Production Clustering and Offshoring

Vladimir Tyazhelnikov*†

January 15, 2017

Abstract

I propose a simple and computationally feasible algorithm to solve a firm's problem for a large class of sequential production models of offshoring. These models allow for production chain of any length, any number of sourcing countries and arbitrary structure of production and trade costs. I show that in this class of models, allocation decisions 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' productivities is endogenous with respect to trade costs: trade liberalization leads to a distribution that stochastically dominates the old one, thus leading to increase in welfare. I use the model to decompose the welfare gains from trade liberalization by two channels: cheaper intermediate inputs and a more efficient production structure. I apply the model to the data and study the case of China joining the WTO. I use a simulated maximum likelihood technique to calibrate the model and find that more efficient production structure accounts for approximately 25% 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, Jaerim Choi, Oleg Itskhoki, Réka Juhász, Dmitry Mukhin, Stephen Redding, 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 and European Trade Study Group 2016.

†University of California, Davis. E-mail: vtyazhelnikov@ucdavis.edu.

1 Introduction

Antràs and Chor (2013) state that "most production processes are sequential by nature." Hummels et al. (1998) estimated world sequential (vertical) trade share between 20% and 25%. A firm with a sequential production technology faces a complicated optimization problem: the location decision on production of a given part affects all subsequent decisions; hence, a firm cannot make production choices independently and has to choose an optimal path instead.

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.[1] Decreases in trade costs allow a firm to fragment its production more and exploit productivity differences between countries. Production clusters of complementary intermediate parts are easily observable in managerial practices of multinationals. Frigant and Lung (2002) describe the strategy of modular production and its prevalence in the car production 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.[2] Snakes and spiders are two limiting cases of perfectly

[1]Not all decreases in trade costs lead to an increase in offshoring; re-shoring, a well-documented phenomenon (Sirkin et al. (2011), Wu and Zhang (2011)), is a situation when a previously offshored part is produced domestically again and is consistent with my model as well. A firm facing high trade costs has to offshore a large cluster. With lower costs, a firm can afford to have smaller clusters and then may choose to re-shore some parts previously produced in the large cluster. It follows that one should be very careful interpreting the impact of re-shoring on domestic employment and wages; re-shoring in this model is driven by the fall in trade costs and, hence, is accompanied by offshoring of parts previously produced domestically.

[2]These two production structures are the most used in the offshoring literature. Sequential production models include: Antràs and Chor (2013), Costinot et al. (2013) and Fally and Hillberry (2014). Examples of non-sequential models are Basco and Mestieri (2013), Antràs and Helpman (2003), Feenstra and Hanson (1996).

sequential and nonsequential technologies, and 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 sub-chain. These sub-chains represent the production technology of complex intermediate parts that are assembled together to become a final good.

Due to the interdependence of production decisions on different stages, firm's snake and tree allocation problems do not have a closed form solution, unless strong assumptions on production and trade costs are made.[3] I propose an optimal control algorithm based on the Bellman optimality principle, that is easy to implement, allows for various extensions, does not have restrictions on the cost structure, and has a short computational time.

I combine the cost minimization problem described above with a general equilibrium model of heterogeneous firms.[4] In order to quantify the model, I employ stochastic production costs formulation: costs of production on each stage and in each country are a random draw from some probability distribution. Firms' productivities then also follow some random distribution that depends on the trade costs parameter. I show that in case of trade liberalization the new distribution of productivities first order stochastically dominates the old one. Higher firms' productivity leads to lower prices and higher variety, thus increasing welfare.

There is extensive empirical evidence on the tight link between access to cheap intermediate inputs and firms' productivity (for example Amiti and Davis (2012) and Goldberg et al. (2008)). The main difference is that in my model the productivity gains are driven not only by cheaper

[3]For example Costinot et al. (2013) assume that all countries have the same productivity, but can be ordered by the probability of mistake that destroys the intermediate good. Fally and Hillberry (2014) assume that costs of production for each supply chain in each country do not depend on the stage of production. Baldwin and Venables (2013) in their baseline snake model assume there are two countries, trade costs do not depend on the stage of production, and production costs can take one of two values.

[4]Snake and tree cost minimization problems can be combined with a variety of general equilibrium models. One of the obvious choices is EK-style Ricardian model. In this case, the model becomes close to Yi (2003, 2010), Johnson and Moxnes (2013) and Ramondo and Rodrıguez-Clare (2013) without an input-output loop, but with a large number of stages and clusterization effect. In this case, sequential production of a good is interpreted as a transformation of an intermediate good along the production chain from upstream to downstream industries.

intermediate inputs, but through the increased fragmentation of production process.[5] I bring my model to the data and estimate the contribution of these two channels for gains from trade in the case of China joining the WTO.

I use the dynamics of Chinese firms' sales distribution before and after China joined the WTO to calibrate the model.[6] Structural estimation via simulated maximum likelihood technique allows me to find the changes in firms' TFP associated with the decrease in trade costs, derive welfare gains and decompose them by two channels: cheaper inputs and higher fragmentation. My calibration suggests that the latter channel, I introduce in this paper, accounts for 25% of total gains for Chinese consumers from joining the WTO.

My model is similar in spirit to Tintelnot (2014) and Antras et al. (2014). Both papers introduce a model with complicated firms' problems, that do not have closed form solution, but can be solved by a numerical algorithm. In Tintelnot (2014) the combination of fixed costs of opening a new plant and variable shipping costs drives the proximity-concentration trade-off. Antras et al. (2014) 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 all intermediate inputs are either perfect substitutes (same part produced in different countries) or perfect complements (different parts); instead I focus on firms' allocation problem that arises in cases when technology exhibits at least some degree of sequentiality.

