimal production location on the stage N. The total costs on both stages of these optimal choices are written down in the value function on the stage $N-1$. Then the same is done at the stage $N-2$ once again for each of M countries to choose where to locate production at stage $N-1$, given the value of the value function in every country at stage $N-1$. The same is done in every stage, until there are M value functions for $c_1 = j,\ j \in M$ that represent optimal trajectories of firms that start in country j. I allow the firm to produce its first part in any country. Then a firm produces the first part in a country associated with lower value of a value function. This path minimizes the costs according to the Bellman principle of optimality. Given that there are just M values of a value function to be stored at every stage of production, and at every stage the algorithm chooses the minimum of these M values for each value of state variable c_i, the number of operations an algorithm has to perform is $M \times N$. [7]

This algorithm works for stage-dependent τ as well (for example, per unit trade costs can be larger for downstream parts).[8] A firm that faces given $1 \times N$ vector of trade costs multiplier τ_i similarly solves the problem (1); the only difference is that now there is a stage-dependent state variable τ_i in the value function. The case of an *ad valorem* tariff, where trade costs depend on the value of transported intermediate good, makes the problem more complicated and is discussed in section 2.3.

Finally, notice that the Bellman equation (2) has a minimum operator at every stage, meaning that either there is a unique optimal choice at every stage, or, if an optimal choice is not unique, a firm is indifferent between two or more sourcing choices. Consequently, the only possible situation when the optimal allocation is not unique is when all non-unique solutions correspond to the same value of total costs. In the stochastic formulation I introduce in Section 4, the probability of this outcome converges to zero, and, for simplicity, from now on I will assume that the optimal path is unique.

2.2 Tree Production Structure

In the previous section, I focused on the sequential production model: I assumed that all parts have to be produced in some exogenously given natural order. This assumption is one of few popular

[7] An alternative way to think about this problem is to interpret it as a tree. Consider a tree of length N and with a choice out of M options in each node. At each stage, the firm makes just one decision: in which country to produce. Assume that when the firm chooses to produce in country j, it chooses to move down along the branch indexed j to the next node. Costs of such a movement are indicated in corresponding nodes. This is a simple decision theory problem with one player, complete information, and absence of randomness. It can be solved by backward induction. Notice that at stage i there are M^i possible choices, but they have only M^2 possible values, so the firm faces a simpler problem: it has to choose what branch it should go to, conditional on its choice on the previous stage. The backward induction algorithm with a tree is more intuitive, but it cannot be directly applied to the more complex problems described in Sections 2.2 and 2.3, so I focus on the Bellman equation interpretation of the problem.

[8] Grossman and Rossi-Hansberg (2008) and Baldwin and Venables (2013) in their extended snake model use trade costs that can vary from task to task, but do not depend on the value of intermediate goods.

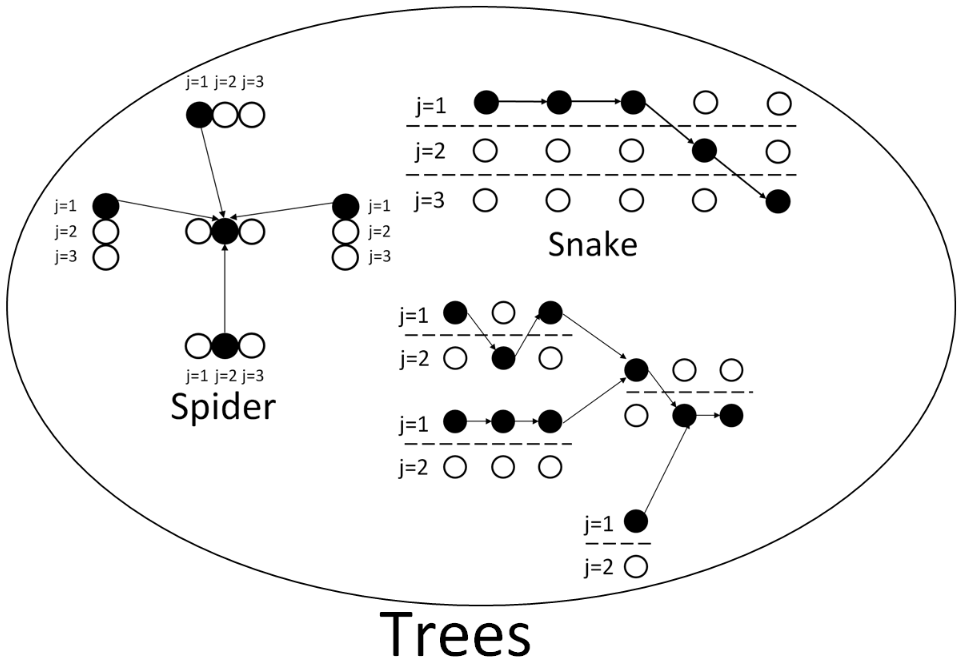

Figure 2: Taxonomy of Technological Assumptions

approaches in the offshoring literature to model a production technology of a complex good. Other popular choices are nonsequential models, where the order of production does not matter (Antras and Helpman (2004), Feenstra and Hanson (1996) and others), and the models where it is costly to produce each part in the country different from the location of the headquarters, as in Grossman and Rossi-Hansberg (2008, 2012).

Tree technology exhibits features of both simultaneous and sequential production technologies. Production of the final good can be represented as a set of sequential subchains that are assembled into more complex intermediate goods. Both snake and spider technologies are particular cases of this tree technology. A tree with one subchain is a snake, and a tree with many subchains of length 1 is a spider. I illustrate the taxonomy of these technological assumptions in Figure 2.

Tree technology relies on only one restrictive assumption: sequentially produced intermediate parts can be assembled together but cannot be disassembled. [9]

[9] This assumption seems reasonable: if a firm made a choice to assemble some parts together, why would it disassemble them and reassemble them later? Surprisingly, this kind of behavior happens in international trade. A classic example is a tariff on light trucks, also known as the chicken tax. In 1963 the United States introduced a 12% tax on light trucks in response to France and Germany increasing their tariffs on U.S. chicken. This tariff has not been changed since, and car producers are using loopholes in order to avoid it. For example, Ford imports its Transit Connect model from its plant in Turkey with rear seats and back windows, so that this vehicle is classified as a wagon and is not subject to the chicken tax. These seats are shredded after Transit Connect crosses the border and windows are replaced with metal panels (http://www.wsj.com/articles/SB125357990638429655). Still, this example is an anecdote rather than a widespread pattern in international trade.

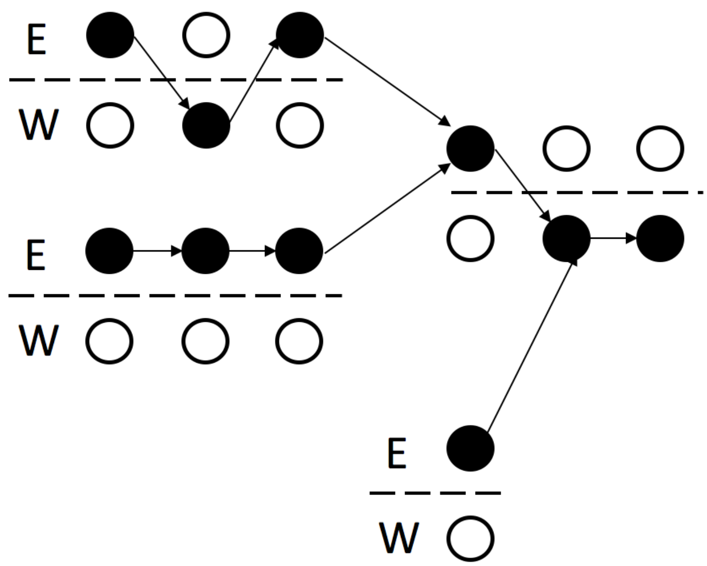

Figure 3: Tree Technology

2.2.1 Technology

I assume that the firm can have an arbitrary number of subchains that can be combined at any stage of production. Figure 3 illustrates the two-country example of such technology: some intermediate goods produced sequentially have to be assembled in the East or in the West at a given stage. Notice that stages do not uniquely identify parts, as more than one subchain can be produced at the same stage. I call a node a stage of production at a given branch, and then there is one-to-one mapping between nodes and parts. An assembly node is a case when the firm combines two or more previously produced parts in one complex part. I imply no restrictions on the number of subchains and the number of assembly nodes.

I assume that intermediate assembly is costly and that these costs of assembly can differ by location. As intermediate parts can be combined but cannot be disassembled, production structure then will look like a tree: there are multiple subchain branches that join at the points of subassembly and become one final-good trunk in the end.

2.2.2 Forward Induction

The algorithm I propose for the baseline model from Section 2.1.2 cannot be directly applied. The problem is that this algorithm uses the production location from the previous stage as a state variable for the given node. In case part i is assembled with L_i intermediate parts, each of which can be produced in one of M countries, it gives L_i^M possible values of the state variable, which can be very large even for a moderate number of countries and assembled parts.

The algorithm I propose here relies on the same optimal control mechanism but reverses the direction of backward induction: rather than moving from the final product to the parts produced in earlier stages, here I choose where to produce a given part for any possible production location of the part produced next. In this case, the state variable at every node will be a production or assembly location in the next stage and, given the tree nature of the technological process, there are M possible values of the state variable, as production or sub-assembly can happen in one of M countries.

The reverse direction of backward induction might seem confusing, so for the sake of clarity, in this subsection I focus on the baseline sequential production model, and explain the intuition behind the algorithm going in the opposite direction.

