The Impact of Trade on Organization and Productivity[*]

Lorenzo Caliendo   Esteban Rossi-Hansberg

Yale University   Princeton University

May 19, 2011

Abstract

A firm's productivity depends on how production is organized given the level of demand for its product. To capture this mechanism, we develop a theory of an economy where firms with heterogeneous demands use labor and knowledge to produce. Entrepreneurs decide the number of layers of management and the knowledge and span of control of each agent. As a result, in the theory, heterogeneity in demand leads to heterogeneity in productivity and other firms' outcomes. We use the theory to analyze the impact of international trade on organization and calibrate the model to the U.S. economy. Our results indicate that, as a result of a bilateral trade liberalization, firms that export will increase the number of layers of management and will decentralize decisions. The new organization of the average exporter results in higher productivity, although the responses of productivity are heterogeneous across these firms. In contrast, non-exporters reduce their number of layers, decentralization, and, on average, their productivity. The marginal exporter increases its productivity by about 1% and its revenue productivity by about 1.8%.

1. INTRODUCTION

Production requires organization. In order to produce firms need to hire managers that organize inputs, that solve the problems that arise in production, and that plan for the future, among other tasks. As a result, the productivity of a firm depends on the way its production is organized. Of course, how much a firm is willing to invest in organizing its production by hiring a hierarchy of managers, with different amounts of knowledge and particular tasks within the firm, depends on the scale of its output. It would be clearly inefficient for a small family business to have the same organizational structure as a large multinational firm that sells all around the world. These differences in organization are obvious for anyone that has seen the complicated organizational hierarchy of a firm like IBM and, at the same time, has bought anything in a mom-and-pop shop around the corner. The implications of organizational decisions for the gains from

1

international trade and other changes in the economic environment are less obvious. In fact, organizational structure is absent in virtually all international trade theories.[1] How do these organizational differences affect productivity? How are organization and productivity impacted by firm decisions that change its scale, like the decision to export? What are the aggregate implications of these within-firm changes?

We study the internal organization of firms in a product market. The market is formed by a continuum of differentiated products. The demand for these products varies depending on their individual characteristics. An entrepreneur can enter the market after designing her product, which conveys a fixed cost. A particular design is associated with a level of demand for the product that is random. Once the entrepreneur observes the resulting demand for the product she envisioned, she can organize production to serve local and international markets. The decision of how to organize production in turn determines the productivity of the firm as well as its input use, price, and profits.

We model an economy in which production requires labor and knowledge. We follow Garicano (2000) and Garicano and Rossi-Hansberg (2004, 2006 and 2011) and model firms as knowledge-based hierarchies where managers deal with exceptions. In this production structure workers in the firm provide work time. For each unit of work time they can *potentially* produce A goods. Problems have to be solved in order for output to be realized. In order to solve problems agents acquire specialized knowledge. Acquiring knowledge is costly, and workers can only use their knowledge to solve the problem they face on their own production activity. So in general, it is not efficient for the firm to make workers learn how to solve infrequent problems. Instead, hierarchies are created where managers learn how to solve the less common problems –the 'exceptions'– and workers deal with the routine problems. When workers are confronted with a problem they do not know how to solve, they ask the manager. The organizational structure determines the knowledge acquired by agents, the order in which different agents confront problems, and how they communicate with each other. Organization economizes on the use of knowledge by leveraging the knowledge of managers. A larger firm can have more than one layer of managers, where managers learn less common problems the higher they are in the hierarchy. All firms have a top manager, the CEO or entrepreneur. Of course, adding a layer of management is costly, since one has to pay managers that do not generate production possibilities, but only solve problems. As a result, adding layers involves an extra cost that is only worth paying if the firm produces a high enough quantity. Adding an extra layer of management can be thought of as reducing the marginal cost of the firm, in exchange for increasing the fixed cost. Nevertheless, given that we provide a fully microfounded theory of the organization of production, the menu of fixed costs and marginal costs faced by the firm is endogenous and depends on the cost of acquiring and communicating knowledge as well as other characteristics of the economy.

We incorporate this production technology into an economy in which agents have CES preferences and where firms compete with each other monopolistically. Hence, firms determine the price that maximizes

---

[1] This is clearly the case for the two main models in international trade today, namely, Melitz (2003) and Eaton and Kortum (2002), as well as virtually all other available theories. An exception is the work of Marin and Verdier (2008, 2010), who study the degree to which free trade can centralize the decision process inside an organization.

their profits. As is common with this structure, they set prices using a constant markup rule. However, given that they are making simultaneous decisions about their organizational structure, their marginal cost is now an endogenous variable. In a closed economy, the only exogenous and heterogeneous characteristic of a firm is the level of demand for its product, which we parameterize with a demand shifter α. In an open economy, the firm can decide to export to other markets as well. We fully characterize the cost minimization problem of a firm given a demand level. Crucially, the level of demand determines the productivity of a firm by determining its optimal organizational structure.

Our theory exploits the basic idea that the productivity of a firm is the result of the way production is organized given the level of demand for its product. This link between demand and productivity is present in any production technology in which a firm needs to make some costly investment decision in order to expand production. A classic example is the decision to expand by increasing the number of plants. Another one is the decision to invest in new technologies (as in Atkeson and Burstein, 2010, Bustos, 2011, Costantini and Melitz (2008), and Lileeva and Trefler, 2010). We model the decision to form a management hierarchy by adding new layers of management, since it allows us to study a rich set of internal characteristics of the firm, as well as the effect of learning costs and communication technology on firms' outcomes. These investments allow the firm to produce more and at the same time lower the firm's average cost. Thus, there is a link between the level of demand, the level of investments of the firm, and the productivity of the firm.

There is another link between the level of demand and productivity that is not present in other papers in the literature, namely, the explicit modeling of the discreteness of many of these investment decisions. Take, for example, the decision of how many plants to use for production. If each plant has an identical minimum efficient scale (MES),[2] the first effect discussed above is not present. However, because the number of plants the firm chooses to built is discrete, the productivity of the firm will depend on its production level per plant relative to the MES of a plant. If the firm is producing below the MES in each plant (and average costs at the plant level are convex), an increase in demand will increase the productivity of the firm. If it is producing above the MES (but not by enough to build a new plant), a small increase in demand that does not trigger the construction of a new plant will decrease its productivity. In this sense, the fact that the decision of how many plants to build is discrete will result in heterogeneous responses in productivity to positive demand shocks (or in differences in productivity across firms).[3]

In our theory, both of these mechanisms are present. Since the number of layers in the firm is discrete, the second effect arises naturally. However, because more layers allow the firm to manage the acquisition of knowledge of its labor force more efficiently, more layers also imply a reduction in the minimum average cost of the firm.

The paper is related to the literature on knowledge-based hierarchies. The main distinction between this

---

[2]The minimum efficient scale is defined as the scale of production at which a firm minimizes average costs given a number of layers.

[3]The literature on multi-product firms has considered the decision to add or drop products and the within-firm revenue productivity changes that result from this type of selection. See for example Bernard, et al. (2010) and Mayer, et al. (2011).

paper and the related literature is that here we model production hierarchies in a product market where firms produce differentiated products. This allows us to study the internal organization and productivity of heterogeneous firms in general equilibrium. Furthermore, it allows us to study the effect of a trade liberalization on organization and firm productivity. The result is a model that can be readily contrasted with the data and that delivers clear testable implications. As far as we know, this is the first paper to incorporate a rich organizational model into a heterogeneous firm context.[4]

The vast literature on trade and heterogeneous firms has been successful at providing a framework with which to analyze individual firm experiences in an international trade context. However, in this literature a firm is just a technology to produce goods at a given marginal cost. The basic equilibrium framework of our model is similar to the one used in this literature but incorporates a much richer organizational structure, as well as endogenous productivity heterogeneity across firms and wage heterogeneity across and within firms. In fact, we show that Melitz (2003) is a special case of our model when knowledge becomes costless.

The paper aims to speak to the empirical literature on the response of firm organization and productivity to trade liberalization. The papers by Clerides, Lach and Tybout (1998) and Bernard and Jensen (1999) show that larger and more productive firms export. They also provide evidence that firms that decide to export are larger and have a larger probability of survival, but they do not find evidence that exporting leads to higher productivity.[5] In our framework firms that export are larger in terms of employment, but their productivity does not necessarily rise. Other papers, such as Baldwin and Gu (2003), Van Biesebroeck (2004) and De Loecker (2007, 2011), have found more significant effects of trade liberalization on productivity.[6] These apparently contradictory findings can be rationalized with our theory. We find that the effect of trade on a firm's productivity is heterogeneous and depends on how large firms are before trade liberalization, the size of the new market, export costs, as well as information and communication technology. For example, the effect of trade on the productivity of large firms (in terms of employment or output) is, in general, smaller than the effect of trade on the productivity of small firms. Even for large and small firms, there are heterogeneous responses depending on whether they add management layers, and their position relative to the MES of production of their chosen organizational design. In our model, all exporting firms that gain productivity produce at a higher scale compared to autarky, and many of them add a layer of management as a consequence of starting to export. On the other hand, non-exporters reduce their scale of production compared to autarky. Non-exporters that were producing in autarky above their MES gain in productivity, while non-exporters that were producing at the MES, or below, reduce their productivity unless they reduce

---

[4] Antràs, et al. (2006 and 2008) study the implications of offshoring on wage inequality using a similar model of organizations. In contrast to this paper they model an economy in which firm heterogeneity is the result of worker heterogeneity and study how the ability to form international teams changes the composition of two-layer production teams, all of which produce a single good. Here we abstract from exogenous worker heterogeneity and let knowledge acquisition and the number of layers be endogenously determined. Perhaps more important, we model an economy with firm heterogeneity in demand in which trade is the result of products being differentiated as in Melitz (2003).

[5] Other papers that document similar findings are Delgado, et al. (2002) and Bernard and Jensen (2004).

[6] Again, other papers are Aw, et al. (2000), Hallward-Driemeier, et al. (2005), Fernandes and Isgut (2006), Park, et al. (2006), Aw, et al. (2007), and Aw, et al. (2008).

the number of management layers in the organization.

Heterogeneous responses in productivity as a result of a bilateral trade liberalization have been recently documented by Lileeva and Trefler (2010). They study the responses of productivity and technology upgrading for Canadian plants after CUSFTA went into effect. They show that plants with low pre-agreement productivity invested more in information and communication technologies, as well as product innovation. Any of these investments could be attributed to changes in their organizational structure, but the first two are more closely related. They also document that new exporting firms that experienced productivity gains increased their domestic sales relative to non-exporters, as implied by our theory. A recent paper by Bustos (2011) also finds a heterogeneous response in technology upgrading. She finds that reductions in tariffs induce process innovation mostly in the 3rd quartile of the firm size distribution and not only in new but also in old exporters, Again, a heterogeneous response as predicted by our theory.

Regarding the internal reorganization of firms, Guadalupe and Wulf (2010) present evidence that increasing competition (as a result of international trade) leads to flatter hierarchies, something our model can explain for the non-exporting firms. They also show that the opportunity to sell in more markets can lead to firms with more management layers, although the results here are more noisy. Given the heterogeneity in the responses to trade liberalization in our theory, looking at just average responses necessarily leads to less accurate results. So again, our theory can rationalize these findings. More important, by providing a structural framework of trade with organization we can relate the different dimensions of firms' actions and so can guide future empirical research on the determinants of within-firm responses to trade policy and technological changes.

The rest of the paper is structured as follows. The next section presents the production technology and solves the cost minimization problem of firms. We also study the profit maximization problem of firms and the equilibrium of the model in a closed economy. Section 3 presents our results for the open economy. In Section 4 we calibrate the model to the U.S. economy and perform several trade liberalization experiments to illustrate our results. Finally, Section 5 concludes. We relegate all proofs to the Appendix.

2. THE MODEL

The economy is formed by $\tilde{N}$ agents that supply one unit of time. All agents are ex-ante identical and order consumption baskets according to constant elasticity of substitution preferences given by

$$U\left(x\left(\cdot\right)\right) = \left(\int_{\Omega} \alpha^{\frac{1}{\sigma}} x\left(\alpha\right)^{\frac{\sigma-1}{\sigma}} M\mu\left(\alpha\right) d\alpha\right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

where $x\left(\alpha\right)$ denotes the consumption of variety α, M is the mass of products available to the consumer, $\mu\left(\cdot\right)$ the probability distribution over the available varieties in Ω, and $\sigma > 1$ the elasticity of substitution. Note that α is a demand shifter that implies that agents like varieties with higher α better.

Agents can work either in the production sector or in the schooling sector. Since agents are ex-ante identical in equilibrium, all of them obtain an equilibrium wage w as compensation for the unit of time they supply in

the market (they are also compensated for the knowledge they acquire, as we discuss below). We call agents that want to start a new firm entrepreneurs. An entrepreneur pays a fixed cost f^E to design its product. After the product is developed, the demand for the product, α, is realized from a cumulative distribution function given by $G(\alpha)$. This draw determines the demand schedule faced by the firm. The entrepreneur then creates an organization if her draw is high enough to obtain profits given the fixed production cost f.

Production requires labor and knowledge. Within an organization agents divide into two main classes: workers and problem solvers or managers. Workers supply their unit of time to generate a production possibility. For this production possibility to become output, the organization needs to solve a problem. Solving problems requires knowledge. In particular, each production possibility draws a problem from the cumulative distribution $F(\cdot)$ with decreasing density. A realization z implies that in order to solve the problem the worker needs to have acquired a set of knowledge that includes z as an element. If the worker's knowledge includes z, the problem is solved and the production possibility becomes A units of output. If the worker cannot solve the problem, she asks a manager one layer above for a solution, if the firm has hired such a manager. Suppose it has. Then the manager spends h units of time listening to the worker's problem (so h is the cost of communication) and solves the problem if her knowledge includes z. If the knowledge of the lowest layer of managers cannot solve the problem, the worker has the option to send the problem one layer higher, and so on, until the problem reaches the top layer, formed by a single agent, namely, the entrepreneur.

The organizational design of the firm involves determining the number of layers as well as the knowledge and number of agents in each layer. The cost of learning is linear in the euclidean length of the knowledge set. So if an agent learns how to solve problems in the interval $[0, z]$, she has to pay wcz. That is, the cost of knowledge is wc per unit of knowledge. Of course, since $F(\cdot)$ has a decreasing density, agents at the bottom of the hierarchy, which attempt to solve the problem first, learn the most common problems, and agents in higher layers learn more rare and exceptional problems progressively. We assume that knowledge is not cumulative so in order to solve a problem z the agent needs to know an interval that includes z, but that interval does not need to include the knowledge of how to solve the most common problems. The cost of knowledge is proportional to the wage, since learning requires teachers in the schooling sector that earn w. So learning one unit of knowledge requires c units of time of a teacher at wage w.[7]

Consider first the case of a self-employed entrepreneur, a one-worker firm. The entrepreneur draws problems from the known exponential distribution of problems, $F(\cdot)$, and acquires knowledge $[0, z]$ to solve them. For simplicity, throughout the paper we let $F(z) = 1 - e^{-\lambda z}$ and we discuss the implications of this choice as we describe our results.[8] The parameter $\lambda > 0$ regulates how common the problems faced in production

---

[7]Note that this interpretation implicitly assumes that every period workers need to be retrained. Since we consider a general equilibrium model in steady state that is calibrated to the share of education workers, one can reinterpret c as the cost of training workers times the share of workers that die every period. In that case, the cost of knowledge is $c = \delta_M \tilde{c}$ where δ_M is the population mortality rate and $\tilde{c}$ is the cost of training workers when only new workers have to be trained. We return to this point in the calibration of the model, where our calibration under this interpretation implies that the stock of knowledge in units of workers is equivalent to about 4 times the population.

[8]The decreasing density of problems is just a normalization, since agents always choose to learn how to solve the most

are. Higher λ implies that problems are less costly to solve since the distribution has more mass close to zero.[9] Since the entrepreneur has only one unit of time and since she is the sole worker in the firm, she has to spend her time generating production possibilities. This implies that even if she knows how to solve all problems, her output is bounded by the amount of output generated by a single production possibility, namely, A. The parameter A reflects the overall technology of the economy. So if the knowledge acquired by the entrepreneur is z_0^0, the fraction of problems that she solves is $F\left(z_0^0\right) = 1 - e^{-\lambda z_0^0}$, and the expected output she produces is $A[1 - e^{-\lambda z_0^0}]$. Throughout, we let z_L^ℓ denote the knowledge of an agent in layer ℓ of a firm with highest layer L (and therefore $L + 1$ layers, since we denote the layer of workers by 0). We follow the same notation convention for all variables.

Suppose now the entrepreneur decides to hire some workers and form a hierarchical organization with two layers where she is at the top of the organization. The entrepreneur hires workers trained to solve the more routine problems, z_1^0. Denote by n_1^0 the number of workers she hires. Since she is the only top manager, $n_1^1 = 1$ (more generally we impose the constraint $n_L^L = 1$). Each worker solves a fraction of problems $[1 - e^{-\lambda z_1^0}]$ and the problems that they cannot solve are directed to the entrepreneur. Since the entrepreneur spends h units of time communicating with each worker that sends her a problem to solve, it follows that she can deal with at most $1/h$ problems. Her time constraint is then given by $h n_1^0 e^{-\lambda z_1^0} = n_1^1 = 1$ and so she can hire at most $n_1^0 = e^{\lambda z_1^0}/h$ workers.

We can easily generalize this structure to a firm with $L + 1$ layers in which workers ask managers to solve their problems sequentially (this flow of questions is optimal, as proven by Garicano, 2000). Then, the span of control of each manager depends on the knowledge of the managers/workers below them, namely, $n_L^\ell / n_L^{\ell+1} = e^{\lambda z_L^\ell}$, for all $0 < \ell < L$, and $n_L^0/n_L^1 = e^{\lambda z_L^0}/h$. We can apply the law of large numbers to this problem and use the expectation of output as the realized level of output.[10]

We solve the organizational problem of the firm in two stages. We determine first the cost-efficient way of producing a given quantity. The solution to the cost minimization problem characterizes the cost function. Then, given a particular draw of α, the entrepreneur acts like a monopolist and decides how to price the goods and how much output to supply.

2.1. Cost Minimization

Consider a firm that produces a quantity q. The variable cost function of such a firm is given by

$$C\left(q; w\right) = \min_{L \geq 0}\left\{C_L\left(q; w\right)\right\} \tag{2}$$

---

common problems first. The exponential specification is not essential but adds some tractability to the model.

[9]In the limit when $\lambda \to \infty$, knowledge is not an input in production since all problems can be solved by agents with an infinitesimal amount of knowledge.

[10]In order to apply the law of large numbers (LLN), suppose that workers generate a production possibility in dt units of time. Each of these production possibilities can yield $A\,dt$ units of output and managers spend $h\,dt$ units of their time with each problem that reaches them. Then, as $dt \to 0$, the number of problems generated by the worker in a unit of time, $1/dt$, converges to ∞. The LLN then implies that the worker can solve a fraction $F\left(z_L^0\right)$ of them, producing $A F\left(z_L^0\right)$ units of output. The rest of the problems are sent to the manager above, who spends a fraction $h\left(1 - F\left(z_L^0\right)\right)$ of her unit of time on them.

where $C(q; w)$ denotes the minimum variable cost of producing q and $C_L(q; w)$ is the minimum cost of producing q with an organization with $L + 1$ layers. Since the fixed costs of production do not change the organizational problem of the firm, in this section we consider only the variable cost $C(q; w)$, to which we would need to add fixed costs to obtain total costs. We do so in the profit maximization section below. Throughout, to simplify the exposition, we refer to this function as the 'cost function.'

The problem of a one-worker firm can help us understand why there are incentives to form organizations with more than one layer. The cost minimization problem of a firm with $L = 0$ is then given by

$$C_0(q; w) = \min_{z \geq 0} w(1 + cz), \text{ subject to } A[1 - e^{-\lambda z}] \geq q,$$

where we impose that an agent that has acquired knowledge z at cost wcz is compensated exactly for her learning cost. This is always the case in our economy given that knowledge can be instantaneously acquired at this cost. An organization with only one layer ($L = 0$) is not able to produce $q > A$. In that case we just set $C_0(q; w) = \infty$. The Lagrangian of the problem is

$$\mathcal{L}(q, w, \phi, z, \theta) = w(1 + cz) + \phi\left[q - A\left[1 - e^{-\lambda z}\right]\right] + \theta z$$

and so the first-order conditions are given by

$$wc = \phi A \lambda e^{-\lambda z_0^0} - \theta \text{ and } A\left[1 - e^{-\lambda z_0^0}\right] = q,$$

where z_0^0 is the optimal amount of knowledge. As noted before, the maximum that a self-employed entrepreneur can produce is A, since $\lim_{z \to \infty} A\left[1 - e^{-\lambda z}\right] = A$. However, she never chooses this limit, since in order to produce A she needs $z \to \infty$ and so her costs converge to infinity. The optimal amount of knowledge that the entrepreneur chooses as a function of quantity is given by $z_0^0(q) = \frac{1}{\lambda} \ln\left(\frac{A}{A-q}\right)$, which is finite since $q < A$. An entrepreneur working alone is not able to solve all the problems that she faces in production and therefore throws away a fraction $e^{-\lambda z_0^0}$. If there was someone else in the organization that knew how to solve some of these more exceptional problems, the organization would be able to produce more output. This creates an incentive to hire agents specializing in different types of problems. Furthermore, an organization with more layers can increase the span of control of the entrepreneur by hiring more workers that create more production possibilities. Of course, since adding layers is costly, for low values of q the cost-efficient way of producing might be to have a one-worker firm.

Using the first-order conditions we can solve for the cost function for a one-layer (and one-worker) firm to obtain a cost function given by

$$C_0(q; w) = w\left(\frac{c}{\lambda} \ln\left(\frac{A}{A-q}\right) + 1\right),$$

with marginal cost

$$MC_0(q; w) = \frac{wc}{\lambda}\frac{1}{A-q} = \frac{wc}{\lambda A}e^{\lambda z_0^0}$$

and average costs equal to $AC_0(q; w) = C_0(q; w)/q$.

Note that the average cost of this problem is strictly convex. This implies that there is a MES of production for a one-layer firm. The level of demand of the firm, α, determines if the firm decides to produce at this efficient scale. Figure 1 presents an example of the average cost and the marginal cost of an organization with $L = 0$. As the firm approaches its maximum capacity of production, A, the average cost converges to infinity (only one-worker firms have a maximum capacity, as we will see below).