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 and estimates the welfare consequences of China joining the WTO. Part 6 concludes.

[5] Caliendo and Rossi-Hansberg (2012) introduce a model where a firm chooses endogenous organizational structure, in particular the number of managerial layers and a level of centralization that depends on heterogeneous demand.

[6] I follow Hsieh and Klenow (2007). They use the changes in firms' sales distribution to estimate resource misallocation and its effect on firms' TFP in China and India. In my model trade costs can be interpreted as frictions that prevent firms from optimally allocating their resources.

2 Firms' Problem

In this part I provide the solution to a firm's cost minimization problem. I start with a standard sequential production technology and then extend it for the more general case of tree technology.

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. 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.[7,8] After all parts are produced, a final good is delivered to a third country, where production is not possible, and shipment costs are same for all M countries.[9]

The firm then minimizes its per unit costs, that I further 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 sub-problems for every

[7] In this paper by trade costs I mean the costs of offshoring. In a narrow sense it is the costs of shipment of intermediate goods between countries.

[8] These are costs to produce one unit of final good. I assume the firm has a technology that exhibits constant returns to scale. It means that τ can be interpreted as a special tariff.

[9] 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.

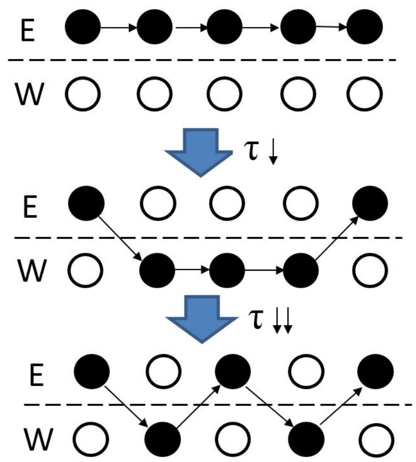

Figure 1: Non-monotonic Impact of Trade Liberalization

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 countries.

Figure 1 illustrates the simplest example, in case of two countries and 5 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 choses 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 re-shores 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. Because of 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:

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

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 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 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 $M\times N$.[10,11]

[10] 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 nod. 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 nod. Costs of such a movement are indicated in corresponding nods. This is a simple decision theory problem with one player, complete information and absence of randomness. It can be solved by backwards 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, conditional on its choice on the previous stage. The backwards induction algorithm with a tree is more intuitive, but cannot be directly applied to more complex problems described in section 2.2 and 2.3. That is why further I focus on the Bellman equation interpretation of the problem.

[11] Matlab code for this and all other optimization algorithms in this paper is available upon request.

This algorithm works for stage-dependent τ as well (for example per unit trade costs can be larger for downstream parts).[12] 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 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.

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 approaches in the offshoring literature to model a production technology of a complex good. Other popular choices are non-sequential models, where the order of production does not matter (Antràs and Helpman (2003), 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 production technologies. Production of a final good can be represented as a set of sequential sub-chains, that are assembled together in more complex intermediate goods. Both snake and spider technologies are particular cases of this tree technology. A tree with one sub-chain is a snake, and a tree with many sub-chains of length 1 is a spider. I illustrate the taxonomy of these technological assumptions in Figure 2.

Tree technology relies only on two restrictive assumptions on the technology. First, parts produced in different countries are perfect substitutes.[13] I find this deviation from Armington assumption quite realistic for such industries as car production and electronics. In case a car

[12]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 good.

[13]Notice that parts produced in different countries are not necessarily identical; cost of production in each country can be interpreted as quality adjusted.

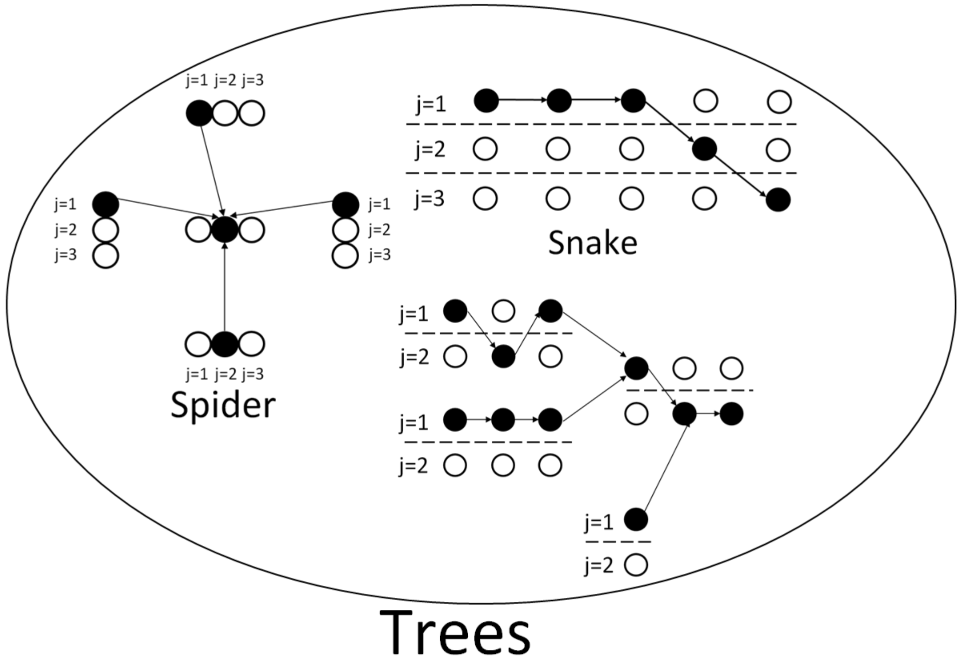

Figure 2: Taxonomy of Technological Assumptions

producer sources input, for example, wheels for a car, it cares about the price and quality of wheels, but does not benefit from a larger variety of wheels sourced from different locations. The second assumption is that sequentially produced intermediate parts can be assembled together, but cannot be disassembled. [14]