The firm is solving the problem of allocating N parts in M countries in order to minimize its production costs. It has technological restrictions on the order of production, but it does not have any terminal conditions on the production location of the first or the last stages. Order of production does not imply any direction either: if the firm produces part i in a location different from $i-1$, it has to pay trade costs τ. But it works in both directions: if the firm produces part $i-1$ in a location different from i, it incurs the same costs τ. The expression (1) can then be re-written as

$$MC = \min_{\{c_i\}_{i=1}^{N}, } \sum_{i=1}^{N} \left(\sum_{k=1}^{M} \mathbb{1}\left(c_i = k\right) a_{ik} + \tau T\left(c_i, c_{i+1}\right) \right), \tag{3}$$

where $c_{N+1} = c_N$. The reversed Bellman equation corresponding to (2) is

$$V_i\left(c_i\right) = \min_{c_i \in M} \left\{ \sum_{k=1}^{M} \mathbb{1}\left(c_i = k\right) a_{ik} + \tau T\left(c_i, c_{i-1}\right) + V_{i-1}\left(c_{i-1}\right) \right\}. \tag{4}$$

2.2.3 Algorithm for a Tree

In order to write down the problem, I need to enumerate production nodes. Every node has a unique index ib that represents at what stage i the part is produced and to what branch b it belongs. Production costs for a part from branch b, produced on stage i in country j, are then a_{ibj}.

I assign number $i = 1$ to the last stage of production; $i = 2$ denotes the second-to-last stage, and so on.[10] In case two or more of the intermediate goods are assembled together, each of the corresponding nodes gets the same stage number i; in addition, each of these nodes gets branch index b, which was not previously assigned to another branch.

[10]This enumeration can seem counterintuitive at first, but, due to the tree nature of a problem, such enumeration greatly simplifies the notation.

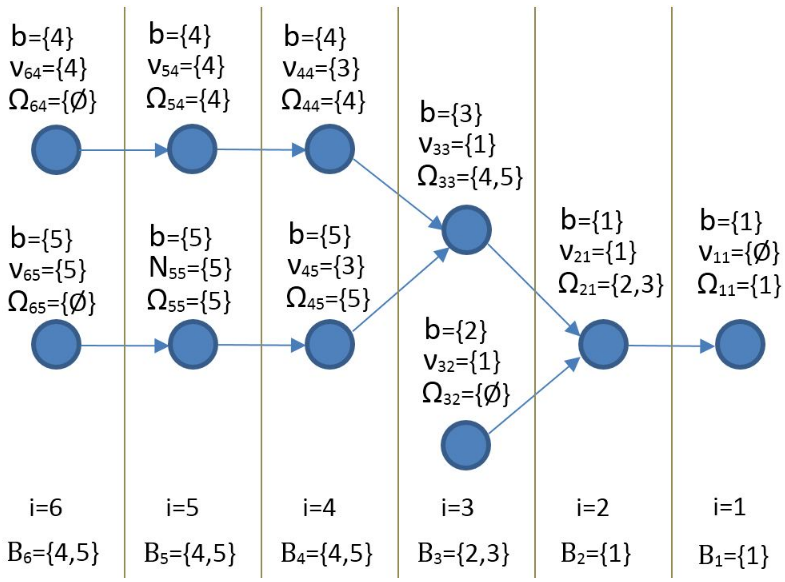

Figure 4: Tree Notation

I define n_b as the last stage of branch b; I call n_b the length of branch b. In addition, for each stage i I introduce an assembly set Ω_{ib}, which is the set of branch indexes b of all parts produced on stage $i+1$, connected to stage ib. v_{ib} is a branch of a part produced at stage $i-1$, a node that ib is connected to. B_i is a set of all branches present at stage i. I present an example of such enumeration in Figure 4.

The tree version of equation (1) (with reversed enumeration) can be written as follows:

$$MC = \min_{\{c_{ib}\}} \sum_{i=1}^{\max\{n_b\}} \sum_{b \in B_i} \left(\sum_{k=1}^{M} \mathbb{1}\left(c_{ib} = k\right) a_{ibk} + \tau T\left(c_{ib}, c_{i-1 v_{ib}}\right) \right). \tag{5}$$

Expressions (1) and (5) look different because of reversed enumeration and due to the more complicated indexing structure of a tree. The main idea remains the same: costs of production of every part depend on the location choice and production location of the next part.[11]

The Bellman equation is then

$$V_{ib}\left(c_{ib}\right) = \min_{c_{ib} \in M} \left\{ \sum_{k=1}^{M} \mathbb{1}\left(c_{ib} = k\right) a_{ibk} + \sum_{l \in \Omega_{ib}} \left[\tau T\left(c_{ib}, c_{i+1l}\right) + V_{i+1l}\left(c_{i+1l}\right)\right] \right\}. \tag{6}$$

For no-assembly nodes, the Bellman equation (6) is similar to equation (2) (the direction of

[11]This is the reason why the reversed algorithm works: every node always has one neighbor node at the next stage but can have multiple neighbors on the previous one.

the algorithm is still reversed): every subchain problem is solved similarly to the baseline chain problem. The difference arises when two or more subchains are combined together. In this case, the value functions for each chain are just added up. The intuition behind it is that all the decisions before the assembly point are made conditional on the location of the assembly. When the location of the assembly is chosen, it should minimize the sum of costs of all subchains being assembled.

Notice that there are separate value functions (with index ib) for different branches. As subchains are assembled together, the value functions add up, and a new value function associated with a new joint branch appears. On the last stage, when the final good is produced, there is one branch (trunk) left, which is associated with a value of the single value function with index $bj = 11$.

2.3 Alternative Trade Costs Functions

When trade costs depend on the value of the intermediate good,[12] the firm's problem becomes more complicated: in order to make an optimal decision, the firm has to know the value of an intermediate good. In this section, I show how this extended problem can be solved by the reversed algorithm.

2.3.1 Algorithm

Under the assumption of iceberg trade costs, every time an intermediate good crosses the border between countries i and j, a fraction $\frac{1}{T^{ice}(i,j)}$ of it "melts," where $T^{ice}(i,i) = 1$ and $T^{ice}(i,j) > 1$ for $i \neq j$.[13] In other words, the firm has to ship $T^{ice}(i,j)$ units of an intermediate good from country i to receive one unit of an intermediate good in country j. As the firm minimizes its per unit costs, I reformulate the problem as follows: a firm produces 1 unit of each intermediate part until it chooses to cross the border. Whenever a border crossing between countries i and j happens, a firm has to produce $T^{ice}(i,j)$ times more of all inputs previously produced, and hence, the transportation costs it incurs is $\left(T^{ice}\left(c_{ib}, c_{i-1v_{ib}}\right) - 1\right)\chi_{ib}$, where χ_{ib} is the cost of an intermediate good crossing the border.

Under the assumption of *ad valorem* tariff, every time an intermediate good crosses the border between countries i and j, it has to pay share $\tau T(i,j)$ of the costs of the intermediate good, where $\tau \geq 0$, $T(i,i) = 0$, and $T(i,j) > 0$ for $i \neq j$. The costs to cross the border will then be $\tau T\left(c_{ib}, c_{i-1v_{ib}}\right)\chi_{ib}$.

One can see that in case $T^{ice}\left(c_{ib}, c_{i-1v_{ib}}\right) - 1 = \tau T\left(c_{ib}, c_{i-1v_{ib}}\right)$, the cost of crossing the border coincide for iceberg and *ad valorem* cases. It means that for any value of iceberg trade costs, an

[12]This is a popular assumption in the offshoring literature. For example, Ramondo and Rodrıguez-Clare (2013), Johnson and Moxnes (2016), and Yi (2010) use iceberg trade costs.

[13]Here, for simplicity I omit the trade costs multiplier τ because the matrix T^{ice} has ones on its main diagonal. The correct way to include τ would be: $\left(T^{ice} - 1\right)\tau + 1$.

equivalent *ad valorem* tariff can be found, such that firm's optimal choice and marginal costs will be the same.[14,15]

The costs of the firm then will depend on the quantity of intermediate inputs it has to produce and can be written down in terms of *ad valorem* tariff as:

$$MC = \min_{\{c_{ib}\}} \sum_{i=1}^{\max\{n_b\}} \sum_{b \in B_i} \sum_{k=1}^{M} \mathbb{1}\left(c_{ib} = k\right) a_{ibk} \chi_{ib} \tag{7}$$

$$\chi_{ib} = \chi_{i-1 v_{ib}} \left(\tau T\left(c_{ib}, c_{i-1 v_{ib}}\right) + 1\right), \chi_1 = 1, \tag{8}$$

where χ_{ib} can be interpreted as the quantity of intermediate inputs that has to be produced at stage ib for iceberg trade costs and as the cumulative tariff rate at stage ib for the *ad valorem* tariff; with multiple border crossings, some parts are taxed more than once.

The expression for the Bellman equation is then straightforward:

$$V_{ib}\left(c_{ib}\right) = \min_{c_{ib} \in M} \left\{ \sum_{k=1}^{M} \mathbb{1}\left(c_{ib} = k\right) a_{ibk} + \sum_{l \in \Omega_{ib}} \tau T\left(c_{ib}, c_{i+1l}\right) V_{i+1l}\left(c_{i+1l}\right) \right\}. \tag{9}$$

The only difference between this Bellman equation and equation (6) is that trade costs now depend on the value of the intermediate good at stage ib, and it is equal to the sum of values of value functions at the previous stage.

2.4 Decentralized Solution

The logic of forward induction suggests that the central planner's problem and a decentralized solution will lead to the same allocation. As noted above, the value of the value function at every stage is equal to the cost of the intermediate good. In other words, the price of an intermediate good at a given point is the sufficient statistic, and the firm does not need to know all the previous path to make an optimal decision.

Decentralized solution can then be represented as the following environment: there are independent firms at every stage in every country. Each of these firms is choosing the lowest-cost supplier, that can offer them the lowest price of an intermediate product that includes costs of shipping and depends on the costs of production and the costs of inputs that each supplier uses. As a result, each potential supplier makes an optimal choice but does not necessarily have a buyer. In

[14] Here I focus on the fact that iceberg trade costs and *ad valorem* tariffs lead to the same optimal path of an individual firm. Still, they can lead to different equilibrium outcomes due to different effects on labor markets and tariff revenue.