Now consider the minimization problem of an organization with $L > 0$ layers:

$$C_L(q;w) = \min_{\{n_L^l, z_L^l\}_{l=0}^L \geq 0} \sum_{l=0}^L n_L^l w \left[c z_L^l + 1 \right]$$

subject to

$$A \left[1 - e^{-\lambda Z_L^L} \right] n_L^0 \geq q,$$

$$n_L^l = n_L^0 h e^{-\lambda Z_L^{l-1}} \text{ for } L \geq l > 0 \text{ and } n_L^L = 1$$

where $Z_L^l = \sum_{\ell=0}^l z_L^\ell$ is the cumulative knowledge of the firm up to layer l. Since the knowledge of agents in two distinct layers never overlaps, if an agent learns an interval of size z_L^l, it knows how to solve problems in the interval $\left[Z_L^{l-1}, Z_L^{l-1} + z_L^l \right]$ or $\left[Z_L^{l-1}, Z_L^l \right]$. Agents at lower layers learn more-common problems in order to economize on the time of their managers. Hence, the resulting distribution of problems can always be normalized to have a decreasing density, like the exponential. So, for a given number of layers, L, the entrepreneur decides the number of workers to hire in each layer (n_L^l) and the knowledge of the employees at each layer (z_L^l) subject to the amount of output she wants to produce and the time constraints.
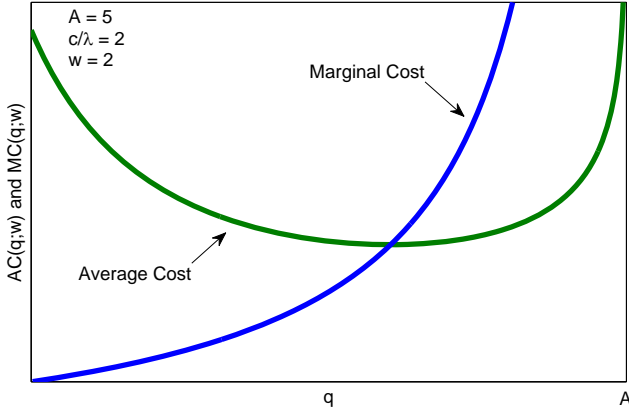

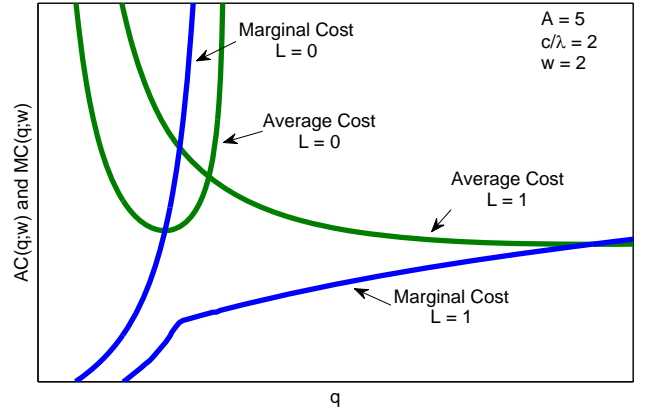Figure 1: Average and marginal cost, $L = 0$

Figure 2: Average and marginal cost, $L = \{0, 1\}$

The entrepreneur is always at the top of the organization, and so $n_L^L = 1$. Therefore, $n_L^0 = e^{\lambda Z_L^{L-1}}/h$.

After substituting the time constraints for labor, the Lagrangian for the cost minimization problem becomes

$$
\mathcal{L}\left(q, w, z, \phi, \left\{z_L^l, \theta_L^l\right\}_{l=0}^L\right) = \frac{w\left(cz_L^0 + 1\right) + h\sum_{l=0}^{L-1} e^{-\lambda Z_L^l} w\left(cz_L^{l+1} + 1\right)}{he^{-\lambda Z_L^{L-1}}}
$$
$$
+\phi\left(q - A\frac{1 - e^{-\lambda Z_L^L}}{he^{-\lambda Z_L^{L-1}}}\right) + \sum_{l=0}^L \theta_L^l z_L^l
$$

where ϕ and $\left\{\theta^l\right\}_{l=0}^L$ are the Lagrange multipliers associated with both constraints. The first-order condition with respect to z_L^0 is

$$
c + \lambda\left(cz_L^0 + 1\right) - \frac{\phi}{w}\lambda A + \frac{\theta_L^0}{wn_L^0} \leq 0.
$$

The first-order conditions for z_L^l, all $0 < l < L$, are

$$
he^{-\lambda Z_L^{l-1}} c + \lambda\left(cz_L^0 + 1\right) + \lambda h\sum_{k=0}^{l-1} e^{-\lambda Z_L^k}\left(cz_L^{k+1} + 1\right) - \lambda A\frac{\phi}{w} + \frac{\theta_L^l}{wn_L^0} \leq 0 \tag{3}
$$

and the first-order condition for the knowledge of the entrepreneur z_L^L is

$$
c - \frac{\phi}{w}\frac{\lambda A}{h}e^{-\lambda z_L^L} + \frac{\theta_L^L}{w} \leq 0. \tag{4}
$$

The rest of the first-order conditions are given by the constraints and

$$
\phi \geq 0, \ \phi\frac{\partial\mathcal{L}}{\partial\phi} = 0, \ \theta_L^l \leq 0, \ \theta_L^l z_L^l = 0 \text{ for all } l.
$$

Note that we allow the entrepreneur to choose zero knowledge at any layer. It is easy to prove that a firm that selects optimally the number of layers never chooses to have intermediate managers with zero knowledge at any layer. The reason is that these managers cost the firm w but, since they are specialized in problem solving but cannot solve any problems, they do not add anything to the production process. Under a restriction on parameters, summarized in Assumption 1, we can guarantee that entrepreneurs and workers will also have positive knowledge. If the parameter restriction does not hold, it could be the case that either workers have zero knowledge or workers and entrepreneurs have zero knowledge. This does not change any of the results that follow, but it does imply that we would need to carry the non-negativity constraints for the knowledge of workers and entrepreneurs throughout. Hence, we impose Assumption 1 in order to simplify the exposition.

Assumption 1 *The parameters λ, c, and h are such that $\frac{c}{\lambda} \leq \frac{h}{1-h}$.*

Proposition 1 shows that Assumption 1 is sufficient to guarantee that, for any organization with $L \neq 1$ management layers, the entrepreneur never sets knowledge to zero at any layer. The intuition of the result is as follows. There are two possible ways in which the firm can reduce production, either by decreasing the knowledge of workers and problem solvers at each layer of the organization, or by reducing the number of layers of the organization. Assumption 1 guarantees that the cost of communication is large enough relative

to the cost of learning, so the firm always prefers to reduce the number of layers before setting the knowledge of its employees to zero.

Proposition 1 *Under Assumption 1, for all $L \neq 1$ and any output level q, the knowledge of agents at all layers is positive ($z_L^l \geq 0$ never binds).*

All proofs are relegated to the Appendix. The case of $L = 1$ is the only one where a simple parameter restriction cannot guarantee that $z_1^0 > 0$. The reason is that an organization with zero management layers is constrained to produce $q < A$. Hence, even if workers do not learn anything, an organization with one layer of managers can dominate the self-employed firm when the quantity produced is close enough to A or larger. However, all the results below hold for the case of $L = 1$ as well. With these results in hand, let us set the Lagrange multiplier for the non-negative knowledge constraint to zero, $\theta_L^l = 0$ for all l. Strictly speaking we still could have zero knowledge for q values and layers that are never optimally chosen by the firm. However, given that these ranges are irrelevant for our purposes, we proceed as if the constraint never binds.

We first characterize the marginal cost curve given L. Note that the marginal cost of a firm is equal to the value of ϕ (the Lagrange multiplier of the output constraint). Hence, the first order condition with respect to z_L^L yields

$$MC_L(q; w) = \frac{\partial C_L(q; w)}{\partial q} = \phi = \frac{wch}{\lambda A} e^{\lambda z_L^L}. \tag{5}$$

As we can see, the marginal cost is an increasing function of the knowledge of the entrepreneur, z_L^L, the cost of acquiring knowledge relative to the knowledge requirements in production, wc/λ, and communication technology h. Recall that the higher h is, the worse the communication technology and so the less efficient are employees at communicating problems. Naturally, the marginal cost is a decreasing function of the level of technology, A. Proposition 2 below shows that, given the number of layers L, the marginal cost of a firm is increasing in output, q. The result implies that the knowledge of all agents is increasing in q. That is, given the number of layers, increasing the quantity produced implies that the chosen organizational structure becomes more restrictive and so the cost of producing an extra unit increases.

Proposition 2 *Given the number of layers L, the marginal cost, $MC_L(q; w)$, and the knowledge of all employees, z_L^l for all l, increase with q. Furthermore, $MC_L(q; w) > 0$, for all L and q, and so the cost function $C(q; w)$ is strictly increasing.*

Now consider the change in marginal costs as the firm decides to increase the number of layers. We show below that when a firm increases the number of layers, it also increases discontinuously the quantity produced. For now, we consider q's at the point of a switch in layers (when the argmin of (2) increases), and study the change in marginal cost as we increase q. Proposition 3 proves that the marginal cost at this point always jumps down discontinuously. The reason is simple: when the firm increases the number of layers, it invests in the ability to produce more output. Hence the marginal cost of producing an extra unit decreases.

Proposition 3 *Suppose that the entrepreneur finds it optimal to increase the number of layers at output level q, then the marginal cost decreases discontinuously at q.*

Since firms operate in a monopolistic competition environment, this characterization of the marginal cost is useful to understand how prices differ across organizations. It also reveals that investing in a new layer is equivalent to expanding the production capacity of the firm.

Let us now study the cost function more generally. From the first-order conditions we can solve for the cost function of any firm with a number of layers L to obtain

$$C_L\left(q;w\right) = w\left(\frac{ch}{\lambda A}e^{\lambda z_L^L}q + \frac{c}{\lambda}\left(1 - e^{\lambda z_L^{L-1}}\right) + cz_L^L + 1\right),$$

where z_L^L and z_L^{L-1} are the solutions to the first-order conditions (3) and (4).[11] The average cost function of a firm with L layers is then given by

$$AC_L\left(q;w\right) = \frac{wc}{\lambda}\left(\frac{h}{A}e^{\lambda z_L^L} + \frac{1 - e^{\lambda z_L^{L-1}} + \lambda z_L^L + \lambda/c}{q}\right). \tag{6}$$

The average cost functions for different L's all share a number of characteristics. They are all homogeneous of degree one with respect to factor prices. The two factor prices are cw for knowledge and w for a worker's time. So this implies that the cost functions are homogeneous of degree one in w. The choice of knowledge levels is invariant to w, since it affects all costs proportionally. Finally, the average cost curves are convex in q. When production is small, an increase in production reduces average costs up to the MES. Further increases in output increase average cost: The firm needs to provide too much knowledge to agents in the lower layers of the organization to achieve such a large scale. These results are proven in Proposition 4.

Proposition 4 *Given the number of layers L, the cost function and the corresponding factor demands have the following properties: i. The cost function is homogeneous of degree one in w. ii. Conditional factor demands are homogeneous of degree zero in w. iii. The average cost curve is convex in q, has its minimum where it crosses the marginal cost curve, and converges to infinity when $q \to 0$ or $q \to \infty$.*

Figure 2 illustrates part iii. of Proposition 4. It shows the convex average cost functions for organizations with one and two layers, together with the corresponding marginal cost curves. It is evident in Figure 2 that the minima of the average cost curves decrease with the number of layers. We show this result formally in the next proposition. The intuition is that firms with more layers are able to economize on the knowledge acquisition of their employees by leveraging on the knowledge of its managers at higher scales of output. The firm is paying more 'fixed' costs to have a large organization and this allows it to produce at lower average costs, as long as its scale is large enough.

---

[11]The cost function for the case of $L = 1$ is slightly different. In particular,

$$C_1\left(q;w\right) = w\left(\frac{ch}{\lambda A}e^{\lambda z_1^1}q + \frac{c}{\lambda h}\left(h - e^{\lambda z_1^0}\right) + cz_1^1 + 1\right).$$

Proposition 5 *The minimum average cost for a given number of layers decreases with the number of layers.*

Figure 3 presents an example of the average cost curves for an organization with up to four layers ($L = 3$). As we can see, the minimum average cost is decreasing in the number of layers. Furthermore, the level of output that minimizes average cost is increasing with the number of layers. This implies that the average cost curves cross between the MES of the firms with adjacent numbers of layers. This is relevant for our purposes, since it implies that the average cost function at the MES, and the region around it, will be part of the lower envelope that determines the average cost curve when the number of layers is optimized as well, as in (2). We prove these results in Proposition 6.
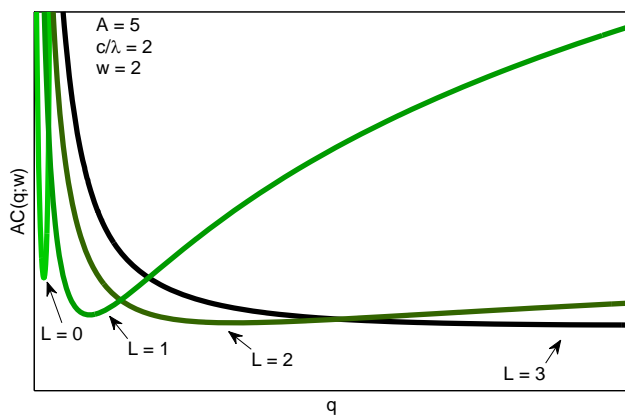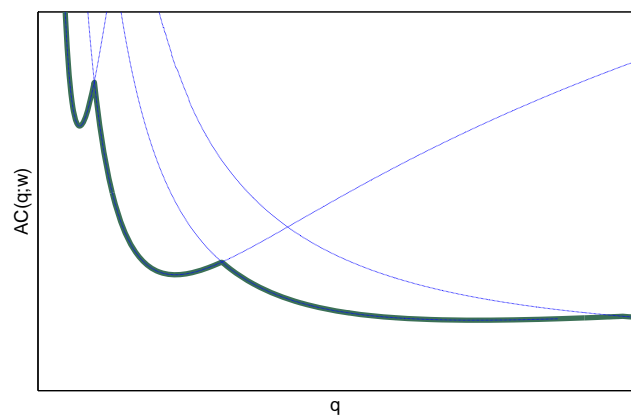

Figure 3: Average cost for several L's

Figure 4: The cost function

Proposition 6 *The level of output that achieves the MES, q_L^*, increases with the number of layers. Furthermore, the average cost curves of an organization with L and $L + 1$ layers cross in the interval $\left(q_L^*, q_{L+1}^*\right).$*

The results we have proven in this section imply that the average cost function that solves the problem in (2) has the shape illustrated in Figure 4. This curve is the lower envelope of the average cost curves given the number of layers. It has an overall decreasing shape, but it is not monotone. Increases in q can increase average costs, since they move the firm away from the MES but not by enough to trigger a change in the number of layers. Note also that Proposition 6 implies that firms that choose to produce higher quantities choose to have a weakly higher number of management layers. The shape of this function is critical in what follows and for our main results on the impact of trade on organization and productivity.[12]

---

[12] Consider the case in which firms can choose continuously their number of layers. Then, the average cost function would be the lower envelope of average cost functions that change continuously. In fact, the average cost function would be formed only by the minimum of the average costs. Since we know that the minimum of the average cost is decreasing in the number of layers, the resulting average and marginal cost functions would be continuous and strictly decreasing in quantities. Such a model is a straightforward extension of our setup, but the interpretation of a continuum of layers is challenging and hard to

Communication technology, h, and the cost of acquiring knowledge about a fraction of problems, c/λ, are key to determining the average cost function. Figures 5 and 6 show an example of the effect of changes in h and c/λ, respectively. First note that in both cases higher h or c/λ increases the average cost for all q's beyond $L = 0$. When the entrepreneur decides to work on her own ($L = 0$), h plays no role in determining average costs, since there is no communication inside the firm. In contrast, the average cost curve increases with c/λ for all q, including the ones where the entrepreneur decides to be self-employed, since even then she has to acquire some knowledge. Note also that the higher h or the lower c/λ, the flatter are the average cost curves beyond layer zero. As we increase the number of layers to more than one, moving beyond the MES is more costly the higher any of these costs.
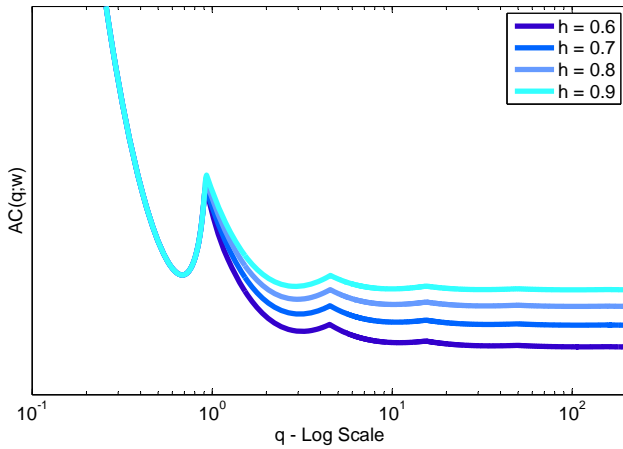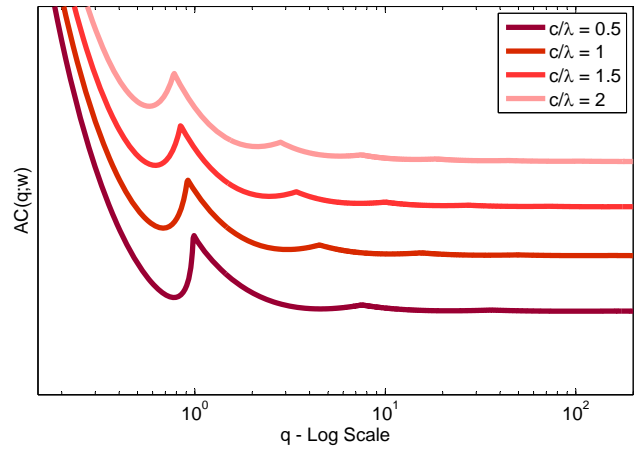

Figure 5: Changing h ($c/\lambda = 1$)

Figure 6: Changing c/λ ($h = 0.7$)

We now turn to the characterization of the cost function when $c/\lambda \to 0$, that is, what happens in this economy when knowledge becomes either unnecessary to solve problems or extremely cheap. Remember that the role of organization in our economy is to economize on the acquisition and communication of knowledge. In this case organization has no purpose. This implies that there would be no firm with more than two layers. Firms would still have two layers instead of one because otherwise their production is limited to A. In fact, if we eliminate the possibility of self-employment, one can show that in this limit the cost function is as in Krugman (1980) or Melitz (2003) with an additional fixed cost. Namely, the marginal cost is constant, the average cost is decreasing due to fixed costs, and the cost function involves a fixed cost plus a variable cost equal to output times the marginal cost. Hence, in the limit without self-employment, our technology includes the standard technology used in trade models as a special case. Overall, since the model specification that follows is identical to that in Melitz (2003), with heterogeneity in demand rather than in marginal costs, the special case of our model when $c/\lambda \to 0$ and when self-employment is not an option is isomorphic to

reconcile with reality, since we observe firms choosing a discrete and small number of layers (as well as plants, products, etc.). Hence, this extension eliminates an essential part of a firm's organizational problem: namely, the many discrete decisions of firms.

Melitz (2003).[13] The result is proven in Proposition 7.

Proposition 7 *In the limit when $c/\lambda \to 0$ and $L \geq 1$, the cost function is given by*

$$C(q;w) = w\left(\frac{q}{A} + 1\right)$$

and so

$$AC(q;w) = w\left(\frac{1}{A} + \frac{1}{q}\right)$$

and

$$MC(q;w) = \frac{w}{A}.$$

2.1.1. Measuring Productivity.—

Measuring productivity at the firm level involves more than just looking at marginal costs. The firm is using labor and knowledge as inputs in production. Hence, to measure productivity we need to aggregate variable factors. Of course, the cost function gives us an appropriate way to do so. Throughout, we measure the productivity of a firm producing a quantity q, $a(q)$, as the inverse of variable average costs, that is, average costs without considering the fixed costs of production. This is done by taking the inverse of the AC function we defined and characterized above, but setting the wage to one (since the input bundle is measured in units of workers so as to keep factor prices constant). Hence, our measure of firm-level productivity is given by

$$a(q) = \frac{q}{C(q;1)} = \frac{1}{AC(q;1)},$$

where the average cost is net of any fixed costs of production.

It is informative to calculate what this measure of productivity implies for the case of Proposition 7 in which $c/\lambda \to 0$ and $L \geq 1$. In this limit the model generates another fixed cost, so total fixed costs are given by $w + f$. Since the cost function, as we defined it above, includes w but not wf, we just need to subtract w. Hence, the inverse of the average cost function net of fixed costs is

$$a(q) = \frac{q}{\lim_{c/\lambda \to 0} C(q;1) - 1} = \frac{q}{\left(\frac{q}{A} + 1\right) - 1} = A.$$

That is, in the case where the marginal cost is given by $1/A$, our measure of productivity results in exactly A. Hence, for this case our measure of productivity coincides exactly with the one in Melitz (2003) and others. Clearly when $c/\lambda > 0$, productivity is not solely determined by A but is a function of how much

---

[13] The need to eliminate the possibility of self-employment to obtain the standard technology arises from the fact that when $L = 0$ the entrepreneur can act as manager and worker at the same time. Allowing for self-employment would amount to having an extra technology in Melitz (2003) where a firm can produce up to A without paying the fixed cost of production. In this case the cost function at the limit when $c/\lambda \to 0$ would be given by

$$C(q;w) = \begin{cases} w & \text{if } q \leq A \\ w\left(\frac{q}{A} + 1\right) & \text{if } q \geq A \end{cases}.$$

the firm produces (and the corresponding organizational structure) given its demand draw α, as well as communication and information technology. That is, in our framework, differences in demand are mapped into differences in technology via the organizational decisions of the firm.