2.2.1 Technology

I assume that the firm can have an arbitrary number of sub-chains that can be combined at any stage of production. Figure 3 illustrates the two country example of such technology: there are

[14]This assumption seems reasonable: if a firm made a choice to assemble some parts together, why would it disassemble them and assemble these parts again 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 25% tax on light trucks as a response to France and Germany increasing their tariffs on US 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 their 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 Connects cross 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.

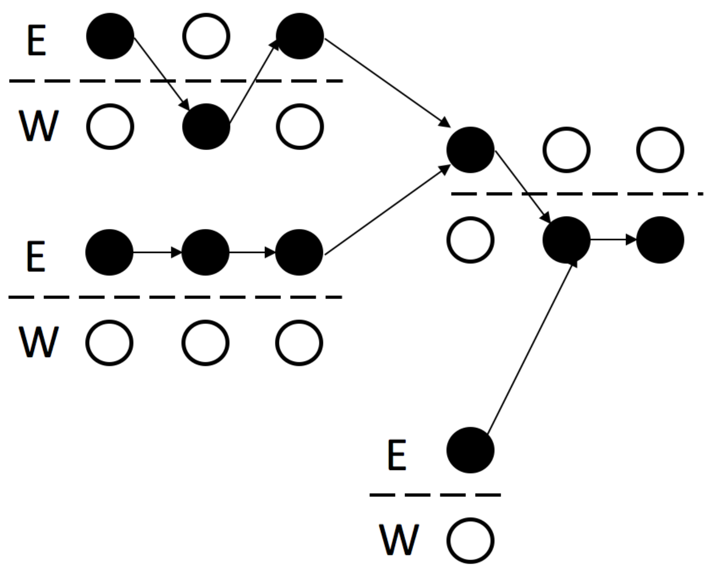

Figure 3: Tree Technology

few intermediate goods produced sequentially that 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 sub-chain 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 sub-chains 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 sub-chain branches that join at the points of subassembly and become one final good trunk in the end.

2.2.2 Reversed 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 on the previous stage as a state variable for the given node. In case part i is assembled of 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 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 previously produced, 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 a problem of allocating N parts in M countries in order to minimize the production costs. It has technological restrictions on the order of production, but 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 as well: 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 the 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 on 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 the last stage and so on.[15] In case more than one 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, that was not previously assigned to another branch.

I define n_b as the 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}. Ω_{ib} is the set of branch indexes b of all parts produced on stage $i+1$, connected to the stage ib. ν_{ib} is a branch of part produced at stage $i-1$, a node ib is connected to. B_i is a set of all branches present at stage i. I present the example of such enumeration in Figure 4.

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

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

Expressions (1) and (5) look different because of reversed enumeration and due to more complicated indexing structure of a tree. The main idea remains the same: costs of production of every part depends on the location choice and production location of the next part.[16]

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

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

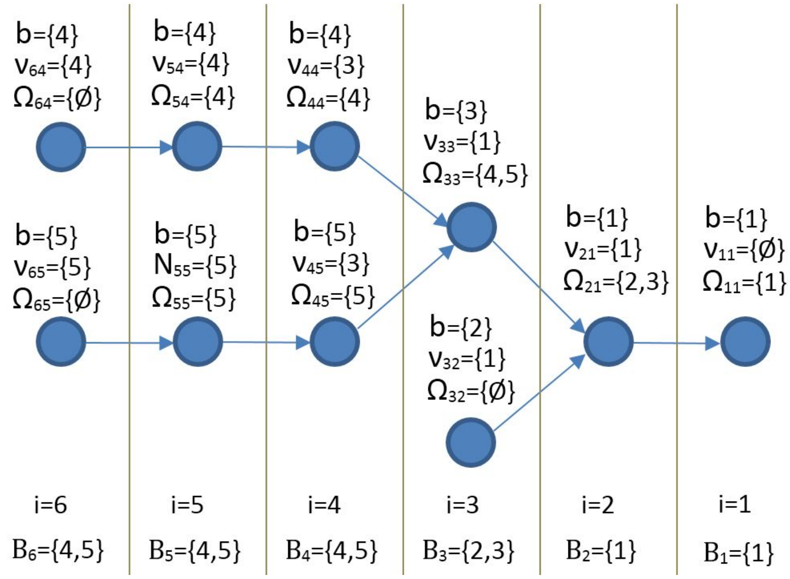

Figure 4: Tree Notation

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 the algorithm is still reversed): every sub-chain problem is solved similarly to the baseline chain problem. The difference arises when two or more sub-chains 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 sub-chains being assembled.

Notice that there are separate value functions (with index *ib*) for different branches. As sub-chains 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, that is associated with a value of the single value function with index

$bj = 11$.