[15] In case of nonconstant returns to scale, the solutions for ad-valorem tariff and iceberg trade costs will be different.

the end, only firms that have buyers survive, and the optimal path is determined by a sequence of surviving firms.

Predictions of the model hence does not depend on the boundaries of the firm, and they hold for any exogenously given organizational structure.

3 Properties of the Model

There is no simple analytical solution for the general cases of chain and tree technology; still, some properties of firms' behavior can be derived. In this section, I introduce theoretical results that hold for any chain and tree problem. Mostly I focus on comparative statics with respect to τ, a single parameter reflecting openness for trade. Changes in τ can be interpreted as multilateral trade liberalization, or more generally as a proportionate decrease in the costs of offshoring.

3.1 Multilateral Trade Liberalization

Proposition 1. *In optimum firm's total costs are nondecreasing in trade costs.*[16]

Proof. Let's assume there is an optimal path A for τ_0, an optimal path B for τ_1, $\tau_0 > \tau_1$, and $MC(A,\tau_0,C) < MC(B,\tau_1,C)$, where C is a vector of the costs of production. Notice that $MC(A,\tau_0,C) \geq MC(A,\tau_1,C)$ and by optimality of B: $MC(A,\tau_1,C) \geq MC(B,\tau_1,C)$. It follows that $MC(A,\tau_0,C) \geq MC(B,\tau_1,C)$, which contradicts the initial assumption. $\square$

This proposition is straightforward: when the firm faces lower trade costs, if it does not change its production decision, it will face the same or lower total costs of production. Then there is no way a new optimal path is more costly than the old one. Notice that the proof does not use any assumption on a production structure and relies on firm's revealed preferences argument. It means that this result is going to hold for a large class of firm's problems.

Now I decompose the costs of the firm. Let function $NTMC(Y) \equiv \sum_{i=1}^{N} \left(a_{ki} 1\{c_i = k\}\right)$ be a value of nontransport marginal costs for path Y. Let $TTMC \equiv MC - NTMC$ be a value of total trade costs that can be represented as $TTMC = \tau TQ$. I call TQ transportation quantity, as it reflects

[16] Here and further, I state propositions and theorems with weak monotonicity, which happens for two reasons: first, if trade costs are so high that offshoring is impossible, some of comparative statics related to offshoring do not work. Second, the firm has a finite number of optimal choices, and then the firm's optimal choice cannot change with any infinitesimal change in a parameter value. To handle the first problem, it is enough to assume that trade costs are not very large and offshoring is possible. The second problem goes away when a large number of firms are taken into consideration: with a continuum of firms, changes in parameter values lead to a change in the optimal path for at least some of the firms.

transportation schedules independent of the trade costs price shifter τ. One can think about TQ as the total number of miles a transportation ship traveled, and τ as the price of gas.[17]

Lemma 1. *The transportation quantity is a nonincreasing function of* τ.

Proof. Let path A with transportation quantity $TQ(A)$ be chosen for $\tau = \tau_0$ and path B with transportation quantity $TQ(B)$ be chosen for $\tau = \tau_1$ and $\tau_0 > \tau_1$. Now assume that the transportation quantity is an increasing function of τ and hence $TQ(A) > TQ(B)$. Then given the choice that the firm made under τ_1: $NTMC(B) + TQ(B)\tau_1 < NTMC(A) + TQ(A)\tau_1$ and under τ_0: $NTMC(B) + TQ(B)\tau_0 > NTMC(A) + TQ(A)\tau_0$. Adding $TQ(B)(\tau_0 - \tau_1)$ to the first inequality, I get: $NTMC(B) + TQ(B)\tau_0 < NTMC(A) + TQ(A)\tau_1 + TQ(B)(\tau_0 - \tau_1) < NTMC(A) + TQ(A)\tau_1 + TQ(A)(\tau_0 - \tau_1) = NTMC(A) + TQ(A)\tau_0$ or $NTMC(B) + TQ(B)\tau_1 < NTMC(A) + TQ(A)\tau_1$, which contradicts the condition on optimality of A under τ_0. □

The intuition behind this proposition is the following: when the price of gas decreases, the firm does not have incentives to decrease the number of miles traveled, even though total transportation expenses can increase or decrease.[18]

Proposition 2. *Provided there is some offshoring, the firm's optimal total costs are increasing in trade costs.*

Proof. Let's assume $\tau_0 > \tau_1$. Let A be an optimal path for $\tau = \tau_0$ and transportation quantity $TQ(A) > 0$. Then by Proposition 1 $MC(A, \tau_0) > MC(A, \tau_1)$. Let B an optimal path for τ_1, then by definition of optimal path $MC(A, \tau_1) \geq MC(B, \tau_1)$, and hence $MC(A, \tau_0) > MC(B, \tau_1)$. □

Proposition 3. *If the firm changes its unique optimal path due to decrease in* τ*, then nontransportation costs of production (NTMC) decrease.*

Proof. Let A be an optimal path for τ_0, B an optimal path for τ_1, $\tau_0 > \tau_1$, and $A \neq B$. By definition of optimality and because of the uniqueness of optimal paths, $MC(A, \tau_0) < MC(B, \tau_0)$ and $MC(A, \tau_1) > MC(B, \tau_1)$. From Lemma 1 $\tau_1 TQ(A) < \tau_1 TQ(B)$. Assume $NTMC(A) < NTMC(B)$, then $MC(A, \tau_1) = NTMC(A) + \tau_1 TQ(A) < NTMC(B) + \tau_1 TQ(B) = MC(B, \tau_1)$, which contradicts the optimality of B under τ_1. □

[17] In case transportation costs are similar for all country pairs $T_{ij} = T_{kl}$ for $\forall i, j, k, l \in \{1, ..., M\}, i \neq j, k \neq l$, τ can be interpreted as the number of border crossings.

[18] In case of similar transportation costs between all country pairs, Lemma 1 means that the number of border crossings is a nonincreasing function of τ. Then by defining a cluster as a sequence of parts produced in the same country, the following statement is true: The average size of a cluster is a nondecreasing function of τ. It simply follows from the fact that the average size of a cluster is equal to $s = \frac{N}{m+1}$, where m is a number of border crossings. As by Lemma 1, m is nonincreasing in τ and, hence, s is nondecreasing in τ.

This is the key proposition that represents gains from fragmentation. Not surprisingly the firm increases its total productivity when trade costs are decreasing: if the firm is engaged in offshoring, it pays less for transportation. But Proposition 3 shows that there is another channel through which productivity increases: the optimal path of the firm depends on trade costs. With a change in trade costs, the firm can choose a different production structure that would lead to higher efficiency of production. Similarly to Proposition 1, this result does not rely on sequentiality of the production structure.

Proposition 4. *Production in a given country can depend on trade costs nonmonotonically. Reshoring is possible.*

Proof. Consider the following numerical example with sequential technology, two countries and five stages of production: $a_1 = \{4,4,4,4,4\}$, $a_2 = \{10,2,5,2,10\}$. Then □

1. If $\tau \geq 1.5$, $c = \{1,1,1,1,1\}$
2. If $1.5 > \tau \geq 0.5$, $c = \{1,2,2,2,1\}$
3. If $0.5 > \tau \geq 0$, $c = \{1,2,1,2,1\}$

Figure 1 illustrates Proposition 4. Black dots represent the firm's choice to produce a part in a given country. The dotted line represents the border between two countries. Arrows represent the firm's optimal path. With high trade costs, the firm chooses to produce the whole good in the first country. With decreased trade costs, it chooses to offshore a large cluster to the second country. When trade costs decreased even further, the firm fragments its production more and reshores the third part back to the first country.

3.2 Limiting Cases

In the general case, there is no closed form solution for the marginal costs of production of the final good in problems (1) and (3). The reason is not a drawback of some modeling assumptions:[19] the interdependence of production on different stages both generate clustering effect and make a problem hard to solve. In Section 2.1, I show that the interdependence of decisions on different stages of production leads to a complicated solution; the optimal path depends on the value of $M \times N + 1$ parameters: costs of production and trade costs. The solutions for limiting cases, however, are trivial:

[19] One reasonable way to simplify the problem is to assume that the firm does not know its costs at a given stage until it builds a plant. This assumption makes the problem trivial: the firm will produce all the parts in the country that has lower *ex ante* costs. In case shipment costs differ for some countries, the firm can face a trade-off between production efficiency and proximity to consumer markets described in Section 3.3; still, under this assumption the clustering mechanism remains redundant.

1. If $\tau=0$, $MC=\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}\min_{j\in M}\left\{a_{ibj}\right\}$

2. If $\tau=\infty$, $MC=\min_{j\in M}\left\{\sum_{i=1}^{N}\sum_{b\in B_i}a_{ibj}\right\}$.

In the free trade case, the firm just chooses to produce each part in the cheapest location. In case $\tau=\infty$, offshoring of parts is impossible, and the firm chooses the cheapest location to produce the whole good. Notice that $\min_{j\in M}\left\{\sum_{i=1}^{N}\sum_{b\in B_i}a_{ibj}\right\}\geq\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}\min_{j\in M}\left\{a_{ibj}\right\}$ with an equality sign only in case there exists such country k that $a_{ibj}\geq a_{ibk}$ for $\forall i,b,j$. These two limiting cases represent two states of the Ricardian world: when countries specialize in production of parts and in production of final goods. When the trade costs decrease, more opportunities arise to exploit productivity differences between countries:

Proposition 5. *For any* $\tau\in(0,\infty)$, $\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}\min_{j\in M}\left\{a_{ibj}\right\}\leq MC(a,\tau)\leq\min_{j\in M}\left\{\sum_{i=1}^{N}\sum_{b\in B_i}a_{ibj}\right\}$.