2.2. Profit Maximization

Consider the profit maximization problem of an entrepreneur with demand draw α and the cost function, $C(q;w)$, discussed in the previous section. The entrepreneur chooses the quantity supplied in the market and its price subject to her demand draw. Households have the iso-elastic preferences in (1), and so the demand schedule faced by an entrepreneur with draw α is given by

$$p(\alpha) = q(\alpha)^{-\frac{1}{\sigma}} (\alpha R)^{\frac{1}{\sigma}} P^{\frac{\sigma-1}{\sigma}},$$

where R denotes total revenues in the economy, $q(\alpha) = \tilde{N}x(\alpha)$, since quantity produced is equal to the quantity demanded by each agent times the number of agents, and P denotes the ideal price index, which we normalize to one throughout. The problem of an entrepreneur with draw α is

$$\pi(\alpha) = \max_{q(\alpha) \geq 0} \tilde{\pi}(q(\alpha);w) = \max_{q(\alpha) \geq 0} p(\alpha) q(\alpha) - C(q(\alpha);w) - wf, \tag{7}$$

subject to $p(\alpha) = q(\alpha)^{-\frac{1}{\sigma}} (\alpha R)^{\frac{1}{\sigma}}$ and where f denotes the fixed costs of production. The first-order condition of this problem is

$$((\sigma - 1)/\sigma) q(\alpha)^{-1/\sigma} (\alpha R)^{1/\sigma} - MC(q(\alpha);w) = 0,$$

which implies a standard constant markup over marginal cost given by

$$p(\alpha) = \frac{\sigma}{\sigma - 1} MC(q(\alpha);w). \tag{8}$$

In contrast to standard models, here the marginal cost, and therefore the price, depend on the quantity produced. The quantity supplied is then implicitly determined by

$$q(\alpha) = \alpha R \left(\frac{\sigma}{\sigma - 1} MC(q(\alpha);w) \right)^{-\sigma}. \tag{9}$$

Denote firm revenue by $r(\alpha)$, then

$$r(\alpha) = p(\alpha) q(\alpha) = \alpha R \left(\frac{\sigma}{\sigma - 1} MC(q(\alpha);w) \right)^{1-\sigma}. \tag{10}$$

Firms with different α's charge different prices and supply different quantities of goods to the market. It is easy to show using equation (9) and Propositions 2 and 3 that $q(\alpha)$ is an increasing function of α. Furthermore, when the firm decides to increase the number of layers, the marginal cost falls discontinuously and so $q(\alpha)$ jumps up. The elasticity of quantity with respect to the decline in marginal cost as we change the number of layers is given by $-\sigma$, and the elasticity of the price with respect to the jump in marginal cost is one. In fact, since the price is proportional to marginal costs, it is increasing in α given a number of layers

16

$L + 1$ and jumps down when the firm increases the number of layers. Hence, prices are not monotone in α, which implies that larger firms (firms with higher α and therefore $q(\alpha)$) do not necessarily charge higher or lower prices.

Proposition 8 *The quantity produced, $q(\alpha)$, and revenue, $r(\alpha)$, increase continuously with α for a given number of layers and increase discontinuously when the firm increases the number of layers. The price, $p(\alpha)$, increases continuously with α for a given number of layers and decreases discontinuously when the firm increases the number of layers.*

The previous result is intuitive. Given α, the entrepreneur decides the number of layers of the organization. The pricing rule is to set prices such that marginal revenue is equal to marginal cost. As we increase α, the marginal revenue schedule changes and the entrepreneur re-optimizes, finding in some cases profit maximizing to restructure the organization and increase the number of layers. Consider an α, say $\hat{\alpha}$, such that the entrepreneur is just indifferent between producing with L or with $L + 1$ layers. Note that by Proposition 3, the organization with $L + 1$ layers has a lower marginal cost than the organization with L layers at $\hat{\alpha}$. This implies that the quantity supplied under organization L cannot be the optimal quantity supplied under organization $L + 1$. In particular, since the $L + 1$ organization has a lower marginal cost, the entrepreneur produces a quantity at which the marginal revenue is lower: thus a higher quantity. Therefore, the quantity supplied under organization L is lower than under organization $L + 1$. A parallel logic then implies the corresponding changes in prices and quantities proven in Proposition 8.

The profits of a firm after obtaining a demand realization α (so not including the cost of entry) are then given by,

$$\pi(\alpha) = r(\alpha) - C(q(\alpha); w) - wf = \frac{r(\alpha)}{\sigma} - w\left(\frac{c}{\lambda}\left(1 - e^{\lambda z_L^{L-1}(\alpha)}\right) + cz_L^L(\alpha) + 1 + f\right), \qquad (11)$$

where $\pi(\alpha) = \tilde{\pi}(q(\alpha); w)$ for the function $q(\alpha)$ solves the fixed point in (9). Note that since $C(\alpha; w)$ is not strictly convex due to changes in the number of layers, in order to maximize profits we need to maximize the profits given a number of layers and then compare these local maxima to determine the global maxima of the overall profit function. It is easy to prove that the profit function for each individual layer is concave and so the local maxima of the problem in (7) are well defined and determined by the first-order conditions of the problem. Note also that since we are comparing local optima that change continuously with α, revenues and profits are continuous and increasing in α. We prove these results in Proposition 9.

Proposition 9 *Given L, the profit function $\tilde{\pi}(q(\alpha); w)$ is strictly concave in q. Furthermore, $\pi(\alpha)$ is increasing and continuous in α.*

An example of the profits of a firm as a function of q is plotted in Figure 7 for different values of α. Clearly $\tilde{\pi}(0; w) = -w(1 + f)$ and $\lim_{q \to \infty} \tilde{\pi}(q; w) = -\infty$. Hence, as is clear from the figure, the profit function has a local maximum for each L. As we increase α, the maximum of these local maxima determines the

optimal number of layers. The figure illustrates cases in which the global maximum is achieved with one, two, or three layers. The figure is also useful to illustrate how $q(\alpha)$ jumps up as the firm decides to add organizational layers.
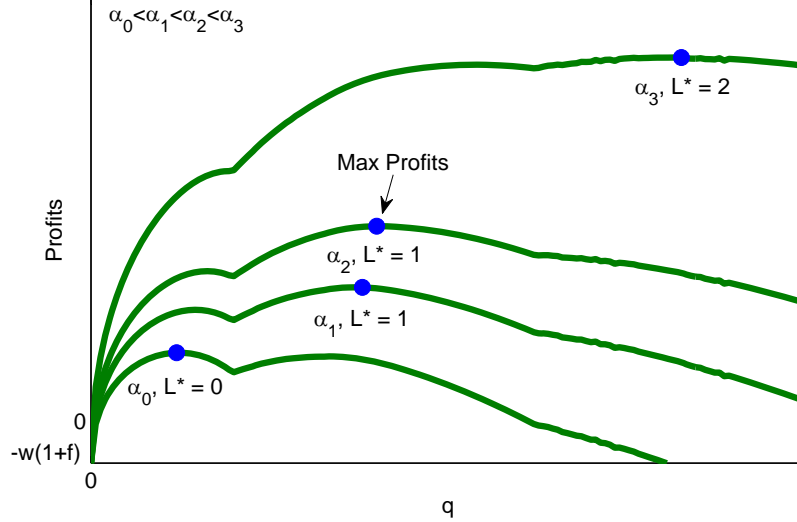

$\alpha_0 < \alpha_1 < \alpha_2 < \alpha_3$

Max Profits

$\alpha_3, L^* = 2$

$\alpha_2, L^* = 1$

$\alpha_1, L^* = 1$

$\alpha_0, L^* = 0$

Profits

0

-w(1+f)

0

q

Figure 7: Profits as a function of α

2.3. Equilibrium in the Closed Economy

Entrepreneurs pay a fixed cost of entry f^E, in units of labor, that allows them to take a draw from the known distributions of α's, $G(\alpha)$. We let $G(\alpha)$ be a Pareto distribution with coefficient γ, so $G(\alpha) = 1 - \alpha^{-\gamma}$. Let M^E denote the total mass of firms that enter in a given period. Firms that enter decide whether to produce or exit. Let M denote the mass of firms operating in equilibrium. Each period a fraction δ of operating firms exits and is replaced by the same number of new entrants.

The value of operating a firm with draw α is then given by

$$v(\alpha) = \max\left\{0, \sum_{t=0}^{\infty}(1-\delta)^t \pi(\alpha)\right\} = \max\left\{0, \frac{\pi(\alpha)}{\delta}\right\}$$

where $\pi(\alpha)$ is given by (11). Note that the fixed cost of production determines a cutoff level $\bar{\alpha}$ such that firms with $\alpha > \bar{\alpha}$ produce. In particular, the marginal firm, $\bar{\alpha}$, makes zero profits. Thus, the zero cutoff profit condition for the marginal firm defines the threshold $\bar{\alpha}$,

$$\pi(\bar{\alpha}) = 0. \tag{12}$$

Of the mass of new entrants, the fraction that operates is $[1 - G(\bar{\alpha})]M^E$. In a stationary equilibrium, the new entrants replace the number of firms that exist, and so

$$[1 - G(\bar{\alpha})]M^E = \delta M,$$

18

where M denotes the total number of active firms. Conditional on entry, the distribution of α is given by $\mu(\alpha) = g(\alpha)/[1 - G(\bar{\alpha})]$ if $\alpha > \bar{\alpha}$ and $\mu(\alpha) = 0$ otherwise, where $g(\cdot)$ is the density of $G(\cdot)$.

Free entry implies that the expected profits are zero accounting for the entry costs, that is

$$\int_0^\infty v(\alpha) g(\alpha) d\alpha = \int_{\bar{\alpha}}^\infty \frac{\pi(\alpha)}{\delta} g(\alpha) d\alpha = w f^E. \tag{13}$$

In equilibrium, labor markets clear. Labor is used for several purposes, as workers and managers in a firm, as teachers to train workers and managers, and in order to design new products and cover the fixed costs of production. Since the total mass of agents in the economy is given by $\tilde{N}$, the labor market clearing condition is given by

$$\tilde{N} = \frac{M}{1 - G(\bar{\alpha})} \left(\delta f^E + \int_{\bar{\alpha}}^\infty (C(\alpha; 1) + f) g(\alpha) d\alpha \right). \tag{14}$$

Finally, equilibrium in the goods market implies that

$$R = w\tilde{N}. \tag{15}$$

The general equilibrium of this economy is characterized by the demand threshold of the firm that obtains zero profits, $\bar{\alpha}$, real wages, w, revenue R, and the mass of firms that operate M. Remember that we have normalized the price index P to one. These four equilibrium variables are obtained by solving the four equations above, namely, (12), (13), (14), and (15). First note that R can be obtained immediately given w from (15). Note further that since the mass of firms, M, does not enter equations (12) and (13), we can solve first for $\bar{\alpha}$ and w, using (12) and (13), and later obtain M from (14). The next proposition shows that there exists a unique equilibrium in this economy.

Proposition 10 *There exists a unique stationary equilibrium.*

The proof of Proposition 10 characterizes the schedules of wages and entry thresholds that satisfy the equilibrium conditions (12) and (13). We show that the schedule implied by the zero cutoff profit condition starts at zero and is strictly increasing. That is, equation (12) implies that, as we increase wages, fewer firms enter, since their profits decrease. In contrast, the schedule defined by the free entry condition implies that, when all firms enter, the wage is positive. Since in the limit where no firm enters wages have to be zero, the curves have to cross.

A stationary equilibrium of this economy exhibits a useful property. In equilibrium an increase in total population from $\tilde{N}$ to $\tilde{N}'$ leads to higher wages given by $w' = \left(\tilde{N}'/\tilde{N} \right)^{\frac{1}{\sigma-1}} w$, but does not change the entry threshold $\bar{\alpha}$ or the quantities produced. Hence, we can easily show that in equilibrium higher population leads to higher prices, revenues and operating profits of active firms. In fact, prices, revenues and profits all change proportionally by the factor $\left(\tilde{N}'/\tilde{N} \right)^{\frac{1}{\sigma-1}}$. Of course, as is standard in this literature, these changes come from the gain in variety associated with the higher population and the increase in the number of firms that the increase in population triggers. The organizational structure enters the result only in that it leads

19

to a cost function that is homogeneous of degree one in wages. However, since quantities produced are not affected by increases in $\tilde{N}$, the organizational structure of firms is unaffected.

Proposition 11 *In equilibrium a larger population size does not affect the entry threshold or the quantities produced (and therefore the organizational structure), but increases the wages, prices, revenues and operating profits of all active firms.*

3. OPEN ECONOMY

Consider now an economy with a domestic (D) and a foreign (F) country, which can trade with each other. Throughout, we assume that the demand draw of a firm applies to both markets. That is, tastes are identical across countries. Entrepreneurs in each country decide how much to supply to the local market and how much to sell abroad. Entrepreneurs in the domestic country pay a fixed cost f_{DD} to produce. If they want to supply the foreign market they also need to pay a fixed cost f_{DF}. Similarly, firms in the foreign country need to pay a fixed cost f_{FF} to produce and f_{FD} to export. Trading goods is costly. Let $\tau_{DF} = \tau_{FD} > 1$ be the 'iceberg' trade cost incurred by firms. These costs entail a physical loss of resources. Finally, $\tilde{N}_i$ for $i = D, F$ denotes the fixed supply of labor in country i. As a result of the fixed cost of production and exporting, there exist thresholds $\bar{\alpha}_{ij}$ such that if $\alpha < \bar{\alpha}_{ij}$, the entrepreneur α from country i does not find it profitable to supply to market j, and vice versa.

Let $x_{ij}(\alpha)$ be the demand of an agent in country j for goods α produced in country i for $i, j \in \{F, D\}$, and $p_{ij}(\alpha)$ denotes its price. Since agents now have access to goods from both countries, the preferences in (1) become

$$U_i\left(x_{ii}\left(\cdot\right), x_{ij}\left(\cdot\right)\right) = \left(M_i \int_{\bar{\alpha}_{ii}}^{\infty} \alpha^{\frac{1}{\sigma}} x_{ii}\left(\alpha\right)^{\frac{\sigma-1}{\sigma}} \mu_{ii}\left(\alpha\right) d\alpha + M_j \int_{\bar{\alpha}_{ji}}^{\infty} \alpha^{\frac{1}{\sigma}} x_{ji}\left(\alpha\right)^{\frac{\sigma-1}{\sigma}} \mu_{ji}\left(\alpha\right) d\alpha\right)^{\frac{\sigma}{\sigma-1}}$$

for $i, j \in \{F, D\}$, where $\bar{\alpha}_{ij}$ is the demand draw of the marginal firm that supplies goods to market j from market i. M_i now denotes the number of operating firms in market i and $\mu_{ji}(\alpha) = g(\alpha)/(1 - G(\bar{\alpha}_{ji}))$ is the conditional distribution of varieties available in market i and produced in country j.

The budget constraint of agents in i is given by

$$M_i \int_{\bar{\alpha}_{ii}}^{\infty} p_{ii}\left(\alpha\right) x_{ii}\left(\alpha\right) \mu_{ii}\left(\alpha\right) d\alpha + M_j \int_{\bar{\alpha}_{ji}}^{\infty} p_{ji}\left(\alpha\right) x_{ji}\left(\alpha\right) \mu_{ji}\left(\alpha\right) d\alpha = w_i.$$

Then, total demand for country j's products in country i is given by

$$\tilde{N}_i x_{ji}\left(\alpha\right) = \alpha \left(\frac{p_{ji}\left(\alpha\right)}{P_i}\right)^{-\sigma} \frac{R_i}{P_i}, \tag{16}$$

for all i, where the price index is given by

$$P_i = \left(M_i \int_{\bar{\alpha}_{ii}}^{\infty} \alpha p_{ii}\left(\alpha\right)^{1-\sigma} \mu_{ii}\left(\alpha\right) d\alpha + M_j \int_{\bar{\alpha}_{ji}}^{\infty} \alpha p_{ji}\left(\alpha\right)^{1-\sigma} \mu_{ji}\left(\alpha\right) d\alpha\right)^{\frac{1}{1-\sigma}}. \tag{17}$$

As before, we normalize the price index of the domestic country to one, so $P_D = 1$.

Consider first the problem of a firm with demand draw α in country i that does not export. It solves,

$$\pi_{ii}(\alpha) \equiv \max_{x_{ii}} \left\{ p_{ii}(\alpha) \tilde{N}_i x_{ii}(\alpha) - C(q_i; w_i) - w_i f_{ii} \right\}$$

subject to (16). For firms that supply only to the domestic market, $q_i = \tilde{N}_i x_{ii}(\alpha)$, and so the first-order condition is given by

$$q_i(\alpha) = \alpha R_i P_i^{\sigma-1} \left(\frac{\sigma}{\sigma-1} MC(q_i(\alpha); w_i) \right)^{-\sigma}, \tag{18}$$

which implies

$$p_{ii}(\alpha) = \frac{\sigma}{\sigma-1} MC(q_i; w_i).$$

Denote by $\pi_{ij}(\alpha)$ the profits of a firm with demand draw α in country i that sells in the domestic and foreign markets. The profit maximization problem becomes

$$\pi_{ij}(\alpha) \equiv \max_{(x_{ii}, x_{ij}) \geq 0} \left\{ p_{ii}(\alpha) \tilde{N}_i x_{ii}(\alpha) + p_{ij}(\alpha) \tilde{N}_j x_{ij}(\alpha) - C(q_i; w_i) - w_i(f_{ii} + f_{ij}) \right\}$$

subject to (16) and where total production is given by

$$q_i(\alpha) = q_{ii}(\alpha) + q_{ij}(\alpha) = \tilde{N}_i x_{ii}(\alpha) + \tau_{ij} \tilde{N}_j x_{ij}(\alpha), \tag{19}$$

where $q_{ij}(\alpha)$ denotes the quantity produced by i for market j. The cost function $C(q_i; w_i)$ solves the cost minimization problem described in Section 2. The first-order conditions of this problem implicitly define the quantities sold in each market,

$$\tilde{N}_i x_{ii}(\alpha) = \alpha R_i P_i^{\sigma-1} \left(\frac{\sigma}{\sigma-1} MC(q_i(\alpha); w_i) \right)^{-\sigma}, \tag{20}$$

and

$$\tilde{N}_j x_{ij}(\alpha) = \alpha R_j P_j^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \tau_{ij} MC(q_i(\alpha); w_i) \right)^{-\sigma}. \tag{21}$$

In contrast with the standard model, $x_{ii}(\alpha)$ and $x_{ij}(\alpha)$ enter the marginal cost function through $q_i(\alpha)$ as well. That is, a firm's level of total production affects its marginal cost and therefore how much it sells in each market. Importantly, the decision to export affects the cost of production of the goods sold in the local market.[14] Hence, as usual, the price in each market is given by a constant markup over marginal cost, namely,

$$p_{ij}(\alpha) = \frac{\sigma}{\sigma-1} \tau_{ij} MC(q_i(\alpha); w_i) = p_{ii}(\alpha) \tau_{ij}.$$

---

[14] This implies that even when $f_{ij} > f_{ii}$ all firms in the economy could enter the exporting market. Of course, if f_{ij} is large enough, then only the most productive firms will export. This is a key distinction with Melitz (2003) where, for the case of two symmetric countries, all firms will export if and only if $f_{ij} \leq f_{ii}$.

3.1. Equilibrium in the Open Economy

As in the closed economy, the zero profit condition for the marginal firm defines the threshold of entering firms in country i, $\bar{\alpha}_{ii}$. So, $\pi_{ii}(\bar{\alpha}_{ii}) \geq 0$. Similarly, indifference between exporting or selling domestically defines the threshold to export $\bar{\alpha}_{ij}$, so $\pi_{ij}(\bar{\alpha}_{ij}) = \max\{0, \pi_{ii}(\bar{\alpha}_{ij})\}$. In parallel to the closed economy, the free entry condition is then given by

$$\int_{\bar{\alpha}_{ii}}^{\bar{\alpha}_{ij}} \frac{\pi_{ii}(\alpha)}{\delta} g(\alpha) d\alpha + \int_{\bar{\alpha}_{ij}}^{\infty} \frac{\pi_{ij}(\alpha)}{\delta} g(\alpha) d\alpha = w_i f_i^E,$$

where f_i^E denotes the fixed cost of entry in market i. Labor market clearing is guaranteed by

$$\tilde{N}_i = \frac{M_i}{1 - G(\bar{\alpha}_{ii})} (\delta f_i^E + \int_{\bar{\alpha}_{ii}}^{\infty} (C(q_i(\alpha);1) + f_{ii}) g(\alpha) d\alpha + \int_{\bar{\alpha}_{ij}}^{\infty} f_{ij} g(\alpha) d\alpha). \tag{22}$$

All these conditions hold for $i \neq j$ and $i,j \in \{D,F\}$. Finally, goods market clearing is guaranteed by $R_i = w_i N_i$ for $i \in \{D,F\}$. It is easy to show that the labor market clearing condition, the free entry condition, and the price index normalization together imply that trade is balanced. An equilibrium is then given by a vector $(\bar{\alpha}_{DD}, \bar{\alpha}_{DF}, \bar{\alpha}_{FF}, \bar{\alpha}_{FD}, M_D, M_F, P_D, P_F, w_D, w_F, R_D, R_F)$, which solves the 8 conditions above together with the two goods market clearing conditions, $P_D = 1$, and P_F determined by (17).

The next proposition guarantees gains from trade as well as an increase in the quantity produced of all exporters and a decline in the quantity produced of all non-exporters. Given that we are normalizing $P_D = 1$, w_D measures welfare in the domestic country. Now consider the effect of an increase in w_i on the quantity produced of non-exporters. We can use equation (18) after substituting $R_i = w_i N_i$ to study the effect of w_i on q_i. Note first that as shown in Proposition 4, $MC(q_i; w_i) = w_i MC(q_i; 1)$. Hence, an increase in wages increases the right-hand side of equation (18) since $\sigma > 1$. If the marginal cost did not depend on q, this would immediately imply that the quantity produced by non-exporters would decrease. In our model there are two other effects. Consider first the case of a firm that does not change the number of layers as a result of the change in wages. In this case the direct effect of w_i is dampened by the fact that the decrease in $q_i(\alpha)$ decreases marginal costs as proven in Proposition 2, which makes firms produce more. However, this new effect can never dominate the first, since it operates through the original effect on $q_i(\alpha)$. If the original effect is strong enough, or the firm is close to switching layers, the original effect may trigger firms to decrease their number of layers. In this second case, the original decline in $q_i(\alpha)$ as a result of the rise in w_i increases discretely the marginal cost of the firm, which reinforces the direct effect and makes the firm sell even less. So the response of quantities produced as a response of changes in wages relative to the standard model is dampened for firms that do not change the number of layers and amplified for firms that do. The same logic applies to the domestic sales of exporting firms, although exporters now have access to a new market. Using Proposition 11, we can prove that access to this new market always increases their sales.

Proposition 12 *In equilibrium a trade liberalization increases welfare in both countries. Furthermore, the*

quantity produced of all non-exporters decreases and the quantity produced of all exporters increases.

Proposition 12 is useful to understand the effect of trade on firm productivity and the internal organization of firms, particularly since in Section 2 we characterized the cost minimization of firms as a function of their quantity produced. The increase in the quantity produced of exporters implies that they increase weakly the number of layers of management. Exporters that do not change the number of layers exhibit a higher marginal cost. These effects are reversed for non-exporters. Productivity is equal to the inverse of average costs at a wage equal to one, as described in Section 2.1.1. Hence, the productivity of exporters can increase or decrease depending on a firm's production relative to the MES of its number of layers. Opening trade to a larger country, which increases sales further tends to raise productivity as it increases layers of management and moves producers to average cost curves with lower minima. Hence, on average, productivity tends to increase for exporters. Non-exporters could also gain or lose in terms of productivity. The smallest firms always become less productive. Others might become more productive, since they get closer to the MES of their chosen L. The average cost curve has a general convex shape, which implies that the productivity gains that result from opening to trade are larger for small exporters than for large exporters (although there are firms that increase and others that decrease their productivity).[15]

Corollary 13 *A trade liberalization increases weakly the number of management layers of all exporters and decreases weakly the number of layers of all non-exporters.*

Corollary 13 is the direct result of Propositions 6 and 12. All exporters produce more after a trade liberalization and firms that produce more minimize their cost when they have weakly more management layers. The converse is true for non-exporters. This corollary and Proposition 12 partially characterize the effect of a trade liberalization. However, since the response of firm productivity and other firm characteristics is heterogeneous, we turn now to a calibrated version of the model where we study a richer set of firm responses. We do this in the following section.