2.3 Alternative Trade Costs Functions

In earlier sections I assumed that trade costs do not depend on the value of the transported intermediate good. I did this for three reasons. First, I follow Baldwin and Venables (2013) and interpret these costs as unbundling costs: costs the firm exhibits to break its production chain and to locate adjacent stages of production in different countries. Besides tariff and transportation costs, these unbundling costs would include organizational costs of opening a factory and organizing a production line in another country as well as potential issues with the timing of delivery of intermediate parts; Harrigan and Venables (2006) find it may be an important issue. Fort (2013) emphasizes the role of coordination costs in fragmentation of production process.

When trade costs depend on the value of the intermediate good,[17] 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 $i \neq j$.[18] In other words, the firm has to ship $T^{ice}(i,j)$ units of intermediate good from country i to receive 1 unit of intermediate good in country j. As the firm minimizes its per unit costs, I reformulate the problem in the following way: a firm produces 1 unit of each intermediate part until it chooses to cross the border. Whenever a border crossing between countries i and

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

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

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 $\left(T^{ice}\left(c_{ib}, c_{i-1\nu_{ib}}\right)-1\right)\chi_{ib}$, 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 $i \neq j$. Costs of crossing the border will then be $\tau T\left(c_{ib}, c_{i-1\nu_{ib}}\right)\chi_{ib}$.

One can see that in case $T^{ice}\left(c_{ib}, c_{i-1\nu_{ib}}\right) - 1 = \tau T\left(c_{ib}, c_{i-1\nu_{ib}}\right)$, costs of the border crossing coincide for iceberg and ad-valorem cases. It means that for any value of iceberg trade costs an equivalent ad-valorem tariff can be found such that firm's optimal choice and marginal costs will be the same.[19,20]

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:

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

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

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 the stage ib for ad-valorem tariff; with multiple border crossings, some parts are taxed more than once.

Expression for the Bellman equation is then straightforward:

[19] Here I focus on the fact that iceberg trade costs and ad-valorem tariff 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.

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

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

The only difference of this Bellman equation from (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.

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 costs of offshoring.

3.1 Multilateral Trade Liberalization

Proposition 1. *In optimum firm's total costs are non-decreasing in trade costs.*[21]

Proof. Let's assume there is optimal path A for τ_0, optimal 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 $MC(A, \tau_0, C) \geq MC(B, \tau_1, C)$ that contradicts the initial assumption. □

[21] Here and further I have propositions and theorems with weak monotonicity. It 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 large number of firms are taken into consideration: with a continuum of firms, changes in parameter values lead to change in the optimal path for at least some of the firms.

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 $NTMC(Y) \equiv \sum_{i=1}^{N} \left(a_{ki} 1\left\{c_i = k\right\}\right)$ be a value of non-transport 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 transportation schedule independent of trade costs price shifter τ. One can think about TQ as a total number of miles a transportation ship traveled, and τ is a price of gas.[22]

Lemma 1. *The transportation quantity is a non-increasing function of τ.*

Proof. Let path A with transportation quantity $TQ(A)$ be chosen for $\tau = \tau_0$ and path B with transportation quantity $TQ(B)$ to 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 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 $NTMC(B) + TQ(B)\tau_1 < NTMC(A) + TQ(A)\tau_1$ that contradicts the condition on optimality of A under τ_0. □

The intuition behind this proposition is the following: the price of gas decreases, so the firm does not have incentives to decrease the number of miles traveled, even though total expenses on transportation can increase or decrease.[23]

[22] 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.

[23] In case of similar transportation costs between all country pairs, Lemma 1 means that the number of border crossings is a non-increasing function of τ. Then by defining cluster as a sequence of parts produced in the same country the following statement is true: The average size of a cluster is a non-decreasing function of τ. It simply

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 $MC(A, \tau_0) > MC(B, \tau_1)$. $\square$

Proposition 3. *If the firm changes its unique optimal path due to decrease in τ, then non-transportation 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 $MC(A, \tau_1) = NTMC(A) + \tau_1 TQ(A) < NTMC(B) + \tau_1 TQ(B) = MC(B, \tau_1)$, that contradicts optimality of B under τ_1. $\square$

This is the key proposition that represents gains from fragmentation. It is not surprising that 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 of increase in productivity: the optimal path of the firm depends on trade costs. With a change in trade costs, the firm can choose different production structure that would lead to higher efficiency of production. Similarly to Proposition 1, this result does not rely on sequentiality of production structure.

Proposition 4. *Production in a given country can depend on trade costs non-monotonically. Re-shoring is possible.*

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 non-increasing in τ and, hence s is non-decreasing in τ.

Proof. Consider the following numerical example with sequential technology, 2 countries and 5 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 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 a high value of trade costs the firm chose to produce the whole good in the first country. With the decrease in trade costs, it chose to offshore a large cluster to the second country. When trade costs decreased even further, the firm fragments its production more and re-shores 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:[24] 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:

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

[24] One reasonable way to simplify the problem is to assume that the firm does not know its costs in 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 vertical offshoring channel remains redundant.

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, there are more opportunities 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 part, 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 section the firm can sell the final good to countries $j \in \{1, ..., M\}$. In this case the cost minimization problem is interacted with a proximity to consumer market consideration. A firm can choose different optimal paths for production of final goods with different destination countries.[25]

[25] Here the firm does not face a complicated export platform problem as in Tintelnot (2014) because in this model there are no fixed costs of opening a plant and the firm just solves horizontal FDI problem independently for each destination country.