Proof. Follows directly from Proposition 1. □

In particular, it means that every firm has a limited potential to gain from offshoring: gains in production efficiency of the firm are limited by $\min_{j\in M}\left\{\sum_{i=1}^{N}\sum_{b\in B_i}a_{ibj}\right\}-\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}\min_{j\in M}\left\{a_{ibj}\right\}$.

3.3 Vertical and Horizontal FDI

At the beginning of this section, I assumed that the final good is exported to a third country. I did so in order to isolate the effects of vertical and horizontal FDI. In this subsection, the firm can sell the final good to countries $j\in\{1,\ldots,M\}$. In this case the cost minimization problem is interacted with proximity to consumer market consideration. A firm can choose different optimal paths for production of final goods with different destination countries.[20]

A firm would have an independent cost minimization problem for each destination country:

$$MC_d=\min_{\{c_i\}_{i=1}^{N},}\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_j}MC_{ib}+\tau^F T^F\left(c_N,d\right)$$
$$MC_{ib}=\sum_{k=1}^{M}\left(\mathbb{1}\left(c_{ib}=k\right)a_{ibk}+\tau T\left(c_{ib},c_{i-1\nu_{ib}}\right)\right),$$

[20]Here the firm does not face a complicated export-platform problem as in Tintelnot (2017), because in this model there are no fixed costs of opening a plant and the firm just solves the horizontal FDI problem independently for each destination country.

where d is the index of destination country, $T^F(i,j)$, and τ^F is a final good shipment-costs matrix and multiplier correspondingly. I introduce τ^F for two main reasons: first, tariffs on final and intermediate goods can be different, and second, the costs of offshoring may include other factors besides the costs of transportation and tariffs.

A firm chooses an optimal path to minimize the sum of production costs and shipment costs. Depending on the relative size of τ and τ^F, the firm will assign different weights to vertical and horizontal FDI considerations.

Proposition 6. *For large enough τ^F, optimal paths with different destinations are different.*

Proof. For $\tau^F > \sum_{j=1}^{M}\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i} a_{ij}$ the firm chooses to produce the final part in the destination country. □

Notice that only one optimal path minimizes the costs of production (or the firm is indifferent between more than one path).

Corollary. *For large enough τ^F, an optimal path does not minimize the marginal costs of production.*

3.4 Effects of FTA

In this section, I show that bilateral trade liberalization in a multicountry world can lead to unexpected results.

Proposition 7. *In the case of three countries, trade liberalization between two countries may increase production in a third country.*

Proof. A numerical example can be provided. Assume there are three countries, country 1 is far from countries 2 and 3, and countries 2 and 3 are close: $\tau_{12}^0 = \tau_{21}^0 = 2$, $\tau_{13} = \tau_{31} = 2$, $\tau_{23} = \tau_{32} = 0.5$. The final good consists of three parts and their costs of production are equal: $a_{11} = 2$, $a_{12} = 7$, $a_{13} = 2$ for country 1; $a_{21} = 8$, $a_{22} = 5$, $a_{23} = 8$ for country 2; and $a_{31} = 8$, $a_{32} = 8$, $a_{33} = 2$ for country 3. Then the cost minimizing decision would be to produce all parts in the first country. Now if trade costs between countries 1 and 2 decrease $\tau_{12}^1 = \tau_{21}^1 = 1$, then the optimal decision is to produce the first part in country 1, second in country 2 and third in country 3. So, a decrease in trade costs between countries 1 and 2 increased production in country 3. □

The third country benefits because before the trade liberalization, the costs of production of part 3 in country 3 were low, but not low enough to make offshoring of this part to country 3 profitable, because of high trade costs. With the decrease in trade costs, production of part 2 became offshored to country 2, but as parts 2 and 3 became adjacent, trade costs between countries

1 and 3 no longer matter, and the firm faces lower trade costs between countries 2 and 3. I provide the following example: with high trade costs, a U.S. firm chooses not to offshore its production. With the decrease in trade costs with Malaysia, this firm may want to offshore some stages of production there. But as these parts are offshored, there may be an advantage to offshore adjacent parts to Indonesia, which is close to Malaysia.

3.5 Endogenous Wages

The problem presented above is the model of absolute advantage as there is no labor market. With a given supply of labor in each country L_j and endogenous wages that are determined through labor market clearing conditions, all countries will produce some parts no matter what production costs are.[21] I normalize the wage in country 1 to $w_1 = 1$. I assume that labor supply is perfectly inelastic and the firm has constant returns to scale production technology. The problem of every firm then looks like

$$MC = \min_{\{c_i\}_i^N,\, i=1} \sum^{I} \left(w_j \mathbb{1}\left(c_i = k\right) a_{ik} + \tau T\left(c_i, c_{i-1}\right)\right), \tag{10}$$

and a firm's labor demand per unit produced is

$$L_{Dk} \equiv \sum_{i=1}^{I} \mathbb{1}\left(c_i = k\right) a_{ik} \text{ for } \forall k \in \{1, ..., M\}.$$

Here for simplicity I assume that transportation services are performed by independent transport companies and do not affect domestic and foreign labor markets.

Lemma 2. *Demand of a firm from country i L_{Di} for labor in country k is a nonincreasing function of w_k.*

Proof. Let the wage in country k decrease, while all other wages remain constant: $w_k^A > w_k^B$ and $w_{j\neq k}^A = w_{j\neq k}^B = w_{j\neq k}$. Let A and B be optimal paths under wage schedules w^A and w^B. In case $A = B$, $L_{Dk}^A = L_{Dk}^B$. Now consider the case $A \neq B$. Then because of the optimality of A and B: (a) $MC\left(A, w^A\right) < MC\left(A, w^B\right)$ and (b) $MC\left(B, w^B\right) < MC\left(A, w^B\right)$. Let $\Delta^{\tau VT} \equiv \tau VT\left(A\right) - \tau VT\left(B\right)$, and $\Delta^L \equiv \sum_{j\neq k} w_j \left(L_{Dj}^A - L_{Dj}^B\right)$. Then (a) and (b) can be rewritten as: $L_k^A w_k^A - L_k^B w_k^A + \Delta^L + \Delta^{\tau VT} < 0$, and $L_k^A w_k^B - L_k^B w_k^B + \Delta^L + \Delta^{\tau VT} > 0$, subtracting the first inequality from the second obtains: $\left(L_{Dk}^A - L_{Dk}^B\right)\left(w_k^B - w_k^A\right) > 0$, and then $L_{Dk}^A < L_{Dk}^B$. □

Notice that if a firm changes its optimal path, then L_{Dk} is decreasing in w_k.

[21] As long as trade costs are not too high for a given firm.

Proposition 8. *There exists a wage schedule that clears the labor market. In a two-country case, this schedule is unique.*

Proof. Existence:

The world economy can be considered as an exchange economy with M agents, where labor supply in country k is the endowment of good k and wage in country k is the price of this good. Then from Lemma 2 demand of each agent for each good is nondecreasing in price of this good, so by proposition 17.C.1 in Mas-Colell et al. (1995) an equilibrium exists.

Uniqueness for the case of two countries:

A firm's relative demand for labor $\frac{L}{L^*}$ is a nonincreasing function of the relative wage w. Every firm takes the wage as given, but decisions of a firm determine the wage through market clearing condition. Here once again I apply the revealed preferences argument. Let's assume there is path A with $\sum_{i=1}^{N} c_i a_{Wi} = R_{WA}$ and $\sum_{i=1}^{N} (1-c)_i a_{Ei} = R_{EA}$ that was chosen for $w = w_0$ and there is path B with $\sum_{i=1}^{N} c_i a_{Ni} = R_{WB}$ and $\sum_{i=1}^{N} (1-c)_i a_{Si} = R_{EB}$ that was chosen for $w = w_1$; $w_1 > w_0$ and $R_{WB} > R_{WA}$. Let function $NPC(Y)$ be a value of nonproduction costs for path Y, then given the choice that the firm made under w_0: $NPC(A) + w_0 R_{WA} + R_{NA} < NPC(B) + w_0 R_{WB} + R_{EB}$ and under w_1: $NPC(A) + w_1 R_{WA} + R_{EA} > NPC(B) + w_1 R_{WB} + R_{EB}$. Adding $R_{WA}(w_1 - w_0)$ to the both parts of the first inequality, I get: $NPC(A) + w_1 R_{WA} + R_{EA} < NPC(B) + w_0 R_{WB} + R_{EB} + R_{WA}(w_1 - w_0) < NPC(B) + w_1 R_{WB} + R_{EB} < NPC(B) + w_1 R_{WB} + R_{EB}$ or $NPC(A) + w_1 R_{WA} + R_{EA} < NPC(B) + w_1 R_{WB} + R_{EB}$, which contradicts the condition of optimality of B under w_1.

□

For the case of multiple countries, proof of uniqueness of the equilibrium is nontrivial: decrease in the wage in one country can increase demand for labor in another country through the bridge FDI channel, similar to Proposition 7. As a result, the gross substitute property does not hold, and the uniqueness cannot be proven using the approach of Allen et al. (2015).

4 General Equilibrium

In this part, I describe a heterogeneous firms model of M countries, where each of a continuum of monopolistically competitive firms produces a complex good that consists of N parts. Each firm faces the same tree technology, determining the order of production, but can have different costs of production in each country.

The main idea of this general equilibrium formulation is that after each firm solved its problem, for a given level of trade costs, this firm's productivity in production of a final good is a sufficient statistic. With the addition of stochastic production costs, I can close the model. This model, which can be solved numerically, is embedded in stochastic empirical estimation.

I use the simplest version of a monopolistic competition model, with quasilinear utility, a homogeneous good sector, and with the absence of fixed costs of production and exporting. I do it in order to eliminate all gains from trade channels not related to the clusterization, so that I can illustrate how the new mechanism operates in a general equilibrium environment.[22]

4.1 Demand

There is a representative consumer in each country who consumes a homogeneous good q_0, and a continuum of differentiated products with real consumption index Q. The real consumption index of differentiated product is the CES aggregator

$$Q = \left[\int_{\omega\in\Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}}, \sigma > 1,$$

where $q(\omega)$ is consumption of variety ω, Ω is the set of varieties available for consumption, and σ is the elasticity of substitution between the varieties.