4. CALIBRATION AND COMPARATIVE STATICS

For the quantitative exercises presented below we require values of the following set of parameters: f_i^E, f_{ii}, f_{ij}, h, c/λ, γ, σ, A, $\tilde{N}_i$, δ, τ_{ij}. We specialize the model to a world with two symmetric countries and calibrate the parameters such that the model matches several moments in the U.S. data for 2002. We use the reference values in the literature for the parameters δ, σ, and t_{ij}. We set the elasticity of substitution to 3.8 following Bernard, et al. (2003) and Ghironi and Melitz (2005). For the level of trade costs, τ, we use 1.3 and for the exogenous firm exit rate we use 10%, both values obtained from Ghironi and Melitz (2005) and

---

[15]The heterogeneous response of firm productivity in our model imply that average productivity can increase or decrease depending on the distribution of α's. We show below that for the calibrated economy with a Pareto distribution of α's, the general shape of the average cost function leads to an average productivity gain for exporters and average productivity loss for non-exporters.

used for the U.S. for 2002. We set $\tilde{N}_i$ equal to the total number of employees in the manufacturing sector and the educational sector as described below.

We normalize $f_{ii} = 1.1$ and calibrate the values of $f_i^E, f_{ij}, h, c/\lambda, A$ and γ to match five moments in the U.S. data and the Pareto coefficient of the size distribution of firms producing manufactured goods in the U.S. The moments that we match are the share of expenditure on domestic goods, total expenditure, the fraction of firms that export, the average size of firms, and the share of education employees in the economy.

The fraction of firms that export is not directly reported in the U.S. Census Bureau. Bernard, et al. (2007) estimate that 18% of the U.S. manufacturing firms exported in 2002. We use this figure since we restrict our attention to the manufacturing sector.

In the model some of the labor used in production is used to train workers. In order to take this into account we create a sector that is a bundle of the manufacturing sector and the share of the educational services sector used in the manufacturing sector. To do so, we first calculate the share of education employees using the figures reported by the Career Guide to Industries (CGI) from the Bureau of Labor Statistics (BLS) Current Population Survey for 2008. The CGI reports the number of employees per occupations in different industries.[16] The total fraction amounts to 11.8% of total private employment in 2008. We calculate the share of total gross production from the education sector using the STAN - OECD database for Structural Analysis. We apply these shares to the manufacturing sector measures of employees and gross production. In this way we obtain that $\tilde{N} = 16.48$ million.[17]

The average firm size is calculated as the total number of employees over the total number of firms from the manufacturing sector.[18] The total number of workers from the manufacturing sector reported by the 2002 Statistics of U.S. Businesses (SUSB) from the U.S. Census Bureau is 14,393,609 and the total number of manufacturing firms is 318,126. Therefore, the average size is 45.24. Finally, we use the TRAINS database to obtain data on imports from the manufacturing sector and the STAN database for gross production from the bundled sector. We find that the share of expenditure on domestic goods is 78.9% and total expenditure is equal to $5.1 billion.

We search over the parameter space for parameter values that match the discussed moments, using as a loss function the norm of the percentage deviation difference between the model and the data. Table 1 presents the moment conditions that we match and the values they take in our preferred calibration.

The calibrated parameter values are presented in Table 2. The cost of knowledge acquisition, c/λ, and communication technology, h, satisfies Assumption 1. Hence, in the calibrated exercises there are no firm with zero knowledge at any layer. In the previous section we suggested an alternative interpretation for the per period cost of knowledge c; it can be reinterpreted as $\tilde{c}\delta_M$ where $\tilde{c}$ is the cost of training workers in an

---

[16]We calculate the share of education employees by the number of workers in the Educational Services sector over total private employment in the year 2008.

[17]We apply the share of education employees from 2008 to the total number of employees in the manufacturing sector in 2002 since the figure for 2002 is not available on the CGI website.

[18]Note that we do not include the education workers as part of the manufacturing sector in order to calculate the average size of firms. We also do not include them in the moment calculated in the model.

economy where only new workers have to be trained every period and δ_M is the the annual death rate. The U.S. death rate is 0.008% and so the cost of training new workers takes a value of 28, implying a stock of human capital equivalent to 4 times the population of the economy.[19]

Table 1: Calibrated moments

Moments	Data	Model
Share of firms that export	18.0	17.53
Average size of firms	45.2	45.44
Share of education employees	11.8	11.85
Share of expenditure on domestic goods	78.9	74.94
Total expenditure	5.1	5.10
Size distribution of firms, Pareto coefficient	-1.095	-1.094

We now describe our results. Figure 8 presents productivity as a function of the idiosyncratic demand parameter α for the case of the calibrated open economy and an economy in autarky with identical fundamental parameter values. Throughout, all light curves refer to the implications of the model in autarky and dark curves refer to the calibrated open economy. The vertical lines in the graph mark the thresholds $\bar{\alpha} < \bar{\alpha}_{DD} = \bar{\alpha}_{FF} < \bar{\alpha}_{DF} = \bar{\alpha}_{DF}$. Namely, they mark the entry threshold in autarky, the entry threshold in the two trading economies, and the export threshold in both economies. The figure shows productivity, measured as $q\left(\alpha\right)/C\left(\alpha;1\right)$, and illustrates how the productivity of firms changes as we move to the open economy. As in any model with fixed exporting costs, only the firms with high enough α's export, and some firms exit as we open the economy.

Table 2: Parameter values

Parameters	A	f^E	f_{ij}	γ	c/λ	h
Values	0.26	35.1	5.4	0.9	0.225	0.26

Among the firms that enter in both scenarios, the ones with the lowest demand realizations experience a reduction in productivity as a result of the trade liberalization. They sell less, at higher prices, and obtain less profits. Note the difference with the standard model in which trade liberalization has no effect on productivity and productivity is identical for all firms (given our setup with exogenous heterogeneity in demand, not productivity). As α increases from its lowest values, productivity increases in both scenarios, since firms are still producing below the MES of a firm with two layers (in this calibration all firms have at least two layers). Firms in autarky reach the MES of a two-layer organization for a lower α than in the open economy (as implied by Corollary 13). Hence, as α increases further, some non-exporting firms experience a gain in productivity.

---

[19]The Pareto coefficient that matched the size distribution of firms has a value lower than 1; therefore, the mean and the variance of the distribution are not well-defined. Nevertheless, since the support of the distribution used in the computations is finite, all the moments we compute are well defined.

The relationship is reversed once the firms change to an organization with three layers. At this point the discontinuous increase in $q(\alpha)$ leads to a discontinuous change in the cost, which results in a discontinuity in productivity (recall Proposition 8). As we move beyond the export threshold $\bar{\alpha}_{DF}$, the productivity of exporting firms jumps up discretely in the open economy. The jump implies an increase in productivity for the marginal firm of 1%. Firms with α's just before the threshold experience a loss in productivity as a result of trade and firms with α's just after the threshold experience a gain. Since exporting firms switch layers for smaller α's than in autarky, there are some large exporting firms that lose productivity as a result of trade. Our parameterization of the model implies that, as firms get larger and have more layers, the gains or losses in productivity are progressively smaller. Productivity in both scenarios converges to a common constant.[20]
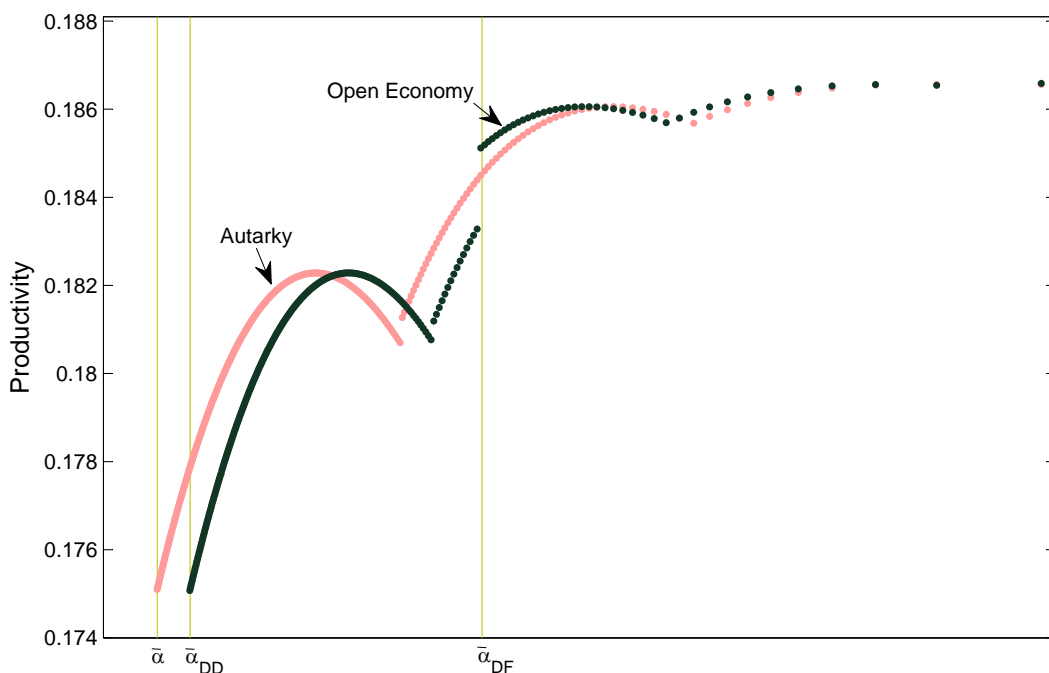

Figure 8: Productivity for all α in autarky and open economy

Figure 8 illustrates clearly the heterogeneous responses in productivity as a result of a trade liberalization from autarky to the calibrated economy. In general, the effects on productivity are below 1%. However, note that these changes in productivity would not exist if we were to move from autarky to an economy without any fixed trade costs (see Figure 15 below). In that case, opening to trade would amount to an increase in population that raises wages, and the number of firms. In contrast, it leaves the quantities produced, and therefore the organizational structure and productivity, unchanged (see Proposition 11). Hence, the productivity implications of our model are the result of the reorganization of production given the change in $q(\alpha)$ generated by the presence of fixed and variable trade costs.

---

[20]These features are a consequence of agents learning the most routine problems first, namely, a distribution of problems with decreasing density.

Figure 9 presents four plots with the main firm outcomes. The upper-left plot presents the average cost function in both scenarios. For the range of α's presented, firms have between two and four layers and the marginal firms that export have three layers. Note how the average cost function becomes flatter as we increase α, which implies that the gains from organizational change beyond four layers are minimal. The main difference between the average cost functions in both scenarios is their level. This change is the result of an increase of 8.1% in wages in the open economy relative to autarky, which increases average costs. The lower-left panel presents firm's profits, which, as expected, are monotonically increasing in α. Profits decrease for all non-exporters and some of the marginal exporters, and increase for the large exporters. These effects are standard in the literature and remain in our setup. The upper-right figure presents production. Non-exporters produce more in autarky than in the open economy and the difference is larger for firms that change layers. Exporters produce more overall, again with the difference between trade and autarky larger for firms in the open economy that change layers.
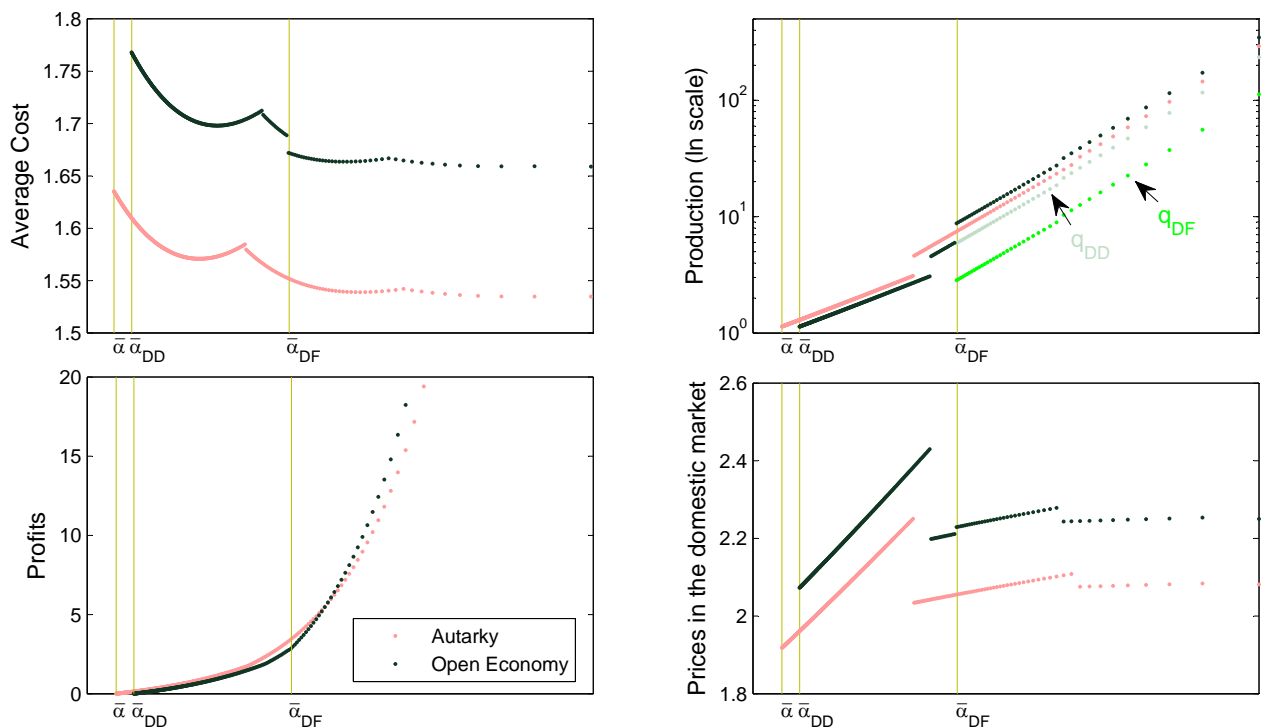
Figure 9. Average cost, production, profits and prices

The upper-right panel of Figure 9 also presents production for the export and domestic markets. Exporters produce less for the domestic market than in autarky. Note the increase in production as a result of trade. This is the main effect that leads to the productivity gains that we have underscored here. The bottom-right figure presents the domestic prices charged by all firms (foreign prices are just multiplied by τ_{DF}). Propositions 2 and 3 describe the behavior of the marginal cost as a function of $q(\alpha)$ and, as we discussed

27

in Section 3, our CES demand system leads to prices that are proportional to marginal costs. Since $q(\alpha)$ is increasing in α as we saw in the upper-right panel, we know that prices increase with α. Note that in the standard model domestic prices would be constant in α. As firms change layers, prices fall discontinuously as described in Proposition 8. The price function also becomes flatter as we increase the number of layers and eventually converges to a constant as we let $\alpha \rightarrow \infty$. Note also the jump in prices for the marginal exporting firm in the open economy. This is the result of the jump in quantity, which increases its marginal cost and therefore its price. We can also calculate revenue productivity by multiplying our productivity measure by the relevant prices. Of course, revenue productivity will also reflect the increase in prices that results from the increase in wages. The jump in prices for the marginal exporting firm together with our analysis of Figure 9 implies that revenue productivity in this economy increases by more than productivity. The overall increase in revenue productivity for the marginal firm that exports is equal to 1.82%.

Figure 10 shows the impact of moving from autarky to the calibrated economy for the distribution of firm sizes, knowledge, income, and productivities. The upper-left panel presents the distribution of firm's employment size and how it fits the data. We plot the size distribution using a standard log-log plot of the cumulative distribution in which a Pareto distribution would appear as a straight line with slope of minus its shape coefficient. As is evident from the figure, the model does a nice job in fitting the distribution of firm sizes. This is, of course, partly the result of our assumption that the idiosyncratic demand shocks are distributed Pareto and the fact that we are calibrating the shape parameter of this distribution to fit the data. However, the fit is by no means guaranteed in a model as rich as this. It is clear from the figure that in autarky firm sizes are smaller and the tail of the distribution is thinner, indicating fewer large firms.

The upper-right panel of Figure 10 presents a histogram of the distribution of knowledge in the economy in our two scenarios. The figure indicates that there are substantially more workers with low levels of knowledge in the open economy, although there are more agents with intermediate levels of knowledge. The effect is the result of the increased number of layers of exporting firms in the open economy. These firms introduce more layers and, as a result, require their workers to know less. However, since they hire more intermediate managers, there is more mass of agents with intermediate levels of knowledge than in autarky. The agents with the highest levels of knowledge are all entrepreneurs and are present in both economies. The lower-left panel of Figure 10 presents the distribution of income. The distribution of income follows closely the distribution of knowledge in the economy. That is, income is given by $w_i \left(c z_L^\ell + 1 \right)$. We know that w_i increases as we open the economy, which shifts the distribution of income in the open economy to the right relative to the closed economy.

Finally, the lower-right panel presents the distribution of productivities in the economy. The figure shows that the fraction of small productivity firms in both economies remains largely unchanged. Recall from Figure 8 that the productivity of the smallest firms decreases, but the entry threshold also shifts to the right. Hence, since the shape of the Pareto distribution of α's is invariant to truncation, we obtain similar shares of small productivity firms in both economies. The main difference between this conclusion and the one we

reached discussing Figure 8 is that the latter conditions on α while the former does not. The figure also illustrates the increase in the mass of the most productive firms as we open to trade. Essentially, firms in the middle of the distribution increase their productivity, thereby raising the mass at the tail. This is the main effect of a trade liberalization: it makes firms with intermediate levels of demand more productive than in autarky because it incentivizes a reorganization of the firm.
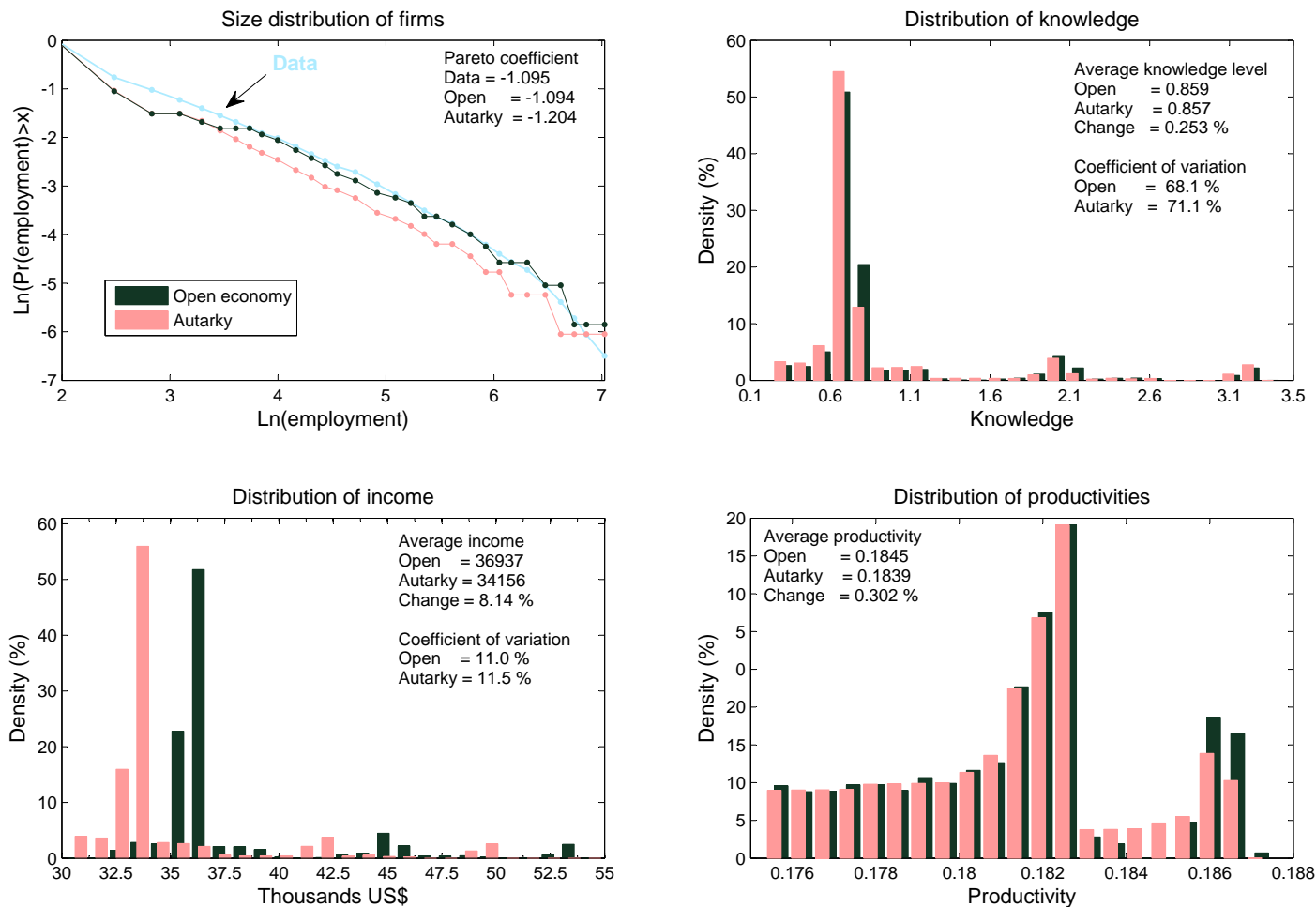
Figure 10: Firm size, knowledge, income, and productivity distributions

4.1. Effects of Trade on Internal Organization

So far our discussion has been limited to firms' and economy-wide changes in outcomes as we open to trade. We now turn to describe the internal reorganization of firms. Figure 11 presents examples of the reorganization of a non-exporter and an exporter as the economy opens to trade. We select firms that change their choice of layers as a result of this reorganization. Of course, many firms do not change their number of layers as the economy opens to trade and so behave differently than what we illustrate in the figure.