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

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

$$MC_{ib} = \sum_{k=1}^{M} \left(\mathbb{1}\left(c_{ib} = k\right) a_{ibk} + \tau T\left(c_{ib}, c_{i-1\nu_{ib}}\right)\right),$$

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

A firm chooses an optimal path in order 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 there is only one optimal path that 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 multi-country 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: $\tau^0_{12} = \tau^0_{21} = 2$, $\tau_{13} = \tau_{31} = 2$, $\tau_{23} = \tau_{32} = 0.5$. 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 cost minimizing decision would be to produce all parts in the West. Now if trade costs between countries 1 and 2 decrease $\tau^1_{12} = \tau^1_{21} = 1$, 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 reason the third country benefits is that 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 are adjacent, now trade costs between countries 1 and 3 do not matter, and the firm faces lower trade costs between countries 2 and 3. I provide the following example: with high trade costs, a US firm chose 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 and advantage to offshore adjacent parts to Indonesia which is located 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.[26] I normalize the wage in country 1 to $w_1 = 1$. I assume that labor supply is perfectly inelastic and the firm has a constant returns to scale production technology. The problem of every firm then looks like:

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

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

and 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. *A firm's labor demand L_{Dk} in each country k is a non-increasing function of w_k.*

Proof. Let 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 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 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.

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

Proof. Work in progress □

4 General Equilibrium

In this part I describe a heterogeneous firms model of M countries, where each of 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 can be solved numerically[27] and is embedded in stochastic empirical estimation.

I use the simplest version of monopolistic competition model with quasilinear utility, 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 clearly illustrate how the new mechanism operates in a general equilibrium environment.[28]

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 a 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 de-

[27]Matlab code to generate the simulated economy is available upon request.
[28] Derivation of richer and more complicated models is straightforward.

scribed by the quasi-linear utility function:

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

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 $q_0 = E - P^{-\frac{\zeta}{(1-\zeta)}}$. Indirect utility function is then

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

and demand function for good ω is:

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

4.2 Technology

All the problems described in part 2 took costs of production as 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. I follow Yi (2010), Ramondo and Rodrıguez-Clare (2013) and Johnson and Moxnes (2013) and assume that costs of production in each stage 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; Hanson et al. (2014) found that absolute advantage between countries can be approximated by the generalized gamma distribution. In this paper I am not making distributional assumptions in order to make 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.[29] 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[30] first order stochastically dominates $MC(\theta, \tau_1)$*

Proof. By Proposition 1 $MC(A, \tau)$ is non-decreasing 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 $MC(\theta, \tau_1)$ □

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 costlessly switch between optimal paths and does not take into account sunk costs the firm paid in order 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.

[29]Correlation between draws within country and correlation of draws at the same stage but for different countries have interesting implications, that are beyond the scope of this paper.

[30]The weak dominance appears in 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.

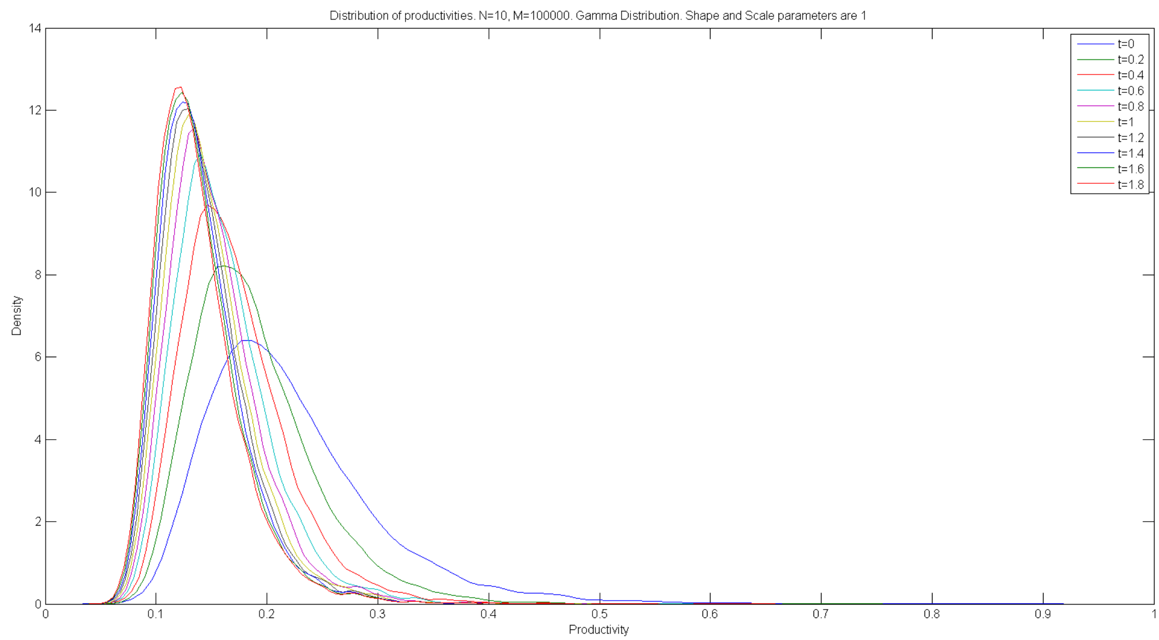

Figure 5: Evolution of Firms' Productivities

Figure 5 illustrates Proposition 9. I simulate 100000 firms with $N = 10$ and generalized gamma distribution of the draws. Notice that the distribution of firms' productivities 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, that would lead to larger entry. Both effects lead to the increase in welfare: through lower prices and larger variety.