Preferences between the homogeneous product q_0 and consumption aggregate Q are described by the quasilinear utility function

$$U = \frac{1}{\zeta} Q^{\zeta} + q_0, 0 < \zeta < \frac{\sigma-1}{\sigma} < 1,$$

which is a real consumption index. Assumption $\zeta < \frac{\sigma}{\sigma-1}$ guarantees that heterogeneous varieties are closer substitutes between each other compared to the homogeneous good. I assume that the consumer has a large enough income to consume a positive quantity of the numeraire good.

Then, the consumer who faces the price index $P = \left[\int_{\omega\in\Omega} p(\omega)^{1-\sigma} d\omega\right]^{\frac{1}{\sigma-1}}$ would have consumption of aggregated product $Q = P^{-\frac{1}{(1-\zeta)}}$ and $q_0 = E - P^{-\frac{\zeta}{(1-\zeta)}}$. The indirect utility function is then

$$W = E + \frac{1-\zeta}{\zeta} P^{-\frac{\zeta}{1-\zeta}},$$

and demand function for good ω is

$$q(\omega) = P^{\frac{\zeta-\sigma}{1-\zeta}} p(\omega)^{-\sigma}. \tag{11}$$

[22] Computational time does not depend on the complexity of the model, and the firm's problem from Section 3 can be combined with richer heterogeneous firms models.

4.2 Technology

All the problems described in Section 2 took costs of production as a given. It is unlikely, however, to get the data on costs in each particular stage of production for each firm. Moreover, to solve this problem, one would need to know not only actual costs of production but also opportunity costs of production of these parts in other countries. Following Yi (2010), Ramondo and Rodrıguez-Clare (2013) and Johnson and Moxnes (2016), I assume that costs of production at each stage in each country follow some random variable. The standard assumption is the Fréchet distribution, popular because it leads to the closed form solution of many models. Fréchet, however, is not the only possible choice; and in this section, I am not making distributional assumptions, in order to make my analysis as general as possible.

Every firm draws $N \times M$ matrix A of costs from some distributions $F_j(a)$, $j \in M$. Here I assume that these draws are i.i.d. Facing the cost matrix A and trade costs τ, the firm solves its problem and has optimal marginal costs that depend on production and transportation costs $MC(A,\tau)$. As elements of A are random variables, $MC(A,\tau)$ is a random variable as well; distribution of optimal marginal costs is then a function of parameters θ_j of distribution $F_j(a)$, $j \in M$: $G_{MC}(\theta,\tau)$, where $\theta = \{\theta_1, ..., \theta_M\}$.

Proposition 9. *If $\tau_1 < \tau_0$, random variable $MC(\theta,\tau_0)$ weakly*[23] *first-order stochastically dominates $MC(\theta,\tau_1)$.*

Proof. By Proposition 1, $MC(A,\tau)$ is nondecreasing in τ. It means that for any given matrix of draws A $Pr(MC(A,\tau_0) < x) \leq Pr(MC(A,\tau_1) < x)$. At the same time, by definition of $MC(\theta,\tau_0)$, random variables $MC(\theta,\tau_0)$ and $MC(A,\tau_0)$ follow the same distribution. It means that $Pr(MC(A,\tau) < x) = Pr(MC(\theta,\tau_1) < x)$. And hence $Pr(MC(\theta,\tau_0) < x) \leq Pr(MC(\theta,\tau_1) < x) \Rightarrow MC(\theta,\tau_0)$ first order stochastically dominates $MC(\theta,\tau_1)$ $\square$

Corollary. *Distribution $MC(\theta,\tau)$ first-order stochastically dominates $MC(\theta,\tau_0)$ and is dominated by $MC(\theta,\tau_1)$ for $\tau \in (\tau_0,\tau_1)$.*

In particular, $MC(\theta,\tau)$ is bounded by two well defined distributions: $\sum_{i=1}^{\max\{n_b\}} \sum_{b\in B_i} \min_{j\in M}\{a_{ibj}\}$ and $\min_{j\in M}\left\{\sum_{i=1}^{N}\sum_{b\in B_i} a_{ibj}\right\}$.

The proof of Proposition 9 relies on Proposition 1, where the firm can at no cost switch between optimal paths and does not account for sunk costs the firm paid to organize its production. Under the presence of sunk costs, I would interpret Proposition 1 in the way in which new entrants or incumbents that expand their production have access to the better technology.

[23]The weak dominance appears in the case of large τ_0 and τ_1 such that there is no offshoring. In this case, changes in trade costs do not affect productivity of the firms.

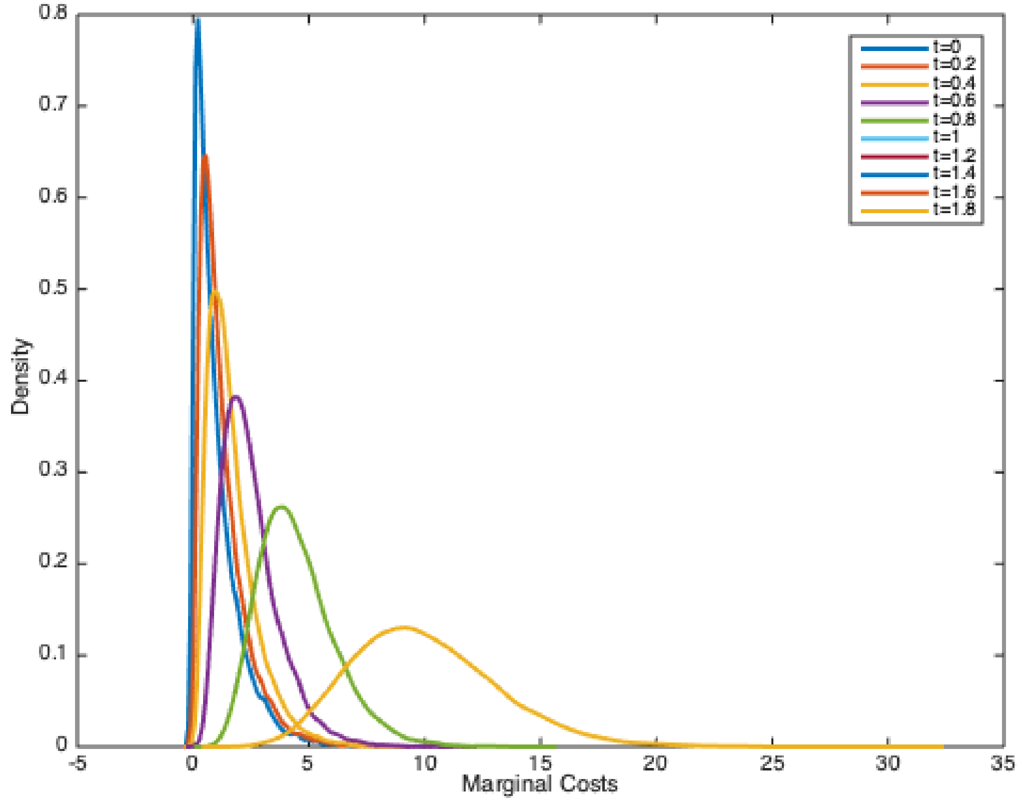

Figure 5: Evolution of Firms' Marginal Costs

Figure 5 illustrates Proposition 9. I simulate 100,000 firms with $N = 10$ and Fréchet distribution of the draws. Notice that the distribution of firms' marginal costs does not just shift, it changes its shape. The change in the shape of the distribution leads to two consequences: first, it lowers costs for each firm on the market; second, it increases profits of every firm and hence expected profit of potential entrants, which would lead to larger entry. Both effects lead to an increase in welfare, through lower prices and greater variety.

4.3 Market Structure

There are two sectors in each country: a homogeneous good sector and a monopolistically competitive sector. A homogeneous good is produced in a competitive market with a unit productivity one, so the wages in the economy are normalized to 1.[24] Firms in a monopolistically competitive sector pay sunk costs f_s to enter the market. In order to separate the effect of trade costs on vertical

[24]In Section 3.5 I show how to endogenize wages and drop homogeneous sector assumption. In this case wages will be determined by exogenously given population in both countries.

and horizontal FDI, I introduce different trade costs of offshoring and shipment costs of a final good. Costs of offshoring are τ, and shipment costs are τ^F.

There is a continuum of firms of mass S. Each of these firms indexed by ω solves the cost minimization problem (5) for individual vectors of production costs and common trade costs parameter τ. Production costs $A(\omega)$ are drawn from distributions $F_j(a)$, $j \in M$ with parameters θ_j. Marginal costs of each firm $MC(\omega)$ are then a random variable and follow endogenous distribution $G_{MC}(\theta, \tau, d)$, where d is the destination country of a final good.[25]

Due to CES utility, each firm sets a constant markup:

$$p(\omega) = \frac{\sigma}{\sigma - 1} MC(\omega). \tag{12}$$

4.4 Equilibrium and Welfare

Here I assume $\tau^F = 0$ the case of nonzero shipment costs is discussed in Section 4.5. To shut down the selection channel, I assume that there are no fixed costs of production or exporting, so each firm operates on all markets and has the same productivity and marginal costs. As firms with the same productivity face the same market outcomes, I index firms by their marginal costs MC.

Combining (11) and (12), every firm's revenue and profits are

$$r(MC) = P^{\frac{\zeta - \sigma}{1 - \zeta}} p(MC)^{1-\sigma}, \pi(MC) = \frac{1}{\sigma} P^{\frac{\zeta - \sigma}{1 - \zeta}} p(MC)^{1-\sigma}.$$

Then, free entry condition is

$$\pi\left(\tilde{MC}\right) = \frac{1}{\sigma} P^{\frac{\zeta - \sigma}{1 - \zeta}} p\left(\tilde{MC}\right)^{1-\sigma} = f_s,$$

where $\tilde{MC} = \int_0^\infty \tilde{MC} \left[\frac{q(MC)}{q\left(\tilde{MC}\right)}\right] dG_{MC}$ are average costs weighted by firms' normalized output.