Nevertheless, even the firms that do not change layers change the knowledge, composition and number of their employees as a result of the change in the quantity produced triggered by the trade liberalization. Each layer is represented by a rectangle where the height represents the knowledge of workers in that layer, and the length represents the number of agents working in that layer. We plot rectangles in higher layers above rectangles in lower ones.
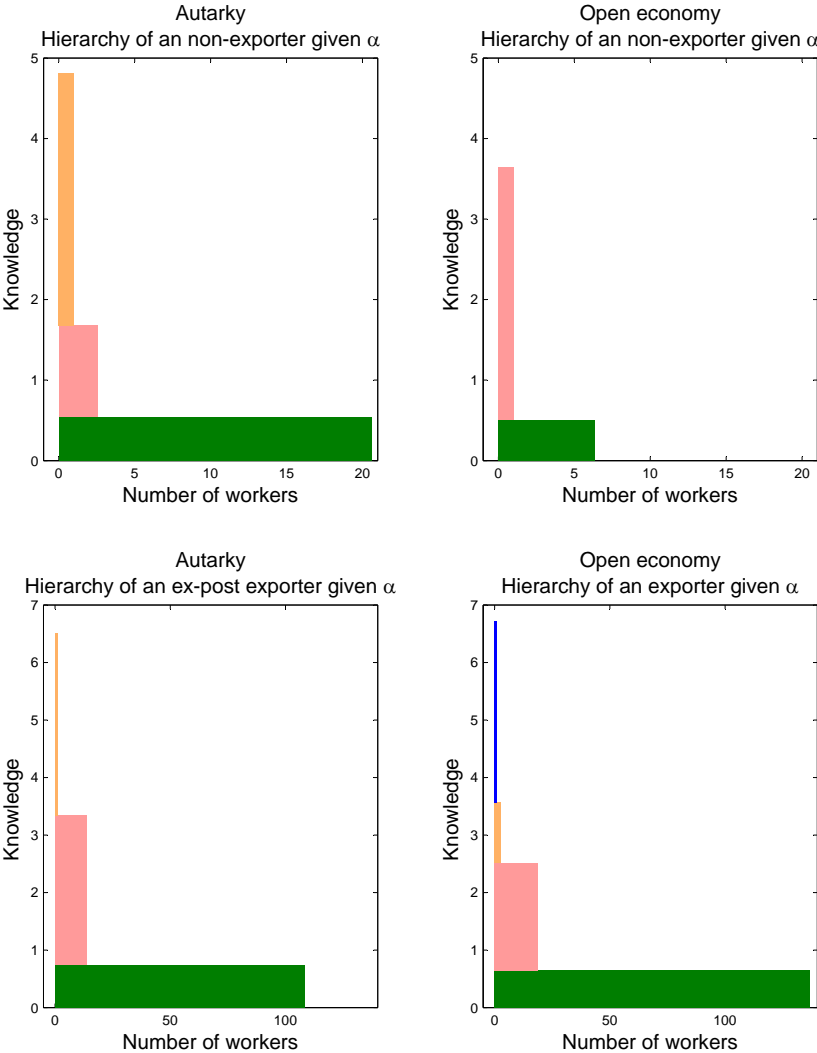
Figure 11: Changes in the organizational design after a trade liberalization

Compare first the two upper panels in the figure, which represent a firm with given demand level, α, that does not export as a result of the trade liberalization. The change in the characteristics of the firm is dramatic. The firm reduces the number of layers from three to two. This implies a reduction in the number of workers (agents in layer zero) from 20 to 6. Note that the main reason to have an intermediate layer is to increase the span of control of the entrepreneur. Since in autarky the firm also hired two agents in the

30

second layer, the total employment size of the firm went from 23 to 7. The levels of knowledge also changed. Workers in the open economy know less and, of course, the firm as a whole solves fewer of the problems faced in production, since the cumulative knowledge of the firm is 25% smaller.

The implications change when we look at the exporter represented in the two lower panels. Now, the number of layers increases as we open to trade, and so does the number of employees in the firm. However, the amount of knowledge of agents in similar layers decreases. Still, since the firm has one more layer, its cumulative knowledge increases. In this case, the main role of adding a layer is to increase the total span of control of the entrepreneur (that is, the total number of employees). However, the firm decides to decrease the knowledge of agents at all comparable layers, since otherwise the increase in output would exceed its optimal size. This decline in the knowledge of the workers of firms in the open economy is also evident in the distribution of knowledge presented in Figure 10.

Figure 11 presents an example of changes in organization but does not allow us to describe these changes across the universe of firms. To do so, we introduce the concept of autonomy as used in Bloom, et al. (2010). In a model of management by exception, like ours, autonomy of a class of workers can be measured by the share of the decisions they make relative to the total set of decisions made by the firm. That is, we can measure the set of problems solved by agents in a given layer, relative to the total amount of knowledge in the firm. So z_L^ℓ/Z_L^L measures the autonomy of agents in layer ℓ. Hence, z_L^L/Z_L^L measures the share of knowledge of the entrepreneur, or its autonomy. Of course, the larger the autonomy of the entrepreneur, the more centralized the decision making in the firm.

Figure 12 presents the measures of autonomy for agents working in different layers as a function of the demand draw of the firm, α. Focus first on the lower-right panel that presents the autonomy of entrepreneurs, or the level of centralized decision making at the firm. The figure shows a clear pattern. The level of centralized decision making is decreasing in α and it decreases discontinuously as firms add more layers. Furthermore, a trade liberalization increases the level of centralization for all non-exporters and decreases it for all exporters. So the model delivers a clear testable implication for this key dimension of the organization of the firm. Bloom, et al. (2010) present evidence that decentralization tends to fall with firm size, as predicted by our theory. As far as we know, the empirical literature has not studied the effect of a trade liberalization on centralization.

The other panels in the figure present the autonomy of workers, first-layer managers (denoted as middle managers) and second-layer managers (denoted as upper-middle managers). The upper-left panel in Figure 12 shows the autonomy of workers. Firms with low α's have only two layers (note how in the lower-left panel there are no upper-middle managers for low levels of α's). For this group, as α increases, the autonomy of workers grows, since this is the only way to raise production. Trade decreases autonomy, since these firms do not export and so the competition from foreign firms reduces their quantities. As we move to firms with three layers, the autonomy of workers remains almost constant and trade has almost no effect. Eventually, once we get to firms with four layers, the autonomy of workers decreases, since the added organizational

layers allow the firm to economize on knowledge by lowering the autonomy of workers.
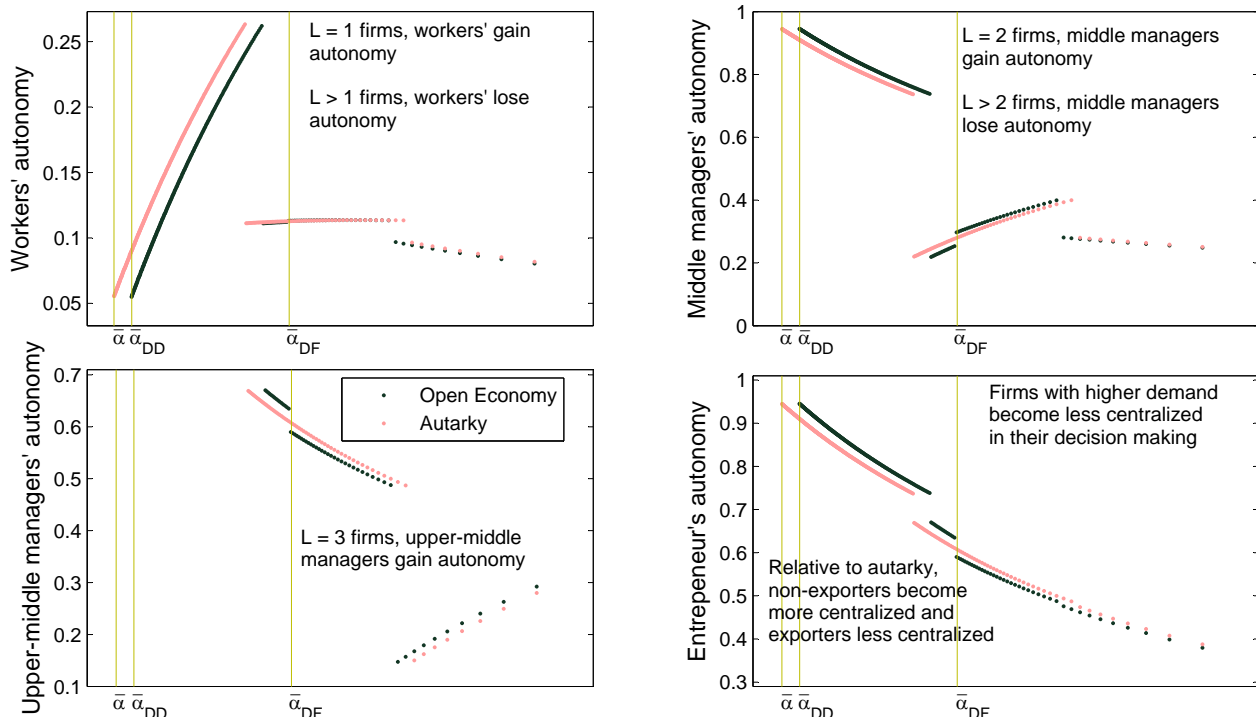

Figure 12: Changes in autonomy before and after a trade liberalization

Now focus on the autonomy of first-layer managers in the upper-left panel. For the firms with the lowest demand draws, these middle managers are also the entrepreneurs, since firms have only two layers. As we argued before, the autonomy of the entrepreneur decreases with α and increases with a trade liberalization. For firms with three layers, the autonomy of first-layer managers increases with α, and with the trade liberalization, since increasing the knowledge of middle managers is the main way in which the firm expands production and leverages the knowledge of the entrepreneur. This is also apparent in the lower-left panel in which now the autonomy of the upper-middle managers is increasing in α for firms with four layers, but not for the ones with three layers, where the upper-middle manager is the entrepreneur. Marin and Verdier (2008) present evidence that the degree of concentration as a function of sales per employee exhibits an inverted U-shape pattern in the data for German and Austrian firms. These patterns imply that this is the case in our theory as well.

4.2. Other Measures of Productivity

We discussed above the implications of our model for the effect of bilateral trade on productivity. In the discussion we used a measure of productivity given by $q(\alpha)/C(\alpha;1)$. Even though this measure is theoretically appealing, in many empirical studies, data on quantities and average costs at constant prices

are not readily available. Hence, we now discuss the predictions of our model for other commonly used measures of productivity. In particular, we study how trade affects revenue productivity (measured as $r(\alpha)/C(\alpha;1) = p(\alpha)q(\alpha)/C(\alpha;1)$), labor productivity (measured as $q(\alpha)/n(\alpha)$ where $n(\alpha)$ denotes the total number of employees in the firm without including education workers or workers used to cover fixed costs), and revenue labor productivity (measured as $r(\alpha)/n(\alpha)$). These measures use progressively more easily available data. Figure 13 presents all of these measures as a function of the demand draw, α.
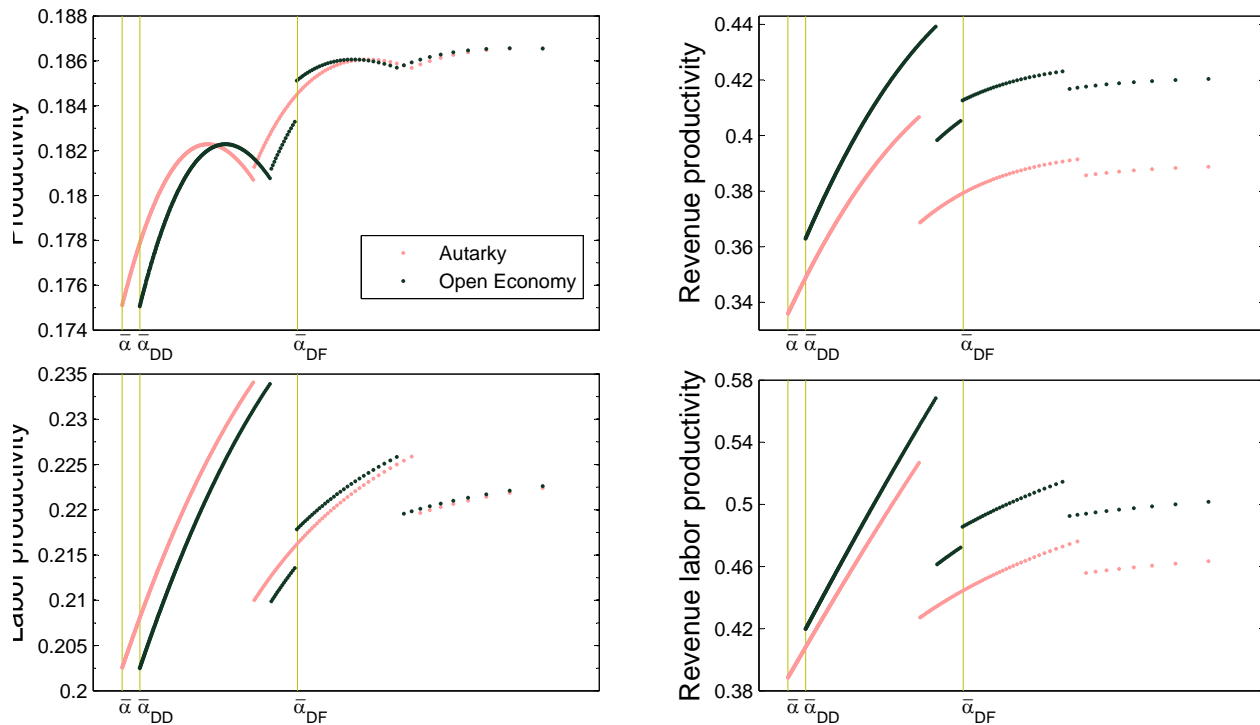
Figure 13: Different measures of productivity

Note first a set of common characteristics of all plots in Figure 13. All measures of productivity increase discontinuously for firms that export. So the marginal exporter always gains in productivity, independently of the measure we use. Of course, the jump in revenue productivity or revenue labor productivity for the marginal firm is larger compared to the case when we use only quantities, Since, as we have shown, the prices of the marginal firm increase discontinuously. The effect of prices also implies that using revenue productivity leads to overall much larger changes in productivity as a result of international trade. In fact, all firms, including the non-exporters, increase their revenue productivity and their revenue labor productivity. Remember that the increase in firms' prices is the result of the increase in wages that results from the extra varieties available to consumers in the open economy. Hence, using revenue and not quantity-based productivity measures overestimates the gains in productivity that result from international trade. This is consistent with the findings in De Loecker (2011).

Table 3: Productivity gains relative to autarky

Weight	Productivity			Revenue productivity		
	$n(\alpha)$	$r(\alpha)$	$q(\alpha)$	$n(\alpha)$	$r(\alpha)$	$q(\alpha)$
All firms	0.302%	0.266%	0.222%	8.627%	8.522%	8.473%
Exporters	0.041%	0.050%	0.051%	8.218%	8.215%	8.216%
Non-exporters	-0.178%	-0.201%	-0.207%	7.868%	7.919%	7.895%
Marginal firm		1.002%			1.822%	

Weight	Labor productivity			Revenue labor productivity		
	$n(\alpha)$	$r(\alpha)$	$q(\alpha)$	$n(\alpha)$	$r(\alpha)$	$q(\alpha)$
All firms	0.348%	0.298%	0.281%	8.654%	8.545%	8.528%
Exporters	0.130%	0.123%	0.125%	8.304%	8.293%	8.295%
Non-exporters	0.022%	0.094%	0.078%	8.096%	8.250%	8.215%
Marginal firm		2.003%			2.831%	

Perhaps the more commonly used measure of productivity in the empirical literature that uses quantity-based measures of productivity is what we call labor productivity. The reason is that it is hard to properly adjust empirically for the knowledge of workers within the firm. This measure yields a very clear empirical implication for the changes in productivity. All non-exporters with the same number of layers should decrease their labor productivity and all exporters with the same number of layers should increase their labor productivity as a result of a trade liberalization. Furthermore, firms with smaller demand draws experience larger changes (positive or negative) in labor productivity than firms with high α's. The results are reversed for firms that experience a change in layers. Non-exporters that decrease their layers gain labor productivity and all exporters that increase their number of layers lose productivity. Table 3 presents the average changes in all of these measures of productivity for all firms, exporters, non-exporters, and the marginal exporter using the number of employees, revenue, or quantities as weights to calculate the average. All results are consistent across weights. The table shows that the increase in labor productivity for the marginal exporter is about 2%.

4.3. Trade Cost, the Costs of Knowledge, and Information Technology

We now consider alternative trade policy experiments and the effect of changes in learning costs and information technology. The first comparative static we analyze is the case where trade costs go to zero, so $t_{ij} = 1$. That is, let us consider the changes that result from moving from autarky to free trade in terms of variable costs. We keep the fixed cost of exporting, since, otherwise, the internal organization of firms does

not change and only wages and the number of firms adjust. Figure 14 presents results that look similar to the transition between autarky and the calibrated economy in Figure 10, but the effects are in general larger.
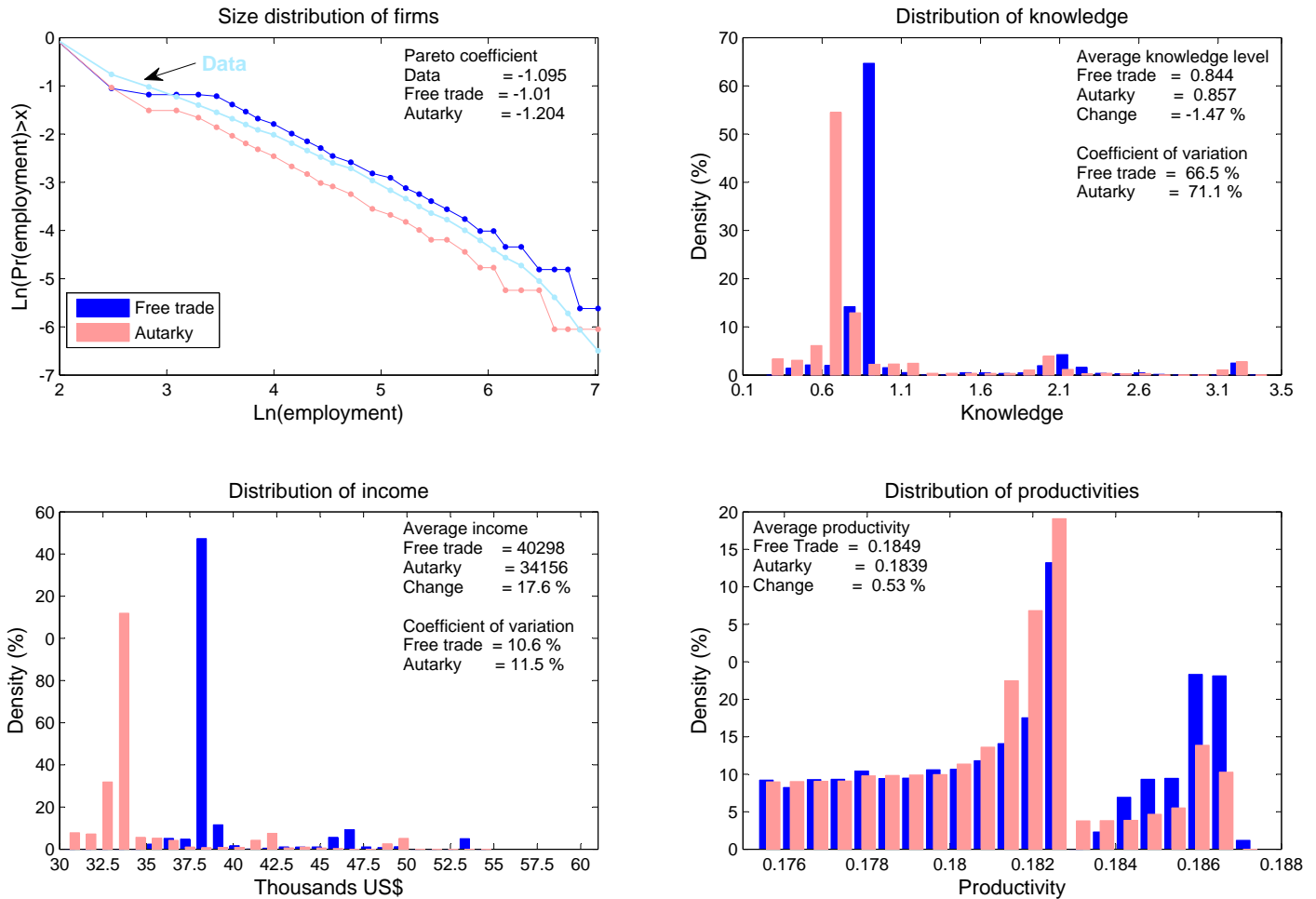
Figure 14: Firm size, knowledge, income, and productivity distributions with free trade

The average increase in productivity doubles. The size distribution has a fatter right tail, and now real wages increase by 17.6%. The larger effects from this transition also allow us to uncover some new effects. Namely, the average amount of knowledge decreases as we move from autarky to free trade. The reason is that exporting firms increase their number of layers and reduce the knowledge of workers in the lower layers. Since there are more of these workers, average knowledge acquisition falls. Furthermore, income and knowledge inequality decline as we open the economy. This was also the case when we moved from autarky to the calibrated economy, but in this case, the effect is stronger and much more significant. The reason is again clear: the top managers of the best and largest firms learn more in autarky, since they use fewer middle managers. Furthermore, with trade, the large spans of control that organizations with many layers permit lead to a large concentration of workers with intermediate skills.

We can also use numerical simulations to understand the effect of changes in the fixed cost of exporting on productivity and welfare. From Proposition 11 we know that if the fixed costs of exporting are small enough so that all firms that produce also export, in a world with two symmetric countries, firms produce identical quantities with and without trade. Thus, only the number of firms and wages adjust. So for low enough transport costs, each firm has a productivity identical to its productivity in autarky. When $f_{ij} = 0$, it is also the case that exporting does not increase the fixed cost of production, and so the production threshold $\bar{\alpha}$ does not change. Hence, when $f_{ij} = 0$ the productivity gain from trade liberalization is equal to zero. Similarly, as $f_{ij} \to \infty$, no firm exports and so the economy is in autarky. Hence, productivity gains are zero as well. For the intermediate fixed cost of exporting, the productivity gains are positive.
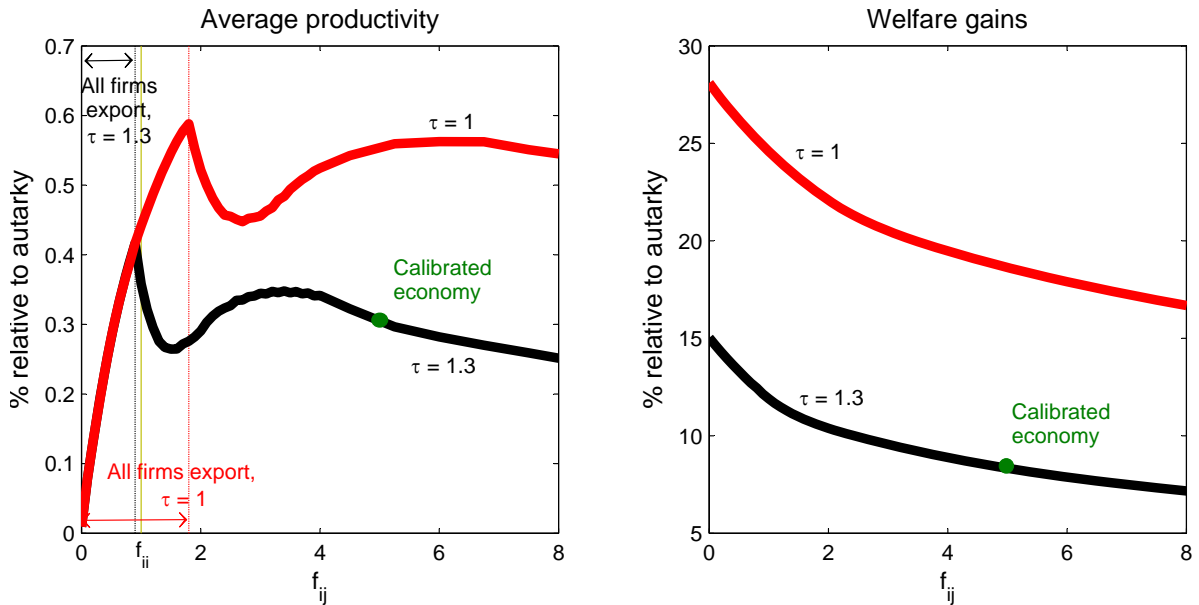
Figure 15: Welfare and productivity gains from trade as a function of f_{ij}

Figure 15 presents average productivity gains and welfare gains relative to autarky for a variety of exporting costs and for two levels of iceberg costs, τ. As can be seen in the figure, there is a range of fixed costs for which all firms export. Note that, in contrast to Melitz (2003) with two symmetric countries, in our theory even if $f_{ij} > f_{ii}$ all firms in the economy might export. This is never the case in Melitz (2003), since there the profits from foreign sales (which are always smaller than the profits from domestic sales) cannot compensate for the fixed cost of the marginal producers. In our theory the profits from domestic and foreign sales are not separable, since it is the total quantity produced that determines the organizational structure (as can be seen in Equations (20) and (21)). So exporting increases the profits from domestic sales by changing the organizational structure of the firm. This can be appreciated in Figure 15 where for the case $\tau = 1$, the largest f_{ij} for which all firms export is about 80% larger than f_{ii}. Of course, as we increase τ, the sales in the export market decrease and the ordering of the threshold is reversed, although this effect is still present.