4.3 Market Structure

There are two sectors in each country: homogeneous good sector and monopolistically competitive sector. Homogeneous good is produced on a competitive market with productivity one and normalizes wages in the economy to 1.[31] Firms in a monopolistically competitive sector

[31] In section 3.5 I show how to endogenize wages and drop homogeneous sector assumption.

pay sunk costs f_s to enter the market. In order to separate the effect of trade costs on vertical 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 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 a destination country of a final good.[32]

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 non-zero shipment costs is discussed in section 4.5. 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:

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

[32] As discussed in section 3.3, in the presence of shipment costs, 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.

where $\tilde{MC} = \int_0^\infty \tilde{MC} \left[\frac{q(MC)}{q(\tilde{MC})} \right] dG_{MC}$ 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 $\frac{\sigma-1}{\sigma} > \zeta$, $-\frac{\zeta(\sigma-1)}{\sigma-1-\zeta\sigma} < 0$, and hence the number of firms is decreasing in average costs. The intuition is simple: in case of technological improvement that decreases each firm's costs, expected profit increases and it becomes larger than sunk costs. This positive expected profit attracts new entrants until expected profit is equal to sunk costs again.

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 $\kappa = \frac{1-\zeta}{\zeta}\left(\sigma f_s\right)^{-\frac{\zeta}{\sigma-1-\zeta\sigma}}$, 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 weighted average of firms' costs and demand parameters are sufficient statistics to measure 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 $\tau^{\mathrm{F}} = 0$ in order to concentrate on the channel of vertical offshoring. In this section I relax this assumption and show that in the presence of 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 delivery, optimal path of the firm can depend on the destination country, and then each firm has different productivities 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 very similar to Helpman et al. (2003). 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 value of τ^F, in this case the firm with lower productivity can choose to offshore and the firm with a higher productivity not to offshore.

This model nests Helpman et al. (2003) 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 $\forall i$.[33] In other words, for the sake of clarity, my model ignores the entry behavior of firms, but is richer in terms of joint distribution of marginal costs in each country (this drives the result on non-strict productivity ordering of offshoring firms).

Under small additions this model also nests Melitz (2003) (using regular CES utility instead

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

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 the increase in welfare for consumers. Amiti et al. (2016) study the effect of China joining WTO and find that lower input tariffs increase TFP of Chinese firms.

In this paper the decrease in trade costs affects firms' productivities through two channels: first, lower trade costs mean that firms can import the same set of inputs for 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 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 sources: lower costs of inputs and more efficient production structure. My calibration indicates, that the second source, introduced in this paper is qualitatively large and accounts for approximately 25% of total gains from trade.

5.1 Data

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

2007 as post-WTO. I use Chinese firms' domestic sales[34] in order to construct the distribution of firms' sales. The mapping between sales and productivity in the model holds only for sales of a final product.

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 that there is a casual link between trade liberalization and the shift from processing to ordinary trade.

In this paper I abstract from the interaction between processing and ordinary trade regimes: I consider firms that have home sales, and hence pure processing firms are excluded from the sample. There can be firms that perform both ordinary and processing trade, but, once again, offshoring activity relevant to domestic sales is subject to tariff duty.

In order to exclude firms producing intermediate goods only, I focus on firms from downstream sectors, found in Antràs et al. (2012).[35]

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 do not only shift or stretch the distribution, but can change its whole shape. Dynamics of this distribution over time hence provides the variation that allows to identify changes in trade costs..

I follow empirical strategy close to Hsieh and Klenow (2007). They use the dynamics of firms' sales distribution and heterogeneous firms' model to measure the impact of resource mis-

[34] I focus on domestic sales because comparing total sales of firms that have different sets of export destinations can be problematic: trade liberalization will have non-trivial heterogeneous effect on such firms.

[35] Here 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 (that can happen due to high trade costs or low production costs in China), changes in trade costs should not affect the distribution much.

allocation 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 not necessarily proportionate for all firms.

Similarly to Hsieh and Klenow (2007) I use the data on sales of the final product. With the assumption on utility or demand system, sales of every firm can be found as a function of its marginal costs. Notice that I introduce utility function in the model anyway in order to derive general equilibrium and welfare implications. In this paper I use quasilinear CES utility, but the model is consistent with any other utility function as well.[36]

5.2.1 Goodness of Fit Measures

Now the goal is to choose the parameters of the model τ, θ_W, and θ_E such that the distribution generated by the model fits the empirical distribution of firms' sales well. Given that the objective is to find theoretical distribution that approximates the empirical distribution the best, it is reasonable to use conventional measures of distance between the distributions. I use Kullback-Leibler divergence (KLD) as the main specification. As a robustness check, I use alternative measures, Kolmogorov-Smirnov (KS) and Cramér-von Mises (CM).

All three measures KS, CM and KLD represent the distance between a theoretical distribution with cdf $F(x)$ and empirical distribution with cdf $F_n(x)$.