Use $P = p\left(\tilde{MC}\right) M^{-\frac{1}{\sigma - 1}}$ to find the expression for the number of firms:

$$M = \left(\frac{1}{\sigma f_s}\right)^{\frac{(1-\zeta)(\sigma-1)}{\sigma-1-\zeta\sigma}} p\left(\tilde{MC}\right)^{-\frac{\zeta(\sigma-1)}{\sigma-1-\zeta\sigma}} = f_s^{\frac{(1-\zeta)(\sigma-1)}{\sigma-1-\zeta\sigma}} \left(\frac{\sigma - 1}{\sigma^{\frac{1}{\zeta}}}\right)^{\frac{\zeta(\sigma-1)}{\zeta-1-\zeta\sigma}} \tilde{MC}^{-\frac{\zeta(\sigma-1)}{\sigma-1-\zeta\sigma}}.$$

As $\frac{\sigma-1}{\sigma} > \zeta$, $-\frac{\zeta(\sigma-1)}{\sigma-1-\zeta\sigma} < 0$, the number of firms is decreasing in average costs. The intuition is simple: in the case of technological improvement that decreases each firm's costs, expected profit

[25] As discussed in Section 3.3, in the presence of shipment costs, the optimal path of the firm can depend on the destination country; in this case, each firm faces imperfectly correlated total productivity draws for each destination country of a final good.

increases, becoming larger than sunk costs. This positive expected profit attracts new entrants until expected profit is again equal to sunk costs.

The price index can be expressed through the average price

$$P=(\sigma f_s)^{\frac{1-\zeta}{\sigma-1-\zeta\sigma}}\,p\left(\tilde{MC}\right)^{\frac{(\sigma-1)(1-\zeta)}{\sigma-1-\zeta\sigma}},$$

and then the welfare is

$$W=E+\kappa p\left(\tilde{MC}\right)^{-\frac{\zeta(\sigma-1)}{\sigma-1-\zeta\sigma}},$$

where $\kappa=\frac{1-\zeta}{\zeta}(\sigma f_s)^{-\frac{\zeta}{\sigma-1-\zeta\sigma}}$, and then gains from trade are

$$d\ln W=-\frac{\zeta(\sigma-1)}{\sigma-1-\zeta\sigma}d\ln\tilde{MC}. \tag{13}$$

It means that the change in the weighted average of firms' costs and demand parameters are sufficient statistic for measuring welfare gains. A one percent decrease in firms' costs leads to $\frac{\zeta(\sigma-1)}{\sigma-1-\zeta\sigma}$ percent increase in costs of living. The share of direct contribution of lower prices is $\frac{\zeta}{(1-\zeta)(\sigma-1)}$, and the share of gains from increased variety is $\frac{\sigma-1-\zeta\sigma}{(1-\zeta)(\sigma-1)}$.

4.5 Horizontal FDI

In the previous sections I assumed that $\tau^{\mathrm{F}}=0$ to concentrate on the channel of vertical offshoring. In this section, I relax this assumption and show that in the presence of a horizontal FDI channel, my model can nest few workhorse trade models.

Assume $\tau^F>0$. As discussed in Section 3.3, in the presence of shipment costs of the final good, the optimal path of the firm can depend on the destination country, and then each firm has different productivity at home and abroad. Then marginal costs of the firm $MC(\theta,\tau,d)$ depend on destination country d and follow endogenous distribution $G_{MC}(\theta,\tau,d)$.

In case $\tau=\infty$ and $M=2$, the model becomes similar to Helpman et al. (2004). Each firm makes sourcing decisions depending on marginal costs of production in each country $\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib1}$ and $\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib2}$ and trade costs τ^F. It chooses to produce in each country if $\tau^d>\frac{\max\left\{\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib1},\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib2}\right\}}{\min\left\{\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib1},\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib2}\right\}}$ and chooses to produce in a lower-cost country and export to a high-cost country if $\tau^d<\frac{\max\left\{\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib1},\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib2}\right\}}{\min\left\{\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib1},\sum_{i=1}^{\max\{n_b\}}\sum_{b\in B_i}a_{ib2}\right\}}$. Notice that each firm makes this decision based on the relative costs of production at home and abroad and the value of τ^F. In this case, the firm with lower productivity can choose to offshore and the firm with a higher productivity can choose not to offshore.

This model nests Helpman et al. (2004) when $\tau = \infty$ under the following assumptions: there are fixed costs of production, exporting, and offshoring $f_d < f_x < f_I$, and productivity draws are perfectly correlated at every stage: $a_{ib1} = a_{ib2}$ for $\forall i$.[26] In other words, for the sake of clarity, my baseline model ignores the selection of firms, but is richer in terms of joint distribution of marginal costs in each country (this drives the result on nonstrict productivity ordering of offshoring firms).

With small additions, this model also nests Melitz (2003) (using regular CES utility instead of quasilinear, with fixed costs of production and exporting $f_d < f_x$, with no possibility for vertical ($\tau = \infty$) or horizontal ($f_I = \infty$) FDI).

5 Empirical Analysis

It is well documented that firms engaged in offshoring benefit when they face lower trade costs. As firms face lower marginal costs, they set lower prices, and this leads to an increase in welfare for consumers. Amiti et al. (2017) study the effect of China joining the WTO and find that lower input tariffs increase the TFP of Chinese firms.

In this paper, the decrease in trade costs affects firms' productivity through two channels: first, lower trade costs mean that firms can import the same set of inputs for a lower price. At the same time, lower trade costs mean that firms are more flexible making choices over their production structure. As I show in Proposition 3, if a firm chooses to change its organizational structure, its costs will be lower.

In this section, I bring the model to the data. I use the case of China joining the WTO in order to calibrate a simplified version of my model. Structural estimation allows me to decompose changes in productivity of Chinese firms by two channels: lower trade costs and changes in technology. Then, I find gains from trade for Chinese consumers, driven by the former channel. Finally, I decompose gains from trade by two channels: lower costs of inputs and more efficient production structure. My calibration indicates, that the second channel, introduced in this paper is qualitatively large and accounts for approximately 12% of total gains from trade.

5.1 Data

I use the dataset from the Chinese National Bureau of Statistics provided by Aghion et al. (2015). It is an annual survey of Chinese manufacturing firms. I use the year 2000 as the pre-WTO period and the year 2007 as the post-WTO period.

[26] Or under a weaker restriction, $\sum_{i=1}^{\max_{n_b}} \sum_{b \in B_i} a_{ib1} = \sum_{i=1}^{\max_{n_b}} \sum_{b \in B_i} a_{ib2}$.

A potential concern would be the presence of processing trade firms in China: since 1987 imports of raw materials, parts, and components used in the production of goods for exports were duty-free (Branstetter and Lardy (2006)). Brandt and Morrow (2013) document a casual link between trade liberalization and the shift from processing to ordinary trade.

In order to exclude firms producing only intermediate goods, I focus on firms from downstream sectors, found in Antràs et al. (2012).[27]

5.2 Approach

As I showed in Section 4.2, the distribution of firms' total marginal costs $G_{MC}(\theta,\tau)$ depends on trade costs τ, and θ, a vector of parameters of productivity distributions; moreover, changes in trade costs not only shift or stretch the distribution, they can change its whole shape. The dynamics of this distribution over time hence provide the variation that allows me to identify changes in trade costs.

I follow the empirical strategy of Hsieh and Klenow (2009). They use the dynamics of firms' sales distribution and a heterogeneous firms model to measure the impact of resource misallocation in India and China on their TFP. In a broad sense, I study misallocation of resources as well: in the presence of trade costs, the firm cannot achieve its first best production allocation. Proposition 9 shows that the decrease in trade costs works as a technology improvement for all firms, though it is not necessarily proportionate for all firms.

Like Hsieh and Klenow (2009), I use the data on firms' output. With the assumption on preferences, the market share of every firm can be found as a function of its marginal costs. In this paper I use quasilinear CES utility, but the model is consistent with any other utility function as well.

5.2.1 Goodness of Fit Measures

Now the goal is to choose the parameters of the model τ and θ such that the distribution generated by the model fits the empirical distribution of firms' market shares well. Given that the objective is to find a theoretical distribution that best approximates the empirical distribution, it is reasonable to use conventional measures of distance between the distributions.

I use the Kolmogorov-Smirnov (KS) criterion as the main specification as it proved to be less dependent on starting values than Kullback-Leibler divergence (KLD) and Cramér-von Mises (CM) measure. Additionally the test provides a p-value of the hypothesis that two samples are drawn from the same distribution.

[27] I do not need all the firms to be engaged in international production. In my model, not all the firms are engaged in offshoring; for some firms, it is cheaper to produce all the parts domestically. If there are many purely domestic firms (which can happen due to high trade costs or low production costs in China), changes in trade costs should not affect the distribution much.

A KS statistic is the maximum absolute distance between the CDFs of two distributions:

$$D_{KS} = \sup_x |F_n(x) - F(x)|$$

and the problem of minimizing this measure is similar to minimax criterion. Minimization of the KS measure is an example of minimum distance estimation; Parr and Schucany (1980) show that it is a consistent way to estimate the parameters of theoretical distribution.

5.3 Implementation

I assume a two-country world, with China and rest of the world.[28] I allow parameters of the model to change over time. Changes in θ_W and θ_E absorb changes in technology and relative wages for both countries, and changes in τ reflect changes in trade costs including nontariff barriers.