As we argued above, as long as all firms export, firms' quantities do not change, and therefore, the productivity of all firms remains constant. However, average productivity increases as a result of selection. The firms with the lowest demand thresholds exit, which increases average productivity. Average productivity increases due to selection until some firms decide not to export. At that point further increases in fixed cost decrease average productivity, since the productivity of non-exporters tends to decrease relative to autarky. Increasing f_{ij} further leads to fluctuations in average productivity as the production threshold moves to the left and the export threshold to the right, changing their position relative to the levels of α at which firms change layers. In any case, as f_{ij} increases further the productivity gains tend to zero. In contrast to the effect on average productivity, the welfare gains from trade decrease monotonically with f_{ij}, since the technology to export becomes progressively worse. Furthermore, a reduction in τ increases welfare for any finite level of fixed exporting costs.
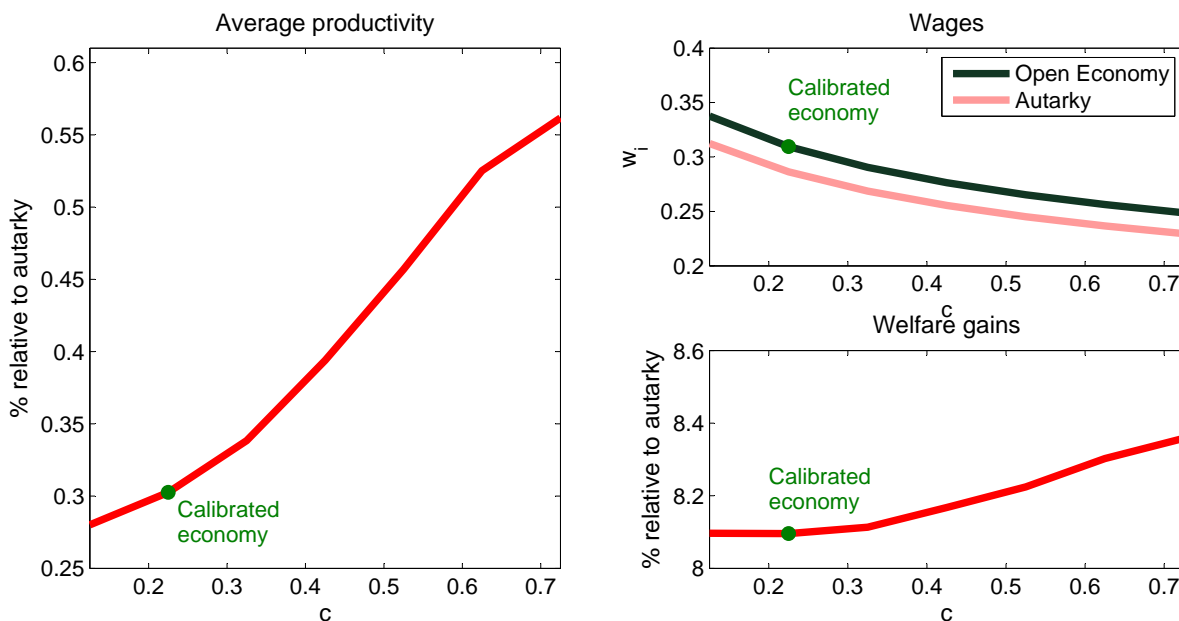
Figure 16: Comparative statics in c

We now turn to the impact of changes in the cost of knowledge, c. As we increase c leveraging knowledge using a complex organization becomes more important. Thus, all firms tend to have taller organizations with more layers, where decision making is more concentrated. Hence, the increase in the quantity produced of exporters associated with international trade leads to more important changes in organization when c is relatively large. In contrast, the organizational structure of non-exporters, which is already relatively simple, does not change much. The result is a monotonically increasing relationship between average productivity and c. Of course, as c increases, the technology to produce knowledge deteriorates and welfare (given by wages since $P_D = 1$) decreases both in autarky and the open economy. However, the decrease in welfare in

autarky dominates, leading to larger gains from trade the higher are learning costs. Organization is more important if c is large, since its sole purpose is to economize on knowledge acquisition. The higher c, the more important is organization, and the increase in scale triggered by trade for exporters. The results of this exercise are presented in Figure 16.
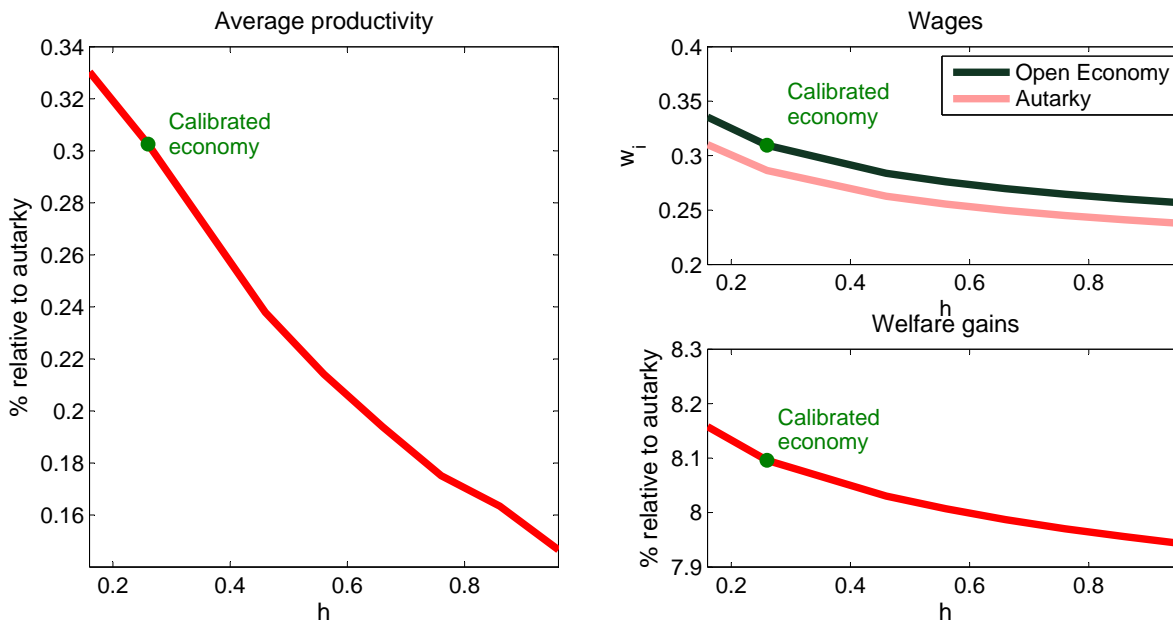
Figure 17: Comparative statics in h

We finish this section with a discussion of the effect of communication technology on productivity, welfare, and the gains from trade, presented in Figure 17. A higher parameter h decreases the span of control of managers. Hence, creating new layers in an organization becomes less attractive. The knowledge of managers can be leveraged less, since the spans of control are smaller given that communication costs are larger. The economy uses less organization and the productivity gains associated with trade become less important. The increase in the quantities produced of exporters changes the organization of exporters by less, leading to smaller average productivity gains in the economy. Since a higher h implies a worse communication technology, it decreases welfare both under autarky and in the open economy. However, the decline for the open economy is larger and so the gains from trade decrease with h. An economy with worse communication technology makes the improvements in the organization of exporters less valuable.

5. CONCLUSIONS

In this paper we propose a theory of an economy in which changes in quantities produced generate changes in the internal organization of firms and therefore firms' productivity, prices, number and knowledge of employees and managers, and the centralization of decisions. The starting point of our theory is the

idea that the level of demand experienced by the firm can affect its decisions on how much to invest to expand its scale, and that many of these investments change discretely. That is, firms decide the integer number of plants they want to use or the integer number of layers they want to have in their organization. These discrete costly investments (in our case adding management layers) lead to heterogeneous productivity responses of firms depending on their level of demand. We implement this idea using an explicit model of a knowledge hierarchy embedded in a general equilibrium setup that resembles Melitz (2003), but where firms are heterogeneous in terms of their level of demand, not their productivity. So, here, heterogeneity in demand leads to heterogeneity in productivity and other characteristics of the firm. As far as we know, this is the first general equilibrium framework with heterogeneous firms and products to incorporate a rich organizational structure. The model can be calibrated or structurally estimated to account for these responses in the data.

With the model in hand, we analyze the effect of a trade liberalization on firm organization and productivity. A trade liberalization is a clear example of an increase in demand for exporting firms combined with more competition from foreign firms. So a trade liberalization is the ideal context in which to use our general equilibrium model of organization to understand the internal reorganization of firms. The empirical trade literature has focused on measuring the increases in productivity that result from trade and has found positive but relatively small changes in productivity once one controls for prices, as in De Locker (2010). A calibrated version of our theory yields similar results. However, we emphasize the heterogeneous responses in productivity (both positive and negative) across producers. The calibrated version of the model yields an increase in productivity of the marginal firm that exports of about 1%. Furthermore, a trade liberalization increases the layers of management of exporting firms and decreases the number of management layers of non-exporters as found by Guadalupe and Wulf (2010). The model yields several other empirical implications that are yet to be contrasted with the appropriate data.

In sum, we believe that our work can be informative for the empirical literature about the mechanisms that can generate changes in firm outcomes as a response to demand shocks. Furthermore, we believe that this theory can help explain the cross-sectional variation in firm productivities as well as other firm outcomes.

REFERENCES

[1] Antràs, P., L. Garicano and E. Rossi-Hansberg, (2006). "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics* 121:1, 31-77.

[2] Antràs, P., L. Garicano and E. Rossi-Hansberg, (2008). "Organizing Offshoring: Middle Managers and Communication Costs," Helpman, E., D. Marin, and T. Verdier, *The Organization of Firms in a Global Economy*, Harvard University Press.

[3] Atkeson, A. and A. Burstein, (2010). "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118:3, 433-484.

[4] Aw, B. Y., S. Chung, and M. J. Roberts, (2000). "Productivity and Turnover in the Export Market: Micro Evidence from Taiwan (China) and the Republic of Korea," *World Bank Economic Review*, 14:1, 65-90.

[5] Aw, B. Y., M. J. Roberts, and T. Winston, (2007). "Export Market Participation, Investments in R&D and Worker Training, and the Evolution of Firm Productivity," *World Economy*, 14:1, 83-104.

[6] Aw, B. Y., M. J. Roberts, and D. Y. Xu, (2008). "R&D Investments, Exporting, and the Evolution of Firm Productivity," *American Economic Review Papers and Proceedings*, 98, 451-456.

[7] Baldwin, J. R. and W. Gu, (2003). "Participation in Export Markets and Productivity Performance in Canadian Manufacturing," *Canadian Journal of Economics*, 36:3, 634-657.

[8] Bernard, A. B. and J. B. Jensen, (1999). "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 47:1, 1-25.

[9] Bernard, A. B. and J. B. Jensen, (2004). "Exporting and Productivity in the U.S.," *Oxford Review of Economic Policy*, 20:3, 343-357.

[10] Bernard, A. B., J. Eaton, B. Jensen and S. Kortum, (2003). "Plants and Productivity in International Trade," *American Economic Review* 93, 1268-1290.

[11] Bernard, A. B, J. B. Jensen, S. Redding and P. Schott, (2007). "Firms in International Trade," *Journal of Economic Perspectives*, 21:3, 105-130.

[12] Bernard, A. B, S. Redding and P. Schott, (2010). "Multiple-Product Firms and Product Switching," *American Economic Review*, 100:1, 70-97.

[13] Bloom, N., R. Sadun, L. Garicano and J. van Reenen, (2010). "The Distinct Effects of Information Technology and Communication Technology on Firm Organization," mimeo, London School of Economics.

[14] Bustos, P., (2011) "Rising Wage Inequality in the Argentinean Manufacturing Sector: The Impact of Trade and Foreign Investment on Technology and Skill Upgrading," forthcoming in *American Economic Review*.

[15] Clerides, S., S. Lach, J. Tybout, (1998). "Is Learning-by-Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico and Morocco," *Quarterly Journal of Economics*, 113:3, 903-947.

[16] Costantini, J. and M. Melitz, (2008). "The Dynamics of Firm-Level Adjustment to Trade Liberalization," in *The Organization of Firms in a Global Economy*, E. Helpman, D. Marin and T. Verdier (eds), Harvard University Press, Cambridge.

[17] Delgado, M. A., J. C. Fariñas, and S. Ruano, (2002). "Firm Productivity and Export Markets: A Non-parametric Approach," *Journal of International Economics*, 57:2, 397-422.

[18] De Loecker, J., (2007). "Do Exports Generate Higher Productivity? Evidence from Slovenia," *Journal of International Economics*, 73:1, 69-98.

[19] De Loecker, J., (2011). "Product Differentiation, Multi-Product Firms and Estimating the Impact of Trade Liberalization on Productivity," forthcoming in *Econometrica*.

[20] Eaton, J. and S. Kortum, (2002). "Technology, Geography, and Trade," *Econometrica*, 70:5, 1741-1779.

[21] Fernandes, A. M. and A. E. Isgut, (2006). "Learning-by-Exporting Effects: Are They for Real?," mimeo, The World Bank and Institute for Competitiveness & Prosperity.

[22] Garicano, L., (2000). "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108:5, 874-904.

[23] Garicano, L. and E. Rossi-Hansberg, (2011). "Organizing Growth," forthcoming in *Journal of Economic Theory*.

[24] Garicano, L. and E. Rossi-Hansberg, (2006). "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 121:4, 1383-1435.

[25] Garicano, L. and E. Rossi-Hansberg, (2004). "Inequality and the Organization of Knowledge," *American Economic Review*, 94:2, 197-202.

[26] Ghironi, F. and M. Melitz, (2005). "International Trade and Macroeconomic Dynamics with Heterogeneous Firms." *Quarterly Journal of Economics* 120.

[27] Guadalupe, M., and J. Wulf, (2010). "The Flattening Firm and Product Market Competition: The Effect of Trade Liberalization on Corporate Hierarchies," *American Economic Journal*: Applied Economics, 2:4, 105-27.

[28] Hallward-Driemeier, M. G. Iarossi, and K. L. Sokoloff, (2005). "Exports and Manufacturing Productivity in East Asia: A Comparative Analysis with Firm-Level Data," mimeo, The World Bank and UCLA.

[29] Krugman P. R., (1980). "Scale Economies, Product Differentiation, and the Pattern of Trade," *The American Economic Review*, 70, 950-959.

[30] Lileeva, A. and D. Trefler, (2010). "Improved Access to Foreign Markets Raises Plant-Level Productivity ... for Some Plants," *Quarterly Journal of Economics*.

[31] Marin, D. and T. Verdier, (2010). "Corporate Hierarchies and International Trade: Theory and Evidence," mimeo, University of Munich.

[32] Marin, D. and T. Verdier, (2008). "Competing in Organizations: Firm Heterogeneity and International Trade" in *The Organization of Firms in a Global Economy*, E. Helpman, D. Marin and T. Verdier (eds), Harvard University Press, Cambridge.

[33] Mayer, T., M. J. Melitz and G. Ottaviano, (2011). "Market size, Competition and the Product Mix of Exporters," mimeo, Harvard University.

[34] Melitz, M. J., (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71:6, 1695–1725.

[35] Park, A., D. Yang, X. Shi, and Y. Jiang, (2006). "Exporting and Firm Performance: Chinese Exporters and the Asian Financial Crisis," mimeo, University of Michigan.

[36] Van Biesebroeck, J., (2004). "Exporting Raises Productivity in Sub-Saharan African Manufacturing Firms," *Journal of International Economics*, 67:2, 373-391.

APPENDIX

Lemma 1: *Given the number of layers L, under Assumption 1, as the multiplier on the output constraint, ϕ, decreases, the constraints over $z_L^l \geq 0$ for all l bind such that the first to hit zero is z_L^{L-1}, then z_L^{L-2} ... until z_L^0 and finally, z_L^L.*

Proof. From the first-order conditions of the cost minimization problem we obtain the following system of $L+1$ equations relating all the z's in a recursive way.

$$e^{\lambda z_L^L} = \phi \frac{\lambda A}{wch} - \frac{\theta_L^L}{wce^{-\lambda z_L^L}}$$

$$e^{-\lambda z_L^L} = \frac{ch}{c + \lambda (cz_L^0 + 1)} + \frac{h\theta_L^L}{w(c + \lambda (cz_L^0 + 1))} - \frac{\theta_L^0}{e^{\lambda z_L^L} n_L^0 w (c + \lambda (cz_L^0 + 1))}$$

$$e^{-\lambda z_L^0} = \frac{c/h}{(c + \lambda (cz_L^1 + 1))} + \frac{\theta_L^0 - \theta_L^1}{n_L^0 hw (c + \lambda (cz_L^1 + 1))}$$

$$e^{-\lambda z_L^{l-1}} = \frac{c}{(c + \lambda (cz_L^l + 1))} + \frac{\left(\theta_L^{l-1} - \theta_L^l\right)}{he^{-\lambda Z_L^{l-2}} n_L^0 w (c + \lambda (cz_L^l + 1))} \text{for all } 1 < l < L$$

Without loss of generality, consider that all knowledge in the organization is positive; therefore, $\theta_L^l = 0$ for all l. Then the system collapses to

$$e^{-\lambda z_L^L} = \frac{hwc}{\phi \lambda A} \tag{23}$$

$$e^{-\lambda z_L^L} = \frac{ch}{c + \lambda (cz_L^0 + 1)} \tag{24}$$

$$e^{-\lambda z_L^0} = \frac{c/h}{c + \lambda (cz_L^1 + 1)} \tag{25}$$

$$e^{-\lambda z_L^{l-1}} = \frac{c}{c + \lambda (cz_L^l + 1)} \text{ for all } 1 < l < L \tag{26}$$

Note that in the cost minimization problem, without restrictions on the number of inputs used, the constraint on quantity is always binding. Therefore $\phi > 0$. We will now solve sequentially, starting from the lowest layer for the value of the Lagrange multiplier ϕ such that it makes each of the $z's$ zero. Let ϕ be large enough. First note from (23) that for any value of $\phi > \phi_L^*$, $z_L^L > 0$, where $\phi_L^* = \frac{hwc}{\lambda A}$. Consider now layer zero and assume that all other z's are positive. Using condition (24) and (23) we can solve for z_L^0 as a function of ϕ. In particular $z_L^0 = \frac{\phi A - w}{wc} - \frac{1}{\lambda}$. This immediately implies that there exists a unique $\phi^* > 0$ given by $\phi_0^* = \frac{w}{A}\left(\frac{c}{\lambda} + 1\right)$ such that $z_L^0 = 0$. Note that $\phi_0^* > \phi_L^*$. This implies that the zero layer constraint binds before the L layer constraint. Now consider layer 1 and solve for ϕ_1^* such that $z_L^1 = 0$ and all other $z's$ are positive. From the first-order conditions with respect to z_L^1 we get $z_L^1(\phi) = \frac{1}{\lambda h} e^{\lambda \left(\frac{\phi A - w}{wc} - \frac{1}{\lambda}\right)} - \frac{1}{\lambda} - \frac{1}{c}$. It immediately follows that there exists a unique ϕ_1^* such that $z_L^1(\phi_1^*) = 0$. Let $\Theta \equiv 1 + \frac{\lambda}{c}$, then $\phi_1^* = \frac{w}{A}\left(1 + \frac{c}{\lambda}(1 + \ln(h\Theta))\right)$. Note that under Assumption 1 $\phi_1^* \geq \phi_0^*$ since $\phi_1^* = \phi_0^* + \frac{1}{A}\left(\frac{wc}{\lambda} \ln(h\Theta)\right)$ and $\ln(h\Theta) \geq 0$. Let's consider now layer 2. We solve for ϕ_2^* such that $z_L^2 = 0$ and consider that all other $z's$ are positive, we get $z_L^2(\phi) = \frac{e^{\lambda z_L^1(\phi)}}{\lambda} - \frac{1}{\lambda} - \frac{1}{c}$. Note that there exists a ϕ_2^* such that $z_L^2(\phi_2^*) = 0$, $\phi_2^* = \phi_0^* + \frac{1}{A}\left(\frac{wc}{\lambda} \ln(h \ln(\Theta) + h\Theta)\right)$. Note that $\phi_2^* > \phi_1^*$ since $\phi_2^* - \phi_1^* = \frac{1}{A}\frac{wc}{\lambda}(\ln(\ln(\Theta) + \Theta) - \ln(\Theta)) > 0$. Continuing in a recursive way we will find that $\phi_{L-1}^* > \phi_{L-2}^*$

since $\phi_{L-1}^* - \phi_{L-2}^* = \frac{1}{A}\frac{wc}{\lambda}\left(\ln\left(\ln\left(\Theta\right)+\Theta\right)-\ln\left(\Theta\right)\right) > 0$. Therefore, $\phi_{L-1}^* > \phi_{L-2}^*... > \phi_1^* \geq \phi_0^* > \phi_L^*$. Hence, as the multiplier over the output constraint, ϕ, decreases, the constraints over $z_L^l \geq 0$ bind in a descending order starting from z_L^{L-1} up to z_L^0 and then z_L^L. ∎

Proposition 1 *Under Assumption 1, for all $L \neq 1$ and any output level q, the knowledge of agents at all layers is positive ($z_L^l \geq 0$ never binds).*

Proof. Let $L > 1$ and consider two organizations, an unconstrained organization L with L layers, and a constrained organization L' with L' layers, where $L' = L + 1$ and constrained to $z_{L'}^{L'-1} = 0$. From the cost minimization problem of the firm, the first-order conditions for the L organization are

$$c + \lambda\left(cz_L^0 + 1\right) - \lambda A\frac{\phi}{w} + \frac{\theta_L^0}{wn_L^0} \leq 0$$

for z_L^0,

$$he^{-\lambda Z_L^{l-1}}c + \lambda\left(cz_L^0 + 1\right) + \lambda h\sum_{k=0}^{l-1}e^{-\lambda Z_L^k}\left(cz_L^{k+1} + 1\right) - \lambda A\frac{\phi}{w} + \frac{\theta_L^l}{wn_L^0} \leq 0$$

for all $0 < l < L$, and

$$c - \frac{\phi}{w}\frac{\lambda A}{h}e^{-\lambda z_L^L} + \frac{\theta_L^L}{w} \leq 0$$

for z_L^L. Now consider the problem of the constrained organization L'. The cost function to minimize is

$$\frac{w\left(cz_{L'}^0 + 1 + h\sum_{l=0}^{L'-3}e^{-\lambda Z_{L'}^l}c\left(z_{L'}^{l+1} + 1\right) + he^{-\lambda Z_{L'}^{L'-2}}\left(cz_{L'}^{L'} + 1\right) + he^{-\lambda Z_{L'}^{L'-2}}\right)}{he^{-\lambda Z_{L'}^{L'-2}}}$$

The first-order conditions of the cost minimization problem are

$$c + \lambda\left(cz_{L'}^0 + 1\right) - \frac{\phi}{w}\lambda A + \frac{\theta_{L'}^0}{wn_{L'}^0} \leq 0$$

for $z_{L'}^0$,

$$he^{-\lambda Z_{L'}^{l-1}}c + \lambda\left(cz_{L'}^0 + 1\right) + \lambda h\sum_{k=0}^{l-1}e^{-\lambda Z_{L'}^k}\left(cz_{L'}^{k+1} + 1\right) - \lambda A\frac{\phi}{w} + \frac{\theta_{L'}^l}{wn_{L'}^0} \leq 0$$

for all $0 < l < L' - 1$, and

$$c - \frac{\phi}{w}\frac{\lambda A}{h}e^{-\lambda z_{L'}^{L'}} + \frac{\theta_{L'}^{L'}}{w} \leq 0$$

for $z_{L'}^{L'}$. Note that these first-order conditions are exactly the same as the ones of the L layer firm since $n_L^0 = \frac{1}{h}e^{\lambda Z_L^{L-1}}$ and $n_{L'}^0 = \frac{1}{h}e^{\lambda Z_{L'}^{L'-2}}$. Therefore, the firms will choose the same $z's$. However, organization L' will have more employees than organization L since it has one more layer. Therefore, the total cost of organization L' is strictly larger than organization L, by w. Note that the number of managers at layer $L' - 1$ that the entrepreneur hires is $n_{L'}^{L'-1} = n_{L'}^0 he^{-\lambda Z_{L'}^{L'-2}} = 1$. Hence, the firm will never choose to have an organization with $z_L^{L-1} = 0$. Lemma 1 then implies that it will never choose to have $z_L^l = 0$ for $l < L - 1$.