[36] I match empirical and theoretical distributions of sales, because sales are directly observed in the data, while finding empirical analog of firms' marginal costs or productivities can be problematic. In particular, I do not use standard productivity measures such as Olley and Pakes (1992), because the firms under consideration can have production facilities abroad. In order to find true productivity of a firm, one has to know not only shipments of intermediate inputs but employment abroad as well. Imagine two identical firms, one of which offshored the production of an input intensive part, and another offshored a part that does not require much resources to be produced. Let's assume that both parts have the same value when produced, and both are needed in the production of the final good. Then the second firm will look more productive as it has lower employment at home, the same value of imported intermediate inputs, and the same total output. Measuring productivity based on value added instead of total output is also not a solution: once again, different firms can offshore the production of different parts, and the best one can find is the TFP of parts produced at home with endogenous choice on what parts to offshore. In a broader context, it means that the standard TFP measures applied to firms engaged in offshoring can be inaccurate.

Kullback-Leibler Divergence is defined as

$$D_{KL} = \int_{-\infty}^{\infty} f_n(x) \log \frac{f_n(x)}{f(x)} dx,$$

where $f_n(x)$ and $f(x)$ are empirical and theoretical cdf's. Eguchi and Copas (2006) show that minimizing KLD between two distributions is similar to finding Maximum-Likelihood estimates of parametric distribution. In this sense, the results of my calibration can be interpreted as ML estimates of distribution $G_{MC}(\theta, \tau)$.[37] I am not the first to use KLD in international trade literature; Mrázová et al. (2016) study how well models with different assumptions on utility function and productivity distributions match the data on the distribution of sales and markups. They use KLD to quantify the information loss from different models.

A Kolmogorov-Smirnov statistic is the maximum absolute distance between the cdf's of two distributions:

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

and the problem of minimizing this measure is similar to minimax criterion.

A Cramér-von Mises measure can be interpreted as the sum of squared differences between empirical and theoretical distributions:

$$D_{CM} = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, dF(x).$$

Minimization of both KS and CM measures are examples of minimum distance estimation; Parr and Schucany (1980) show that it is a consistent way to estimate the parameters of theoretical distribution.

Minimization of these 3 measures gives qualitatively similar results.

[37] I prefer simulated maximum likelihood estimation (SML) to more common simulated method of moments used, for example, in Eaton et al. (2011). SML does not depend on the choice of the moments, uses all available information and has a simple and intuitive objective function.

5.3 Implementation

I assume a two country world, with China and rest of the world.[38] I allow all 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 non-tariff barriers.

In order to simplify the analysis, I assume that ex-ante technology (distributions) in China and in the rest of the world is the same, but the relative wage w can change over time $F_{Ct}\left(\theta_{Ct}\right) = w_t F_{Wt}\left(\theta_{Wt}\right)$, where t is an index for time period.

After the model is calibrated and parameters (θ_0, τ_0, w_0) and (θ_1, τ_1, w_1), where $t = (0, 1)$ are pre- and post-liberalization time periods are found, welfare consequences of the trade liberalization can be analyzed. First, 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 welfares for pre-liberalization economy with parameters (θ_0, τ_0, w_0) and counterfactual economy with the old technology parameters, but post-liberalization level of trade costs (θ_0, τ_1, w_0). 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. In order 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 or smaller than total

[38]This assumption might seem problematic in a multi-country world, but here I consider China joining WTO, which I interpret as a case of trade liberalization between China and rest of the world. Firms' decision after the trade liberalization is then 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.

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 describe the main steps of estimating my structural model on data.

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

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

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

4. I choose parameter values (θ_0, τ_0, w_0) such that the distance from (3) is minimized.

5. I perform steps (1-4) for the data for post-treatment period ($t = 1$) and find (θ_1, τ_1, w_1).

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

7. I simulate the economy with parameters (θ_0, τ_0, w_0), fix optimal paths for each firm and find average marginal costs for τ_0 and τ_1: $\tilde{MC}\left(\theta_0, \tau_0, w_0\right) |\tau_0$ and $\tilde{MC}\left(\theta_0, \tau_0, w_0\right) |\tau_1$, and find the direct effect of log change in marginal costs $d \ln \tilde{MC}_{Direct} \equiv \ln \left(\frac{\tilde{MC}(\theta_0, \tau_0, w_0)|\tau_1}{\tilde{MC}(\theta_0, \tau_0, w_0)|\tau_0} \right)$.

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

In order 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 choice becomes unresponsive to trade shocks. In case N is large, firms' productivity distribution, by the law of large numbers, converges to a single valued degenerate distribution. My simulations

Table 1: Calibration Results

Year	σ	ζ	Shape	Scale	τ	w
2000	4	0.5	0.42	0.13	0.49	1.18
2007	4	0.5	0.43	0.18	0.45	1.16

suggest that $N = 10$ generates enough flexibility for firms and is consistent with high variability in firms' sizes.[39]

Another key assumption is the distribution of firms' costs on every stage. I use most popular distributions in trade literature such as Pareto (Chaney (2008)), Fréchet (Ramondo and Rodrıguez-Clare (2013)), log-normal (Head et al. (2014)) and generalized gamma (Hanson et al. (2014)) distributions. I find that fat tailed Pareto distribution gives the best results; it happens because empirical distribution of firms' sales has fat tails as well, and in order to generate large number of very productive firms, probability that a firm does not have any bad draws should be high. Empirical distribution of firms' sizes has extremely fat tails, that is why I trim 5% of largest firms in both simulated and empirical samples.

Finally, I need values of demand parameters σ and ζ, that determine the mapping between productivities, sales and welfare. I use standard values from the literature, $\sigma = 4$ and $\zeta = 0.5$.

The results of the calibration are presented in Table 1.