After the model is calibrated and parameters (θ_0, τ_0) and (θ_1, τ_1) are found, where $t = (0,1)$ are pre- and post-liberalization time periods, welfare consequences of the trade liberalization can be analyzed. First, a fall in the costs of production can be driven by changes in technology and relative wages. In this paper, I focus on the effect of trade liberalization. I ignore the potential effect of trade liberalization on technology in both countries and focus on the direct effect of a decrease in trade costs on welfare. I define the welfare effect of trade liberalization as the difference in welfare for a pre-liberalization economy with parameters (θ_0, τ_0) and counterfactual economy with the old technology parameters but a post-liberalization level of trade costs (θ_0, τ_1). Changes in welfare can be found by equation (13) from the difference between average marginal costs of these two economies. Now these gains can be decomposed by two channels: cheaper inputs and more efficient production structure. To perform this decomposition, I simulate the economy with (θ_0, τ_0, w_0), fix the production path of every firm, and for this set of production paths, recalculate marginal costs of each firm for τ_1. The difference between average marginal costs of these two economies will reflect the direct effect of lower trade costs. By Propositions 1 and 3, the size of this effect is equal to or smaller than the total reduction of marginal costs for each firm. Then the difference between total gains from trade and gains from cheaper inputs will be the gains from fragmentation.

In this section, I outline the steps I take to estimate my structural model on data.

1. I use the data for the pre-treatment period ($t = 0$).

[28] This assumption might seem problematic in a multicountry world, but here I consider China joining the WTO, which I interpret as a case of trade liberalization between China and rest of the world. After trade liberalization, firms need to decide whether to produce a given part in China or abroad. Costs of production abroad can be interpreted as a minimum of production costs across all available locations.

2. I simulate an economy with parameters (θ, τ) and find the simulated distribution of firms' market shares.

3. I calculate the distance between the simulated and empirical distributions of firms' market shares.

4. I choose parameter values (θ_0, τ_0) such that the distance from (3) is minimized.

5. I perform steps (1–4) for the data for the post-treatment period ($t = 1$) and find (θ_1, τ_1).

6. I simulate the economies with parameters (θ_0, τ_0) and (θ_0, τ_1, w_0) and find log change in marginal costs due to the fall in trade costs $d\ln\tilde{MC}_{Total} \equiv \ln\left(\frac{\tilde{MC}(\theta_0,\tau_1)}{\tilde{MC}(\theta_0,\tau_0)}\right)$.

7. I simulate the economy with parameters (θ_0, τ_0), fix optimal paths for each firm and find average marginal costs for τ_0 and τ_1: $\tilde{MC}(\theta_0, \tau_0)|\tau_0$ and $\tilde{MC}(\theta_0, \tau_0)|\tau_1$, and find the direct effect of log change in marginal costs $d\ln\tilde{MC}_{Direct} \equiv \ln\left(\frac{\tilde{MC}(\theta_0,\tau_0,)|\tau_1}{\tilde{MC}(\theta_0,\tau_0)|\tau_0}\right)$.

8. The share of welfare gains due to fragmentation and direct effect can thus be found as $1 - \frac{d\ln\tilde{MC}_{Direct}}{d\ln\tilde{MC}_{Total}}$ and $\frac{d\ln\tilde{MC}_{Direct}}{d\ln\tilde{MC}_{Total}}$, respectively.

5.4 Results

To calibrate the model, I make several assumptions. First, I choose the number of stages N in a production chain. With small N, technology is simple and firms' production choices become unresponsive to trade shocks. When N is large, firms' productivity distribution, by the law of large numbers, converges to a single valued degenerate distribution. My simulations suggest that $N = 10$ generates enough flexibility for firms and is consistent with high variability in firms' market shares.

A potential concern here is that many complex goods consist of thousands of intermediate parts and then, according to the model, firms would not differ much in terms of their market shares. Notice, though, it is true only in case cost draws are independent. When there is correlation between production costs of different parts, there can be a large variation in firms' market shares. In other words, the number of parts and the correlation between cost draws work in the opposite direction: the former decreases the variance of firms' shares, the latter increases it. In this paper for the sake of tractability, I focus on independent draws and consequently choose a not-too-high N.

Antràs and De Gortari (2017) reject the hypothesis $N > 2$; they obtain this result, however, with the productivity draws uncorrelated along the chain; more importantly, they degenerate the clustering channel, assuming that firms at every stage choose suppliers based on their productivity only. It is not surprising that models that do not account for clustering predict shorter values of a chain as production processes in these models are less sticky, so for the same technology and trade costs parameters values, they will predict higher trade volumes.

Another key assumption is the distribution of firms' costs on every stage. Two most popular distributional assumptions in trade literature are Pareto (Chaney (2008)) for heterogeneous firms models and Fréchet (Ramondo and Rodrıguez-Clare (2013), Eaton and Kortum (2002)) for Ricardian models. Even though my model includes heterogeneous firms, the market of parts (intermediate inputs) is Ricardian, so in my empirical exercise I stick to Fréchet distribution. Notice, however, that my model can work under any distributional assumptions.

I assume that domestic and foreign productivity draws have the same shape parameter but have different scale parameters thus capturing different technological capabilities of China and the rest of the world and accounts for potential differences in relative wage. I allow both scale parameters to vary over time but keep shape parameter constant — I assume there were no major technological shifts between 2000 and 2007.

For simplicity, I consider i.i.d productivity draws, but correlation between draws has rich interpretation. Correlation along the chain reflects similarity between adjacent stages of production and is a parameter working in the opposite direction than the length of chain. As discussed in Section 5.5, the value of this parameter determines whether snake or spider technology is associated with higher gains from trade. Correlation of draws for the same part but between different countries reflects technology transferability. In the baseline case, a firm's productivity in a domestic country does not depend on its productivity abroad. Another limiting case is Helpman et al. (2004), where a firm can perfectly transfer its technology abroad. Finally, correlation of draws between firms determines the country-specific component of productivity.

To make my results consistent with the most of the literature, I make an assumption of iceberg trade costs. I discuss the behavior of the model under iceberg and specific trade costs in Section 5.5.

Finally, I need values of demand parameters σ and ζ, which determine the mapping between productivities, sales, and welfare. I use standard values from the literature, $\sigma = 5$ and $\zeta = 0.5$.

There are two sources of gains from trade in this model — higher average productivity and higher variety associated with larger entry. Average productivity is determined by technology and trade costs, while the number of firms on the market depends on the value of sunk costs. It is not possible to identify whether the number of firms on the market increased due to improved technology or because of lower sunk costs of production, so I abstract from the variety effect and focus only on welfare changes associated with the different shape of productivity distribution.

The results of this calibration are presented in Table 1. My calibration indicates that trade costs decreased by approximately 16%, so that the share of increase in welfare associated with the trade liberalization accounts for 18%. With the estimates (θ_0, τ_0) and (θ_1, τ_1), I can follow steps 6

Table 1: Calibration Results

Parameters								Welfare shares	
Year	σ	ζ	Shape	Scale 1	Scale 2	τ	KS statistic	Trade Share	Fragmentation share
2000	5	0.5	0.88	1.32	1.17	2.29	0.04	0.18	0.12
2007				1.67	1.06	1.94	0.06		

through 8 and find $1-\frac{d\ln\tilde{MC}_{Direct}}{d\ln\tilde{MC}_{Total}}$ and $\frac{d\ln\tilde{MC}_{Direct}}{d\ln\tilde{MC}_{Total}}$. I find that gains from fragmentation account for $1-\frac{d\ln\tilde{MC}_{Direct}}{d\ln\tilde{MC}_{Total}}=11.7\%$ and gains from lower costs of offshoring are $\frac{d\ln\tilde{MC}_{Direct}}{d\ln\tilde{MC}_{Total}}=88.3\%$.

5.5 Snakes and Spiders

In this section, I compare welfare for sequential and nonsequential production. The implications are different for the cases of iceberg and specific trade costs. The results are presented in Figure 6.

Let's assume that a firm produces a whole good domestically, but then decides to offshore one part i abroad. For the spider technology, transportation costs associated with this decision will be $\tau\, a_i$, while for the snake technology it will be $\tau\left((2+\tau)\,\chi_i+a_i\right)$, where χ_i is the total value of the intermediate good at the moment of transportation. It means that for $i>1$ trade costs associated with offshoring of a single part will always be higher for a snake and will increase downstream. No wonder that welfare is always lower in the case of snake technology, and gains from trade are also lower when trade costs are high. In this case, a significant part of differences in welfare is driven by the fact that iceberg trade costs amplify the costs of transportation for more downstream sectors.

The clustering effect, however, is not driven by the assumption of iceberg trade costs, so it makes sense to compare welfare under the assumption of specific trade costs. Costs associated with offshoring of a single part $N>i>1$ will then be τ for a spider and 2τ for a snake, because the good has to be shipped there and back. At the same time, if a firm chooses to offshore K adjacent parts, costs associated with this decision will be $K\tau$ for a spider and 2τ for a snake. In other words, the production technology that will be associated with higher welfare gains depends on a particular realization of costs. In particular, if there is a positive correlation among productivity draws along the chain, a snake will typically be associated with the higher gains from trade, and if the correlation is negative, a spider will generate higher gains from trade.

At the same time, a firm decides to offshore only if the difference in costs of production exceeds the costs of transportation. When the variance of costs is low, the probability that cost difference of a single part exceeds a given value of τ is low and the chances are that a firm will offshore production of some cluster.