The case for $L = 0$ is trivial, since $z_0^0 = 0$ implies that the output of the firm is equal to zero. ∎

Proposition 2: *Given the number of layers L, the marginal cost, $MC_L\left(q;w\right)$, and the knowledge of all employees, z_L^l for all l, increase with q. Furthermore, $MC_L\left(q;w\right) > 0$, for all L and q, and so the cost function $C\left(q;w\right)$ is strictly increasing.*

Proof. From the first-order conditions of the cost minimization problem we have that

$$\frac{\partial \phi}{\partial q} = \frac{\frac{hwc}{A\lambda}e^{\lambda z_L^L}}{A\left[e^{\lambda Z_L^L}\left(1 + \frac{n_L^0}{h}\sum_{l=1}^{L-1}\left(1/n_L^l\right)\right) + \frac{1}{h}e^{-\lambda z_L^L}\right]} > 0$$

where we used that

$$\frac{\partial z_L^l}{\partial z_L^L} = \begin{cases} \frac{\partial z_L^0}{\partial z_L^L} = he^{\lambda z_L^L} > 0 & \text{for } l = 0 \\ \frac{\partial z_L^0}{\partial z_L^L}\prod_{v=1}^{l}\frac{\partial z_L^v}{\partial z_L^{v-1}} = e^{\lambda z_L^L}e^{\lambda Z_L^{l-1}} > 0 & \text{for } l > 0 \end{cases},$$

$$\frac{\partial Z_L^{L-1}}{\partial z_L^L} = \sum_{l=0}^{L-1}\frac{\partial z_L^l}{\partial z_L^L} = \frac{\partial z_L^0}{\partial z_L^L}\left(1 + \sum_{l=1}^{L-1}\prod_{v=1}^{l}\frac{\partial z_L^v}{\partial z_L^{v-1}}\right) = he^{\lambda z_L^L}\left(1 + \frac{1}{h}\sum_{l=1}^{L-1}e^{\lambda Z_L^{l-1}}\right),$$

and

$$\frac{\partial z_L^L}{\partial \phi} = \frac{A}{hwce^{\lambda z_L^L}} > 0.$$

Note also that since $MC_L(q;w) = \frac{hwc}{\lambda A}e^{\lambda z_L^L} > 0$ for any z_L^L, the cost function $C_L(q;w)$ is strictly increasing. Since this is the case for all L, $C(q;w)$ is strictly increasing. ∎

Proposition 3: *Suppose that the entrepreneur finds it optimal to increase the number of layers at output level q, then the marginal cost decreases discontinuously at q.*

Proof. Consider an organization with L layers producing q and let ϕ_L be the multiplier under organization L. Now suppose that we are at the q such that the entrepreneur finds it optimal to increase the number of layers, say, to $L' = L + 1$ layers. Note that the cost minimization problem of organization L' has an extra first-order condition compared to organization L, the first-order condition with respect to $z_{L'}^{L'-1}$. Let the multiplier under organization L' be $\phi_{L'}$. First note $\phi_{L'} = \phi_L$ is inconsistent with fixing q. The reason is that if $\phi_{L'} = \phi_L$ then from the first-order conditions we have that $z_{L'}^l = z_L^l$ for $l = 0, ..., L-1$, $z_{L'}^{L'} = z_L^L$ and $z_{L'}^{L'-1} > 0$ by Proposition 1, since it was optimal to increase the number of layers. This implies that $Z_{L'}^{L'-1} > Z_L^{L-1}$ and the organization will produce $\frac{A}{h}[e^{\lambda Z_{L'}^{L'-1}} - e^{-\lambda z_{L'}^{L'}}] > q$. Following the same argument, $\phi_{L'} > \phi_L$ is inconsistent with fixing q, since the organization will also produce more than q given that each z is an increasing function of $\phi_{L'}$. Therefore, the only way in which L' is optimal and it can produce q is if $\phi_{L'} < \phi_L$. The marginal cost (ϕ) decreases as the number of layers (L) increases. ∎

Proposition 4 *Given the number of layers L, the cost function and the corresponding factor demands have the following properties: i. The cost function is homogeneous of degree one in w. ii. Conditional factor demands are homogeneous of degree zero in w. iii. The average cost curve is convex in q, has its minimum where it crosses the marginal cost curve, and converges to infinity when $q \to 0$ or $q \to \infty$.*

Proof. The proof of i. is immediate from (6) and ii.

The claim in ii. follows from the first-order conditions of the cost minimization problem. Note that equations (3) and (4), with $\theta_L^l = 0$ for all l and L depend on w only through the term ϕ/w. Thus, if the system of first-order conditions is satisfied with w, and ϕ. If $w' = tw$ for $t > 0$, the system is satisfied with $t\phi$ and the rest of the allocation does not change. Hence, factor demands do not change with w.

iii) consider first the case of a self-employed entrepreneur, $L = 0$. The average cost is given by

$$AC_0\left(q;w\right) = w\left(\frac{c}{\lambda q}\ln\left(\frac{A}{A-q}\right) + \frac{1}{q}\right)$$

The partial derivative of the $AC_0\left(q;w\right)$ with respect to q is

$$\frac{\partial AC_0\left(q;w\right)}{\partial q} = -\frac{AC_0\left(q;w\right) - MC_0\left(q;w\right)}{q}$$

Note that the average cost is decreasing if $AC_0\left(q;w\right) > MC_0\left(q;w\right)$ and increasing if $AC_0\left(q;w\right) < MC_0\left(q;w\right)$. At $AC_0\left(q;w\right) = MC_0\left(q;w\right)$ it has a unique stationary point q^* that solves $\frac{c}{\lambda}\ln\left(\frac{A}{A-q^*}\right) + w = \frac{c}{\lambda}\frac{q^*}{A-q^*}$. This stationary point is a minimum since

$$\frac{\partial^2 AC_0\left(q;w\right)}{\partial q} = \frac{c}{\lambda}\frac{1}{q^2} + \left(\frac{2AC_0\left(q;w\right) + \frac{c}{\lambda A}\left(A-q\right)}{q^2}\right) > 0 \text{ since } q < A,$$

and

$$\lim_{q\to 0} AC_0\left(q;w\right) = +\infty, \quad \lim_{q\to A} AC_0\left(q;w\right) = +\infty,$$

which shows that the average cost function is strictly convex. The marginal cost is

$$MC_0\left(q;w\right) = \frac{wc}{\lambda}\frac{1}{A-q} = \frac{wc}{\lambda A}e^{\lambda z_0^0}.$$

Now consider the cost function of an organization with L layers producing q

$$C_L\left(q;w\right) = w\left(\frac{ch}{\lambda A}e^{\lambda z_L^L}q + \frac{c}{\lambda}\left(1 - e^{\lambda z_L^{L-1}} + \lambda z_L^L + \frac{\lambda}{c}\right)\right).$$

Recall that the marginal cost is given by $MC_L\left(q;w\right) = \frac{wch}{\lambda A}e^{\lambda z_L^L}$. This implies that

$$\Rightarrow \left(\frac{ch}{A}e^{\lambda z_L^L}\frac{\partial z_L^L}{\partial q} + \frac{c}{q}\left(\frac{\partial z_L^L}{\partial q} - e^{\lambda z_L^{L-1}}\frac{\partial z_L^{L-1}}{\partial q}\right)\right)q = 0. \tag{27}$$

Now consider the average cost of this organization,

$$AC_L\left(q;w\right) = w\left(\frac{ch}{\lambda A}e^{\lambda z_L^L} + \frac{c}{\lambda}\frac{1 - e^{\lambda z_L^{L-1}} + \lambda z_L^L + \lambda/c}{q}\right).$$

Some properties of this function are

$$\lim_{q\to 0} AC_L\left(q;w\right) = +\infty, \lim_{q\to +\infty} AC_L\left(q;w\right) = +\infty,$$

where we use the fact that $\partial z_L^L/\partial q > 0$ and $\partial z_L^{L-1}/\partial q > 0$, which follows from Proposition 2. The partial derivative of the $AC_L\left(q;w\right)$ with respect to q is

$$\frac{\partial AC_L\left(q;w\right)}{\partial q} = -\frac{1}{q}\left(AC_L\left(q;w\right) - MC_L\left(q;w\right)\right),$$

where we use (27). Note that the average cost is decreasing if $AC_L\left(q;w\right) > MC_L\left(q;w\right)$ and increasing if $AC_L\left(q;w\right) < MC_L\left(q;w\right)$. It has a stationary point where they are equal. In particular, there is a unique stationary point q^* such that $\frac{\partial AC_L\left(q^*;w\right)}{\partial q} = 0$ and q^* solves

$$\Rightarrow AC_L\left(q^*;w\right) = \frac{wch}{\lambda A}e^{\lambda z_L^L\left(q^*\right)}.$$

By evaluating the second derivative of the average cost function, we show that the stationary point is a minimum. That is

$$\frac{\partial^2 AC_L\left(q;w\right)}{\partial q} = \frac{2}{q^2}\left(AC_L\left(q;w\right) - \frac{wch}{\lambda A}e^{\lambda z_L^L}\right) + \frac{wch}{qA}e^{\lambda z_L^L}\frac{\partial z_L^L}{\partial q},$$

and so at q^*,

$$\frac{\partial^2 AC_L\left(q^*;w\right)}{\partial q} = \frac{wch}{qA}e^{\lambda z_L^L(q^*)}\frac{\partial z_L^L}{\partial q} > 0.$$

This implies that the average cost function is convex and it reaches a minimum when it intersects with the marginal cost function. Also note that the MES is implicitly given by

$$e^{-\lambda z_L^{L-1}(q^*)} = \frac{c}{c + \lambda\left(cz_L^L\left(q^*\right) + 1\right)}.$$

∎

Proposition 5 *The minimum of the average cost for a given number of layers decreases with the number of layers.*

Proof. Consider an organization with L layers producing q^* at the MES. From the first-order conditions of the cost minimization problem, together with the fact that the firm is producing at its MES, the $L+1$ knowledge levels of the organization solve the following set of $L+1$ equations

$$\begin{aligned}
\lambda z_L^L &= \ln\left(\lambda z_L^0 + \Theta\right) - \ln h, \\
\lambda z_L^0 &= \ln\left(\lambda z_L^1 + \Theta\right) + \ln h, \\
\lambda z_L^{l-1} &= \ln\left(\lambda z_L^l + \Theta\right) \quad \text{for all } 1 < l \leq L,
\end{aligned}$$

where $\Theta \equiv 1 + \frac{\lambda}{c}$. Define the function $g\left(x\right) = \ln x + \Theta$ and abusing notation we will refer to $g^L\left(v\right)$ to the operator applied to the function such that it evaluates the function L times over itself, namely, $g^L\left(v\right) = g\left(g\left(.....g\left(v\right)\right)\right)$ where we are applying the g function L times.

Some properties of the function g are:

Property 1: The function $g^l\left(v\right)$ evaluated at Θ, $g^l\left(\Theta\right)$, is increasing in l. So

$$g^L\left(\Theta\right) > g^{L-1}\left(\Theta\right) > \ln\Theta + \Theta > 1 \text{ for all } L.$$

Property 2: There exists a unique v^*, such that $v^* > \Theta$ and $g^L\left(v^*\right) = g^{L+1}\left(v^*\right)$. Note that the function $g\left(v\right)$ is continuously increasing. The $\lim_{v\to\infty} g\left(v\right) = +\infty$, $\lim_{v\to 0} g\left(v\right) = -\infty$, $g'\left(v\right) = \frac{1}{v}$. $g'\left(v\right) \geq 1$ for all $v \leq 1$ and $g'\left(v\right) \leq 1$ for all $v \geq 1$ and also $\lim_{v\to\infty} g'\left(v\right) = 0$. Note that at $v = \Theta$, $g\left(\Theta\right) = \ln\Theta + \Theta > \Theta$. By the intermediate value theorem, the function $g\left(v\right)$ has two stationary points. In particular, there exists a unique $v^* > \Theta$ such that $g^L\left(v^*\right) = g^{L+1}\left(v^*\right)$.

Property 3: $\lim_{v\to\infty} g\left(v\right) = +\infty$ and this implies that $\lim_{v\to\infty} g^l\left(v\right) = +\infty$ for all l.

Property 4: The partial derivative of $g^l\left(v\right)$ with respect to v is given by

$$\frac{\partial g^L\left(v\right)}{\partial v} = g'\left(v\right)\prod_{i=1}^{L-1} g'\left(g^i\left(v\right)\right) = g'\left(g^{L-1}\left(v\right)\right)\frac{\partial g^{L-1}\left(v\right)}{\partial v}.$$

Property 5: $g'(v) = \frac{1}{v} \Rightarrow g'(v) < 1$ for all $v > 1$. $g'(g(v)) = \frac{1}{g(v)} < 1$ for all $v > \exp(1 - \Theta)$ since $g(v) > 1$ for all $v > \exp(1 - \Theta)$.

These 5 properties of g are going to be used in the remaining part of the proof. We define the following two functions

$$f(v) = \frac{v - \Theta + \ln h}{\lambda},$$

$$G(v, L) = \frac{1}{\lambda} \ln \left(g^L(v) + \ln h \right).$$

Note that the solution for z_L^L at the MES solves the following equation

$$f\left(\lambda z_L^L + \Theta\right) = G\left(\lambda z_L^L + \Theta, L\right).$$

We will now show that there is a unique solution to this equation and that the solution decreases with L. Consider the following change of variables, $v = \lambda z + \Theta$. Note that when $z = 0$, $v = \Theta$; and when $z \to \infty$, $v \to \infty$. First note that at $v = \Theta$,

$$f(\Theta) = \frac{\ln h}{\lambda} < 0,$$

and

$$
\begin{aligned}
G(\Theta, L) &= \frac{1}{\lambda} \ln \left(g^L(\Theta) + \ln h \right) \\
&> \frac{1}{\lambda} \ln \left(\ln \Theta + \Theta + \ln h \right) \text{ by Property 1} \\
&> 0 \text{ by Assumption 1.}
\end{aligned}
$$

Note that both functions converge to infinity as v converges to infinity

$$
\begin{aligned}
\lim_{v \to +\infty} f(v) &= +\infty, \\
\lim_{v \to +\infty} G(v, L) &= \frac{1}{\lambda} \ln \left(g^L(+\infty) + \ln h \right) = +\infty \text{ by Property 3.}
\end{aligned}
$$

Now note that both functions are strictly increasing,

$$f'(v) = \frac{1}{\lambda}$$

$$G'(v, L) = \frac{1}{\lambda} \frac{1}{g^L(v) + \ln h} \frac{\partial g^L(v)}{\partial v} = \left(\frac{\prod_{i=1}^{L-1} g'\left(g^i(v)\right)}{g^L(v) + \ln h} \right) \frac{g'(v)}{\lambda} \text{ by Property 4,}$$

and

$$G'(v, L) < \frac{1}{\lambda} \text{ for all } z > 0 \text{ by Property 5.}$$

We have just shown that the functions cross only once. This follows since, at $z = 0$, $f(\Theta) < G(\Theta, L)$ and both functions are continuously increasing where f increases at a higher rate than G. By the intermediate Value theorem, this immediately implies that there exists a unique z_L^L that solves $f\left(\lambda z_L^L + \Theta\right) = G\left(\lambda z_L^L + \Theta, L\right)$.

48

We now characterize how z_L^L changes as we change L. We show that the value of z_L^L such that $f\left(\lambda z_L^L + \Theta\right) = G\left(\lambda z_L^L + \Theta, L\right)$ decreases as we increase L. In order to do so, we first show that the function $G\left(v, L\right)$ increases in L at $z = 0$ $(v = \Theta)$. Then we show that there exists a unique $z^* > 0$ $(v^* > \Theta)$ such that $G\left(v^*, l\right)$ has the same value for all l. Then we show that the z_L^L that solves $f\left(\lambda z_L^L + \Theta\right) = G\left(\lambda z_L^L + \Theta, L\right)$ has to be larger than z^*. Finally we show that z_L^L decreases with L.

First note that $G\left(\Theta, l\right)$ increases with l, since

$$
\begin{aligned}
G\left(\Theta, l\right) &= \frac{1}{\lambda} \ln\left(g^l\left(\Theta\right) + \ln h\right) \\
&> \frac{1}{\lambda} \ln\left(g^{l-1}\left(\Theta\right) + \ln h\right) = G\left(\Theta, l-1\right) \text{ by Property 1}
\end{aligned}
$$

From property 2, there exists a unique $v^* > \Theta$ such that $g^l\left(v^*\right) = g^{l+1}\left(v^*\right)$ and this implies that $G\left(v^*, l\right) = G\left(v^*, l+1\right)$, for all $l > 2$. Note that at v^*, $G\left(v^*, L\right) > f\left(v^*\right)$. Since

$$
G\left(v^*, L\right) = \frac{1}{\lambda} \ln\left(v^* + \ln h\right)
$$

and

$$
f\left(v^*\right) = \frac{v^* - \Theta + \ln h}{\lambda} = \frac{\ln v^* + \ln h}{\lambda} \text{ using that } v^* = \ln v^* + \Theta,
$$

and so this implies that $G\left(v^*, L\right) > f\left(v^*\right)$ for all $1 > h > 1/v^*$. Note that $h > 1/\Theta > 1/v^*$ by Assumption 1 and Property 2.

Finally note that $G'\left(v, L-1\right) > G'\left(v, L\right)$. Since

$$
\begin{aligned}
G'\left(v, L\right) &= \frac{1}{\lambda} \frac{1}{g^L\left(v\right) + \ln h} g'\left(g^{L-1}\left(v\right)\right) \frac{\partial g^{L-1}\left(v\right)}{\partial v}, \\
G'\left(v, L-1\right) &= \frac{1}{\lambda} \frac{1}{g^{L-1}\left(v\right) + \ln h} \frac{\partial g^{L-1}\left(v\right)}{\partial v},
\end{aligned}
$$

and so

$$
\begin{aligned}
\frac{G'\left(v, L\right)}{G'\left(v, L-1\right)} &= \frac{g^{L-1}\left(v\right) + \ln h}{g^L\left(v\right) + \ln h} g'\left(g^{L-1}\left(v\right)\right) \\
&= \left(\frac{1}{g^L\left(v\right) + \ln h}\right)\left(\frac{g^{L-1}\left(v\right) + \ln h}{g^{L-1}\left(v\right)}\right) < 1
\end{aligned}
$$

Therefore, the knowledge of the entrepreneur, z_L^L, at the MES, q^*, decreases with L. Since at the MES the average cost is equal to the marginal cost and the marginal cost is an increasing function of z_L^L, we have shown that the minima of the average cost decrease with the number of layers, L. ∎

Proposition 6 *The level of output that achieves the MES, q_L^*, increases with the number of layers. Furthermore, the average cost curves of an organization with L and $L+1$ layers cross in the interval $\left(q_L^*, q_{L+1}^*\right)$.*

Proof. Consider an organization with L layers producing at the MES, q_L^*. Given q_L^*, this organization has no incentives to modify its knowledge level at any layer. To see this, consider adding a top layer to this organization with a span of control of one. Assigning positive knowledge to the new top layer is sub-optimal,

since it has the same cost and benefit than increasing the knowledge of the manager at layer L. Given that the knowledge of this manager was already optimal and that the cost of acquiring knowledge is linear, a new top manager with span of control of one will increase the total cost by w but will not change production.