The technology can be described by Pareto distribution with fat tails and a significant value of trade costs. One can see the difference between 2000 and 2007: productivity in both countries changed ambiguously: both shape and scale parameters increased the former decreased the expectation of productivity draw, the latter increased it; relative wage decreased and, most importantly, trade costs fell by approximately 10%.

[39]Potential concern here is that many complex goods consist of thousands of intermediate parts and then, according to the model, all the firms would not differ much in terms of their market sizes. Notice, though, it is true only in case cost draws are independent. In case there is correlation between production costs of different parts, there can be a large variation in firms' market size. 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' sales, the latter increases it. In this paper for the sake of tractability I focus on independent draws and consequently choose not too high N.

With the estimates (θ_0, τ_0, w_0) and (θ_1, τ_1, w_1) I can follow steps 6-8 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 $1 - \frac{d \ln \tilde{MC}_{Direct}}{d \ln \tilde{MC}_{Total}} = 24.6\%$ and $\frac{d \ln \tilde{MC}_{Direct}}{d \ln \tilde{MC}_{Total}} = 75.4\%$.

6 Conclusion

Understanding how global firms make decisions and what consequences it has for trade outcomes is an important task. Baldwin and Venables (2013) and Antràs and Rossi-Hansberg (2008) show that predictions of the models at the intersection of organizational economics and international trade strongly depend on the assumptions on the production structure. In this paper I consider a large class of firm's problems and introduce a way to numerically solve them.

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

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

I show that irrespective 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, increase their productivity and lower their prices. Better technology increases entry and leads to higher variety for consumers.

I propose a simple general equilibrium framework in order 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 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 25% 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.

AMITI, M., M. DAI, R. C. FEENSTRA, AND J. ROMALIS (2016): "How Did China?s WTO Entry Benefit US Consumers?" Tech. rep., Working Paper.

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," Tech. rep., National Bureau of Economic Research.

ANTRAS, P., T. C. FORT, AND F. TINTELNOT (2014): "The margins of global sourcing: theory and evidence from US firms," Tech. rep., National Bureau of Economic Research.

ANTRÀS, P. AND E. HELPMAN (2003): "Global sourcing," Tech. rep., National Bureau of Economic Research.

ANTRÀS, P. AND E. ROSSI-HANSBERG (2008): "Organizations and trade," Tech. rep., National Bureau of Economic Research.

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.

BASCO, S. AND M. MESTIERI (2013): "Heterogeneous trade costs and wage inequality: A model of two globalizations," *Journal of International Economics*, 89, 393–406.

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.

CALIENDO, L. AND E. ROSSI-HANSBERG (2012): "THE IMPACT OF TRADE ON ORGANIZATION AND PRODUCTIVITY," *The Quarterly Journal of Economics*, 1393, 1467.

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., S. KORTUM, AND F. KRAMARZ (2011): "An anatomy of international trade: Evidence from French firms," *Econometrica*, 79, 1453–1498.

EGUCHI, S. AND J. COPAS (2006): "Interpreting kullback–leibler divergence with the neyman–pearson lemma," *Journal of Multivariate Analysis*, 97, 2034–2040.

FALLY, T. AND R. HILLBERRY (2014): "A Coasian Model of International Production Chains," .

FEENSTRA, R. C. AND G. H. HANSON (1996): "Globalization, outsourcing, and wage inequality," Tech. rep., National Bureau of Economic Research.

FORT, T. (2013): "Breaking up is hard to do: Why firms fragment production across locations," *US Census Bureau Center for Economic Studies Paper No. CES-WP-13-35*.

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. KHANDELWAL, N. PAVCNIK, AND P. TOPALOVA (2008): "Imported intermediate inputs and domestic product growth: Evidence from India," Tech. rep., National Bureau of Economic Research.

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.

HANSON, G. H., N. LIND, AND M.-A. MUENDLER (2014): "The Dynamics of Comparative Advantage," .

HARRIGAN, J. AND A. J. VENABLES (2006): "Timeliness and agglomeration," *Journal of Urban Economics*, 59, 300–316.

HEAD, K., T. MAYER, AND M. THOENIG (2014): "Welfare and trade without pareto," *The American Economic Review*, 104, 310–316.

HELPMAN, E., M. J. MELITZ, AND S. R. YEAPLE (2003): "Export versus FDI," Tech. rep., National Bureau of Economic Research.

HSIEH, C.-T. AND P. J. KLENOW (2007): "Misallocation and manufacturing TFP in China and India," Tech. rep., National Bureau of Economic Research.

HUMMELS, D. L., D. RAPOPORT, AND K.-M. YI (1998): "Vertical specialization and the changing nature of world trade," *Economic Policy Review*, 4.

JOHNSON, R. C. AND A. MOXNES (2013): "Technology, trade costs, and the pattern of trade with multi-stage production," .

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

MRÁZOVÁ, M., J. P. NEARY, AND M. PARENTI (2016): "SALES AND MARKUP DISPERSION: THEORY AND EMPIRICS," .

OLLEY, G. S. AND A. PAKES (1992): "The dynamics of productivity in the telecommunications equipment industry," Tech. rep., National Bureau of Economic Research.

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.

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

TINTELNOT, F. (2014): "Global production with export platforms," Tech. rep., WTO Staff Working Paper.

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. (2003): "Can vertical specialization explain the growth of world trade?" *Journal of political Economy*, 111, 52–102.

——— (2010): "Can multistage production explain the home bias in trade?" *The American Economic Review*, 100, 364–393.