This logic is consistent with the results in Figure 6. Snake technology delivers higher welfare

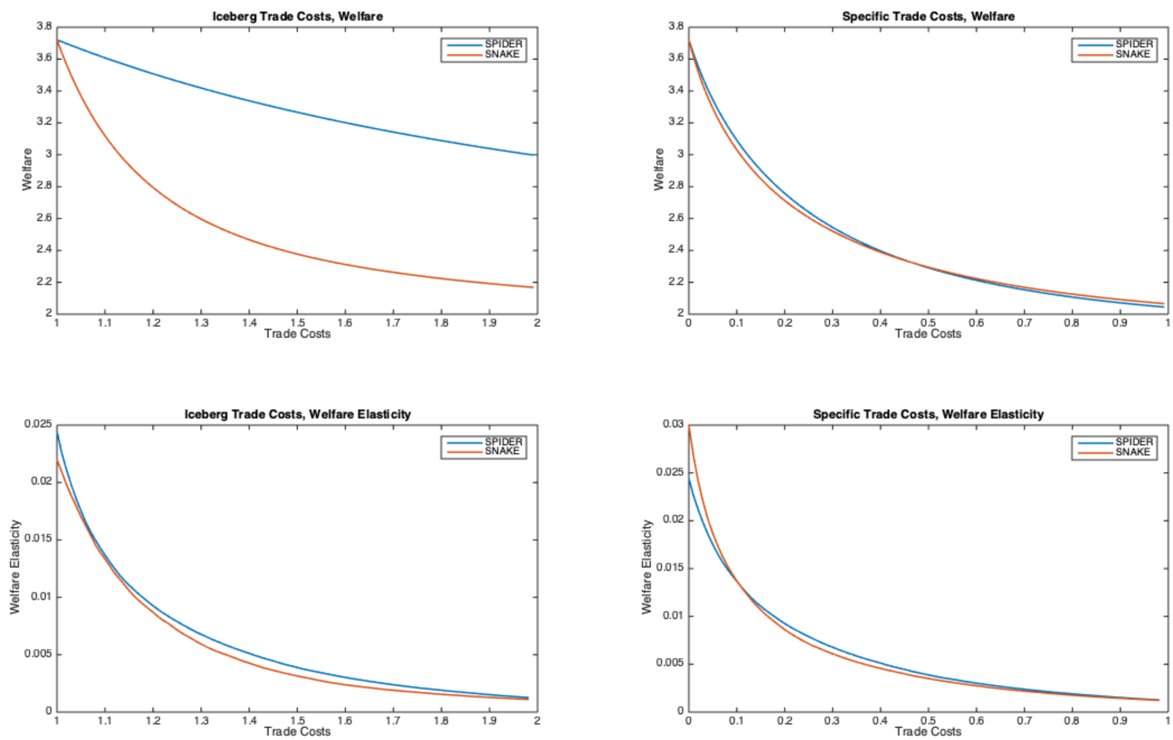

Figure 6: Welfare Under Snake and Spider Technology

for higher values of trade costs and lower for smaller trade costs, reflecting the fact that estimated parameter values suggest that cost distribution has a relatively low variance.

6 Conclusion

Understanding how global firms make decisions and how those decisions affect trade outcomes is important. Baldwin and Venables (2013) and Antràs and Rossi-Hansberg (2009) show that predictions of the models at the intersection of organizational economics and international trade strongly depend on assumptions about the production structure. In this paper, I consider a large class of firm's problems, I call trees, and introduce a way to numerically solve them.

I introduce a general equilibrium model of sequential production with an arbitrary number of stages and without restrictions on trade and production costs, or on the number of countries. The interdependence of the stages of production generates the clustering effect: firms choose to organize their production in large clusters ito save on trade and unbundling costs. This managerial strategy is well documented for car and bicycle manufacturing and electronics, but it was previously ignored by the trade literature.

The interdependence of the stages of production that generates the clustering effect makes a firm's problem hard to solve analytically. I provide a simple algorithm based on the Bellman optimality principle that solves this problem; it can be easily modified for various extensions.

I show that, regardless of the cost structure, any case of trade liberalization leads to the fall in firm's marginal costs. It creates a new channel for the gains from trade: firms that can allocate their production facilities more efficiently, will increase their productivity and lower their prices. Better technology increases entry of firms and leads to higher variety for consumers.

I propose a simple, general equilibrium framework to illustrate the gains from trade channels the model generates. This framework can be easily extended so that it can nest popular trade models. Computational algorithms proposed in this paper combined with the simulated maximum likelihood estimation allows me to decompose the gains from lower costs of offshoring by two channels: cheaper inputs and more efficient production structure. Calibration of the model on Chinese firm-level data indicates that the latter channel is sizable and accounts for up to 12% of total gains from trade.

References

Aghion, P., J. Cai, M. Dewatripont, L. Du, A. Harrison, and P. Legros (2015): "Industrial policy and competition," *American Economic Journal: Macroeconomics*, 7, 1–32.

Allen, T., C. Arkolakis, and X. Li (2015): "On the existence and uniqueness of trade equilibria," *Manuscript, Yale Univ*.

Amiti, M., M. Dai, R. C. Feenstra, and J. Romalis (2017): "How did China?s WTO entry benefit US consumers?" Tech. rep., National Bureau of Economic Research.

Amiti, M. and D. R. Davis (2012): "Trade, firms, and wages: Theory and evidence," *The Review of economic studies*, 79, 1–36.

Antràs, P. and D. Chor (2013): "Organizing the global value chain," *Econometrica*, 81, 2127–2204.

Antràs, P., D. Chor, T. Fally, and R. Hillberry (2012): "Measuring the upstreamness of production and trade flows," *The American Economic Review*, 102, 412–416.

Antràs, P. and A. De Gortari (2017): "On the geography of global value chains," Tech. rep., National Bureau of Economic Research.

Antras, P., T. C. Fort, and F. Tintelnot (2017): "The margins of global sourcing: Theory and evidence from us firms," *American Economic Review*, 107, 2514–64.

Antras, P. and E. Helpman (2004): "Global sourcing," *Journal of political Economy*, 112, 552–580.

Antràs, P. and E. Rossi-Hansberg (2009): "Organizations and trade," *Annu. Rev. Econ.*, 1, 43–64.

Baldwin, C. Y. and K. B. Clark (2000): *Design rules: The power of modularity*, vol. 1, MIT press.

Baldwin, R. and A. J. Venables (2013): "Spiders and snakes: offshoring and agglomeration in the global economy," *Journal of International Economics*, 90, 245–254.

Brandt, L. and P. Morrow (2013): "Tariffs and the Organization of Trade in China," *Available at SSRN 2285628*.

Branstetter, L. and N. Lardy (2006): "China's embrace of globalization," Tech. rep., National Bureau of Economic Research.

CHANEY, T. (2008): "Distorted gravity: the intensive and extensive margins of international trade," *The American Economic Review*, 98, 1707–1721.

COSTINOT, A., J. VOGEL, AND S. WANG (2013): "An elementary theory of global supply chains," *The Review of Economic Studies*, 80, 109–144.

EATON, J. AND S. KORTUM (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

FALLY, T. AND R. HILLBERRY (2015): *A coasian model of international production chains*, The World Bank.

FEENSTRA, R. C. AND G. H. HANSON (1996): "Globalization, outsourcing, and wage inequality," *The American Economic Review*, 86, 240.

FRIGANT, V. AND Y. LUNG (2002): "Geographical proximity and supplying relationships in modular production," *International Journal of Urban and Regional Research*, 26, 742–755.

GALVIN, P. AND A. MORKEL (2001): "The effect of product modularity on industry structure: the case of the world bicycle industry," *Industry and Innovation*, 8, 31.

GOLDBERG, P. K., A. K. KHANDELWAL, N. PAVCNIK, AND P. TOPALOVA (2010): "Imported intermediate inputs and domestic product growth: Evidence from India," *The Quarterly journal of economics*, 125, 1727–1767.

GROSSMAN, G. M. AND E. ROSSI-HANSBERG (2008): "Trading Tasks: A Simple Theory of Offshoring," *The American Economic Review*, 98, 1978–1997.

——— (2012): "Task trade between similar countries," *Econometrica*, 80, 593–629.

HARMS, P., O. LORZ, AND D. URBAN (2012): "Offshoring along the production chain," *Canadian Journal of Economics/Revue canadienne d'économique*, 45, 93–106.

HELPMAN, E., M. J. MELITZ, AND S. R. YEAPLE (2004): "Export versus FDI with heterogeneous firms," *American economic review*, 94, 300–316.

HSIEH, C.-T. AND P. J. KLENOW (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124, 1403–1448.

JOHNSON, R. C. AND A. MOXNES (2016): "Technology, trade costs, and the pattern of trade with multi-stage production," .

JOHNSON, R. C. AND G. NOGUERA (2012): “Accounting for intermediates: Production sharing and trade in value added,” *Journal of International Economics*, 86, 224–236.

MAS-COLELL, A., M. D. WHINSTON, J. R. GREEN, ET AL. (1995): *Microeconomic theory*, vol. 1, Oxford university press New York.

MELITZ, M. J. (2003): “The impact of trade on intra-industry reallocations and aggregate industry productivity,” *Econometrica*, 71, 1695–1725.

PARR, W. C. AND W. R. SCHUCANY (1980): “Minimum distance and robust estimation,” *Journal of the American Statistical Association*, 75, 616–624.

RAMONDO, N. AND A. RODRIGUEZ-CLARE (2013): “Trade, Multinational Production, and the Gains from Openness,” *Journal of Political Economy*, 121.

RAZHEV, M. (2015): “Input-Output Structure and Trade Elasticity,” in *Working Paper*.

SIRKIN, H. L., M. ZINSER, D. HOHNER, ET AL. (2011): *Made in America, again: Why manufacturing will return to the US*, Boston Consulting Group Boston.

TINTELNOT, F. (2017): “Global production with export platforms,” *The Quarterly Journal of Economics*, 132, 157–209.

WANG, W. (2017): “Intermediate Goods and Misallocation in China?s Manufacturing Sector,” .

WU, X. AND F. ZHANG (2011): “Efficient supplier or responsive supplier? an analysis of sourcing strategies under competition,” in *China Business Initiative conference, sponsored by the Chazen Institute of International Business at Columbia Business School*.

YI, K.-M. (2010): “Can multistage production explain the home bias in trade?” *The American Economic Review*, 100, 364–393.