Now consider an organization with $L+1$ layers constrained to produce q_L^* where the span of control of the manager is greater than one, as in our model. Consider the case where the knowledge of the top manager at layer $L+1$ is zero. First note that this organization, from layer L down the hierarchy, is a replica of the organization with L layers. The difference is that there are n_{L+1}^L 'managers at the top' replicas of these $L-$suborganizations. This immediately implies that in order to produce q_L^* each of these $L-$ suborganizations has to produce less than q_L^*. That is, the cost of the $L+1$ organization with zero knowledge at the top that produces q_L^* is equal to

$$w + AC_L \left(\frac{q_L^*}{n_{L+1}^L}; w \right) q_L^* > AC_L \left(q_L^*; w \right) q_L^*,$$

where the inequality is the result of the convexity of $AC(\cdot)$ (Proposition 4) and that q_L^* achieves the MES with L layers. Note that if each of these $L-$suborganizations produces at its MES, q_L^*, then the $L+1$ organization will be producing more than q_L^*, namely $n_{L+1}^L q_L^*$. Such an organization will have an average cost of $AC_L \left(q_L^*; w \right)$ which is lower than $AC_{L+1} \left(q_L^*; w \right)$ with $z_{L+1}^{L+1} = 0$, but it will be producing strictly more than q_L^* violating the constraint of producing q_L^*.

In general, $z_{L+1}^{L+1} > 0$, since the span of control of the manager at the top is larger than one, which results in suboptimal knowledge acquisition in the $L-$suborganization. However, this adjustment is just the result of the inability of the organization to have $L-$suborganizations that produce q_L^*. Hence, the $L+1$ organization necessarily incurs a higher average cost when producing q_L^*. Since by Proposition 5 we know that organizations with higher layers have lower minimum average costs, this immediately implies that the $L+1$ organization is not at its MES at q_L^*. Moreover, we also just argued that if we relaxed the constraint over the quantity produced, the $L+1$ organization is able to produce at a lower average cost but with a higher scale, for example, by producing $n_{L+1}^L q_L^*$. Since the average cost is a convex function, and since when producing more than q_L^* the $L+1$ layer organization has strictly lower average costs, this implies that $q_{L+1}^* > q_L^*$. Thus, the MES of the $L+1$ organization has a higher scale of production than the L layer organization. $\blacksquare$

Proposition 7 *In the limit when $c/\lambda \to 0$ and $L \geq 1$, the cost function is given by*

$$C \left(q; w \right) = w \left(\frac{q}{A} + 1 \right)$$

and so

$$AC \left(q; w \right) = w \left(\frac{1}{A} + \frac{1}{q} \right)$$

and

$$MC \left(q; w \right) = \frac{w}{A}.$$

Proof. Consider an organization with $L = 1$, fix c and let $\lambda \to \infty$ so that $c/\lambda \to 0$. We guess that $\lim_{\lambda \to \infty} \lambda z_1^1 = \infty$ and verify our guess. Note that this guess combined with the first-order condition implies that

$$\lim_{\lambda \to \infty} \lambda z_1^1 = \infty \Rightarrow \lim_{\lambda \to \infty} \lambda z_1^0 = \ln \frac{qh}{A}.$$

Then from the first-order condition with respect to z_1^0

$$\lim_{\lambda \to \infty} \lambda z_1^1 = \lim_{\lambda \to \infty} \ln \frac{1}{h}\left(1 + \ln \frac{qh}{A} + \frac{\lambda w}{c}\right) = \infty, \tag{28}$$

which verifies our guess. This also implies that $\lim_{\lambda \to \infty} z_1^1 = 0$. Note that from (28) we get

$$\lim_{\lambda \to \infty} \frac{\ln \frac{1}{h}\left(1 + \ln \frac{qh}{A} + \frac{\lambda w}{c}\right)}{\lambda} = \lim_{\lambda \to \infty} \frac{\frac{w}{c}}{\left(1 + \ln \frac{qh}{A} + \frac{\lambda w}{c}\right)} = 0.$$

Then as $\lambda \to \infty$

$$\lim_{\lambda \to \infty} MC_1(q; w) = \lim_{\lambda \to \infty} \frac{hc}{A\lambda e^{\ln h/\left(1 + \ln \frac{qh}{A} + \frac{\lambda w}{c}\right)}} = \lim_{\lambda \to \infty} \frac{c + c\ln \frac{qh}{A} + \lambda w}{A\lambda} = \frac{w}{A},$$

$$\lim_{\lambda \to \infty} AC_1(q; w) = \lim_{\lambda \to \infty} \frac{w + \frac{A}{q}w + \frac{A}{q}cz_1^1}{A} = w\left(\frac{1}{A} + \frac{1}{q}\right),$$

$$\lim_{\lambda \to \infty} C_1(q; w) = w\left(\frac{q}{A} + 1\right) = w\frac{q}{A} + w.$$

In the limit economy, when $\lambda \to \infty$, the entrepreneur is able to solve all problems without any knowledge. Therefore, the problems that the workers generate are all converted into output but $z_1^0 = 0$. Since no knowledge is needed for production, there are no costs incurred in acquiring knowledge. The only costs are the wages that the manager has to pay to the workers and to himself. The number of workers in a limit economy is given by $n_1^0 = q/A$; therefore, the cost of production is wq/A plus w which is the wage of the entrepreneur. The marginal cost of increasing production is the extra cost of hiring more workers, w/A. Note that there are no gains from adding an extra layer of problem solvers. This will only increase the cost of the organization, since it will have to pay more salaries. ∎

Proposition 8 *The quantity produced, $q(\alpha)$, and revenue, $r(\alpha)$, increase continuously with α for a given number of layers and increase discontinuously when the firm increases the number of layers. The price, $p(\alpha)$, increases continuously with α for a given number of layers and decreases discontinuously when the firm increases the number of layers.*

Proof. From Proposition 2 we know that the marginal cost increases with q, and at the q such that the entrepreneur finds it optimal to switch layers, the marginal cost decreases (Proposition 3), since the price set by the entrepreneur is a constant markup over marginal cost. The properties of the marginal cost schedule are reflected in the pricing rule of the monopolist, namely, that prices increase within a number of layers and they jump down as we switch layers. We now need to show that q increases with α. Within a number of layers,

$$\frac{\partial q(\alpha)}{\partial \alpha} = \frac{q(\alpha)/\alpha}{1 + \frac{\sigma q(\alpha)}{MC(q(\alpha);w)}\frac{\partial MC(q(\alpha);w)}{\partial q}} = \frac{q(\alpha)/\alpha}{1 + \sigma\lambda q(\alpha)\frac{\partial z_L^L}{\partial q}} > 0$$

which follows from $\partial MC\left(q\left(\alpha\right);w\right)/\partial q\left(\alpha\right)>0$. Note that when the organization increases layers, $q\left(\alpha\right)$ does not move continuously with α. This follows from the dependence of the production supplied on the marginal cost. In particular, q is a decreasing function of the marginal cost,

$$\frac{\partial q\left(\alpha\right)}{\partial MC\left(q\left(\alpha\right);w\right)}=-\sigma\frac{q\left(\alpha\right)}{MC\left(q\left(\alpha\right);w\right)}$$

Therefore, when the marginal cost decreases as a consequence of adding a new layer, the new quantity supplied goes up discontinuously. The elasticity of the increase in $q\left(\alpha\right)$ with respect to the marginal cost is given by σ.

Since $r\left(\alpha\right)=p\left(\alpha\right)q\left(\alpha\right)$, the elasticity of $r\left(\alpha\right)$ with respect to $MC\left(q\left(\alpha\right);w\right)$ is equal to $1-\sigma<0$, and so revenues jump up when $MC\left(q\left(\alpha\right);w\right)$ jumps down as a result of an increase in the number of layers. ∎

Proposition 9 *Given L, the profit function $\tilde{\pi}\left(q\left(\alpha\right);w\right)$ is strictly concave in q. Furthermore, $\pi\left(\alpha\right)$ is increasing and continuous in α.*

Proof. The result follows immediately from

$$\frac{\partial^{2}\pi\left(\alpha;w\right)}{\partial q} = -\left(\frac{\sigma-1}{\sigma^{2}}\right)q\left(\alpha\right)^{-1/\sigma-1}\left(\alpha R\right)^{1/\sigma}P^{(\sigma-1)/\sigma}-\frac{\partial MC\left(q\left(\alpha\right);w\right)}{\partial q}<0$$

$$\text{since } \frac{\partial MC\left(q\left(\alpha\right);w\right)}{\partial q} = \begin{cases} \frac{wc}{A}e^{\lambda z_{0}^{0}}\frac{\partial z_{0}^{0}}{\partial q}>0 & \text{for } L=0 \\ \frac{wch}{A}e^{\lambda z_{L}^{L}}\frac{\partial z_{L}^{L}}{\partial q}>0 & \text{for } L>0 \end{cases}.$$

Note also that since the profit function is given by

$$\pi\left(\alpha\right)=\alpha R\left(\frac{\sigma}{\sigma-1}MC\left(q\left(\alpha\right);w\right)\right)^{1-\sigma}-C\left(q\left(\alpha\right);w\right)-wf$$

and the cost and marginal cost functions are not directly affected by α, $\pi\left(\alpha\right)$ is increasing in α. Since at a change in layers profits have to be identical, $\pi\left(\alpha\right)$ is continuous. ∎

Proposition 10 *There exists a unique stationary equilibrium.*

Proof. Define the following functions,

$$\Delta^{ZCP}\left(\alpha,w\right) = \pi\left(\alpha\right)=0,$$
$$\Delta^{FE}\left(\alpha,w\right) = \int_{\alpha}^{\infty}\pi\left(x\right)g\left(x\right)dx-\delta wf^{E}=0.$$

An example with the properties that we prove is presented in Figure A1.

First we characterize the zero cut-off profit (ZCP) schedule. Note that the schedule starts at $(0,0)$ since

$$\Delta^{ZCP}\left(0,0\right)=0$$

Then note that it is increasing in the (α,w) plane, since

$$\frac{\partial w}{\partial\alpha}=-\frac{\Delta_{\alpha}^{ZCP}}{\Delta_{w}^{ZCP}}>0$$

which follows from

$$\Delta_{\alpha}^{ZCP} = \frac{\partial \pi (\alpha)}{\partial \alpha} = \left(\frac{\partial r (\alpha)}{\partial q} - MC (\alpha; w) \right) \frac{\partial q (\alpha)}{\partial \alpha} + \frac{\partial r (\alpha)}{\partial \alpha} = \frac{1}{\sigma} \frac{r (\alpha)}{\alpha} > 0$$

$$\Delta_{w}^{ZCP} = \frac{\partial \pi (\alpha)}{\partial w} = \frac{1 - \sigma}{\sigma} \frac{r (\alpha, w)}{w} < 0$$

Implicitly define $\bar{\alpha} (w)$ by $\Delta^{ZCP} (\tilde{\alpha} (w), w) = 0$. Then we have proven that $\tilde{\alpha} (0) = 0$ and that $\tilde{\alpha} (w)$ is strictly increasing.
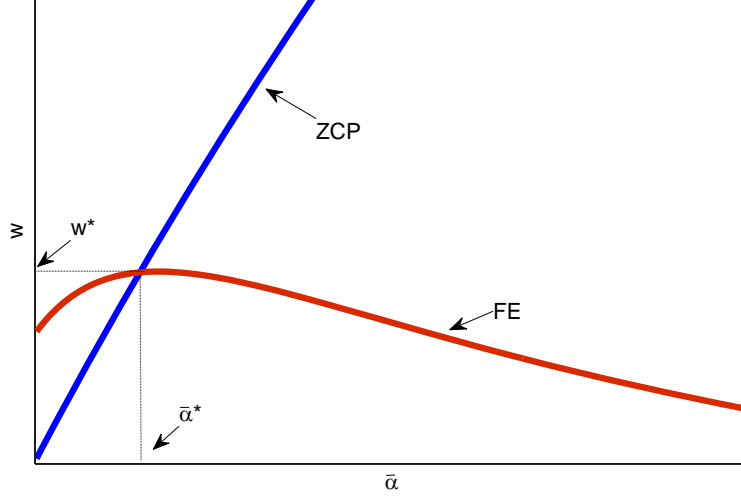

Figure A1: Equilibrium w^* and $\bar{\alpha}^*$

Let's focus now on the $\Delta^{FE} (\alpha, w)$ schedule. Note that it starts at $(0, \hat{w})$ where $\hat{w} > 0$, since $\Delta^{FE} (0, \hat{w}) = 0$, where $\hat{w}$ is defined as

$$\hat{w} = \left(\frac{\sigma - 1}{\sigma} \right)^{\frac{1}{1 - \sigma}} \left(\frac{\tilde{C} + (f + \delta f^E)}{\hat{N} \left(\frac{\sigma}{\sigma - 1} \right)^{-\sigma} \widetilde{MC}} \right)^{\frac{1}{1 - \sigma}} > 0$$

and where $\tilde{C} = \int_{0}^{\infty} \frac{C(\alpha; w)}{w} g (\alpha) d\alpha$ and $\widetilde{MC} = \int_{0}^{\infty} \alpha \left(\frac{MC(\alpha; w)}{w} \right)^{1 - \sigma} g (\alpha) d\alpha$ are constants.
    Also note that in the limit, as $\alpha \to \infty$, $w \to 0$, so

$$\Delta^{FE} (\infty, 0) = 0.$$

This implies, by the intermediate value theorem, that there exists an $\bar{\alpha}^*$ and w^* such that $\Delta^{ZCP} (\bar{\alpha}^*, w^*) = 0$ and $\Delta^{FE} (\bar{\alpha}^*, w^*) = 0$.

We now show that the $\Delta^{FE} (\alpha, w)$ schedule increases until it reaches the $\Delta^{ZCP} (\alpha, w)$ schedule and it declines thereafter. Note that

$$\frac{\partial w}{\partial \alpha} = -\frac{\Delta_{\alpha}^{FE}}{\Delta_{w}^{FE}} \quad \begin{array}{ll} > 0 & \text{if } \alpha < \tilde{\alpha} (w) \\ < 0 & \text{if } \alpha > \tilde{\alpha} (w) \end{array}$$

which follows from

$$\Delta_\alpha^{FE} = -\pi\left(\alpha\right)g\left(\alpha\right) \quad \begin{array}{ll} > 0 & \text{if } \alpha < \tilde{\alpha}\left(w\right) \\ < 0 & \text{if } \alpha > \tilde{\alpha}\left(w\right) \end{array}$$

and

$$\Delta_w^{FE} = \int_\alpha^\infty \frac{\partial \pi\left(x\right)}{\partial w} g\left(x\right) dx - \delta f^E = \frac{1-\sigma}{\sigma} \int_\alpha^\infty \frac{r\left(x\right)}{w} g\left(x\right) dx < 0.$$

Hence, the pair $(\bar{\alpha}^*, w^*)$ that solves $\Delta^{ZCP}\left(\bar{\alpha}^*, w^*\right) = 0$ and $\Delta^{FE}\left(\bar{\alpha}^*, w^*\right) = 0$ is unique. Equations (14) and (15) can then be directly solved to obtain M and R. This implies that an equilibrium of this economy exists and is unique. $\blacksquare$

Proposition 11 *In equilibrium a larger population size does not affect the entry threshold or the quantities produced (and therefore the organizational structure), but increases the wages, prices, revenues and operating profits of all active firms.*

Proof. Let (α^*, w^*), be the solution under $\tilde{N}$. To show that $\left(\alpha^*, \left(\frac{\tilde{N}'}{\tilde{N}}\right)^{\frac{1}{\sigma-1}} w^*\right)$ is the solution under $\tilde{N}'$, and $q\left(\alpha\right)$ are the same under $\tilde{N}$ than under $\tilde{N}'$. Note that if (α^*, w^*) are the solution under $\tilde{N}$, then

$$\Delta^{ZCP}\left(\alpha^*, w^*, \tilde{N}\right) = q\left(\alpha^*\right)^{1-\frac{1}{\sigma}} \left(\alpha^* w^* \tilde{N}\right)^{\frac{1}{\sigma}} - w^* C\left(\alpha^*\right) - w^* f = 0$$

which implies

$$\frac{\left(\tilde{N}\right)^{\frac{1}{\sigma}}}{w^{*1-\frac{1}{\sigma}}} = \frac{C\left(\alpha^*\right) + f}{q\left(\alpha^*\right)^{1-\frac{1}{\sigma}} \left(\alpha^*\right)^{\frac{1}{\sigma}}},$$

and

$$\Delta^{FE}\left(\alpha^*, w^*, \tilde{N}\right) = \int_{\alpha^*}^\infty \left(q\left(\alpha\right)^{1-\frac{1}{\sigma}} \left(\alpha w^* \tilde{N}\right)^{\frac{1}{\sigma}} - w^* C\left(\alpha\right) - w^* f\right) g\left(\alpha\right) d\alpha - \delta w^* f^E = 0$$

which implies

$$\frac{\left(\tilde{N}\right)^{\frac{1}{\sigma}}}{w^{*1-\frac{1}{\sigma}}} = \frac{f\left[1 - G\left(\alpha^*\right)\right] + \delta f^E + \int_{\alpha^*}^\infty C\left(\alpha\right) g\left(\alpha\right) d\alpha}{\int_{\alpha^*}^\infty q\left(\alpha\right)^{1-\frac{1}{\sigma}} \left(\alpha\right)^{\frac{1}{\sigma}} g\left(\alpha\right) d\alpha}.$$

Now change $\tilde{N}$ to $\tilde{N}'$. Let $q\left(\alpha\right)$ be the same under $\tilde{N}'$ as under $\tilde{N}$. Set the same α^* and find w such that $\Delta^{ZCP}\left(\alpha^*, w, \tilde{N}'\right) = 0$ where $\tilde{N}' \neq \tilde{N}$. Note that $w^{*\prime} = \left(\frac{\tilde{N}'}{\tilde{N}}\right)^{\frac{1}{\sigma-1}} w^*$ solves the zero cut-off profit condition,

$$\begin{aligned} \Delta^{ZCP}\left(\alpha^*, w^{*\prime}, \tilde{N}'\right) &= \left(\frac{q\left(\alpha^*\right)^{1-\frac{1}{\sigma}} \left(\alpha^* w^{*\prime} \tilde{N}'\right)^{\frac{1}{\sigma}} P^{\frac{\sigma-1}{\sigma}}}{w^{*\prime}} - C\left(\alpha^*\right) - f\right) w^{*\prime} \\ &= \Delta^{ZCP}\left(\alpha^*, w^*, \tilde{N}\right) \frac{w^{*\prime}}{w^*} = 0 \end{aligned}$$

Note that this solution also solves the free entry condition,

$$\int_{\alpha^*}^\infty \pi\left(\alpha, w^{*\prime}\right) g\left(\alpha\right) d\alpha = \delta w^{*\prime} f^E,$$

$$\int\limits_{\alpha^*}^{\infty} \frac{\pi\left(\alpha, w^{*\prime}\right)}{w^{*\prime}} g\left(\alpha\right) d\alpha = \delta f^E + f\left(1 - G\left(\alpha^*\right)\right),$$

$$\int\limits_{\alpha^*}^{\infty} \left(q\left(\alpha\right)^{1-\frac{1}{\sigma}} \left(\alpha\right)^{\frac{1}{\sigma}} P^{\frac{\sigma-1}{\sigma}} \frac{\left(\tilde{N}'\right)^{\frac{1}{\sigma}}}{w^{*\prime 1 - \frac{1}{\sigma}}} - c\left(\alpha\right) \right) g\left(\alpha\right) d\alpha = \delta f^E + f\left(1 - G\left(\alpha^*\right)\right),$$

$$\frac{\left(\tilde{N}'\right)^{\frac{1}{\sigma}}}{w^{*\prime 1 - \frac{1}{\sigma}}} = \frac{\delta f^E + f\left(1 - G\left(\alpha^*\right)\right) + \int\limits_{\alpha^*}^{\infty} c\left(\alpha\right) g\left(\alpha\right) d\alpha}{\int\limits_{\alpha^*}^{\infty} q\left(\alpha\right)^{1-\frac{1}{\sigma}} \left(\alpha\right)^{\frac{1}{\sigma}} P^{\frac{\sigma-1}{\sigma}} g\left(\alpha\right) d\alpha} = \frac{\left(\tilde{N}\right)^{\frac{1}{\sigma}}}{w^{*1 - \frac{1}{\sigma}}}.$$

We still need to show that our guess does not change $q\left(\alpha\right)$. Use equation (9) to conclude that for any α

$$q\left(\alpha\right) = \alpha \left(w^*\right)^{1-\sigma} \tilde{N} \left(\frac{\sigma}{\sigma - 1} MC\left(q\left(\alpha\right); 1\right) \right)^{-\sigma} = \alpha \left(w^{*\prime}\right)^{1-\sigma} \tilde{N}' \left(\frac{\sigma}{\sigma - 1} MC\left(q\left(\alpha\right); 1\right) \right)^{-\sigma}.$$

Hence $q\left(\alpha\right)$ remains unchanged after the change in $\tilde{N}$. Equation (8) implies that given $q\left(\alpha\right)$, prices are proportional to wages. Then, an increase in the price with $\tilde{N}$, together with constant quantity, increases revenues. Since revenues change proportionally to the wage, and so do costs, operating profits change proportionally as well. ∎

Proposition 12 *In equilibrium trade liberalization increases welfare in both countries. Furthermore, the quantity produced of all non-exporters decreases and the quantity produced of all exporters increases.*

Proof. Consider first the case in which $f_{DF} = f_{FD} = 0$. Then, all firms export in equilibrium. In this case a trade liberalization is equivalent to an increase in the population size of a closed economy. Proposition 11 then implies that wages increase but the selection threshold and quantities produced remain unchanged. Now consider an economy in which $f_{DF} = f_{FD} > 0$ and some firms do not export. Since not all firms expand their market, w_D then increases by less than $\left(\left(\tau_{ij} \tilde{N}_j + \tilde{N}_i \right) / \tilde{N}_i \right)^{\frac{1}{\sigma - 1}}$ (the increase that would result if all firms were to export), since non-exporters are not selling to foreign firms and they are not demanding labor accordingly. As a result the effect of wages and population sizes that affect production according to equations (19), (20), and (21) goes from $\bar{w}_D^{1-\sigma} \tilde{N}_i$ in the closed economy to $w_D^{1-\sigma} \left(\tau_{ij} \tilde{N}_j + \tilde{N}_i \right)$ (where $\bar{w}_D$ denotes the wage in the closed economy). The new term is larger in the open economy, since wages adjust less than what Proposition 11 predicts. This implies that the total production of exporters has to necessarily increase. Non-exporters do not get access to the foreign market so the relevant effect on sales is necessarily decreasing in w_D by equation (18). Toward a contradiction, suppose that opening to trade leads to a decline in w_D. Then, by the reasoning above, production of non-exporters and exporters increases, which by Proposition 2 increases their labor demands since by Proposition 4 the cost function is homogeneous of degree one. This contradicts the labor market clearing condition in (22). Hence, w_D strictly increases, and the production of non-exporters decreases and that of exporters increases with trade liberalization. Of course, since our choice of a domestic country is arbitrary, this applies to the other country as well. ∎