SIEPR Discussion Paper No. 02-38

Organization and Inequality in a Knowledge Economy

By
Luis Garicano
University of Chicago and CEPR
And
Esteban Rossi - Hansberg
Stanford University

June 2003

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

Organization and Inequality in a Knowledge Economy*

Luis Garicano† — University of Chicago and CEPR

Esteban Rossi-Hansberg‡ — Stanford University

First Draft: October 2002
This Draft: June 2003

Abstract:

How does information technology affect wages and organizations? To address this question, we model an economy formed by a continuum of agents, with different cognitive skills, who may produce on their own or may join in organizations with other agents to better utilize their knowledge. Our model generates an assignment of workers to positions, a wage structure, and a universe of knowledge-based hierarchies with different allocations of tasks, spans of control, and number of layers. We show that commonly observed phenomena such as positive sorting by ability, an increasing relationship between rank and cognitive ability, an increasing and convex wage schedule, and a positive relation between wages and firm size are natural consequences of the organization of knowledge in the economy. We then use our model to study the impact of information technology on the labor market and on the structure of firms. We show that the evolution of wage inequality and firm size is consistent with decreases in the cost of accessing information in the 80's and early 90's and decreases in the cost of communicating information in the late 90's. Our theory is also consistent with the evidence on decentralization, flatter hierarchies, and larger spans of control.

JEL Classification Code: D2, J3, L2

---

*We have benefitted from discussions with Daron Acemoglu, Fernando Alvarez, Gary S. Becker, Jacques Crémer, Daniel Ferreira, Bob Gibbons, Erik Brynjolfsson, Pierre-André Chiappori, Wouter Dessein, Mathias Dewatripont, Doug Diamond, Robert Gertner, Robert Hall, Narayana Kocherlakota, Dirk Krueger, Felix Kubler, Kevin M. Murphy, John Pencavel, Canice Prendergast, Michael Raith, Raghu Rajan, Luis Rayo, Tano Santos, Jean Tirole, Ed Vytlacil, Ivan Werning and seminar participants at Stanford, Chicago, Toulouse and the CEPR seminar on Organizational Economics. We thank David Card and Kevin J. Murphy for providing us with data on Inequality and on CEO pay respectively. Garicano is particularly grateful to his late advisor Sherwin Rosen, on whose work this paper builds.

†luis.garicano@gsb.uchicago.edu. http://gsb.uchicago.edu/fac/luis.garicano. (773)8342862. 1101 E. 58th Street, Chicago, IL, 60637, USA.

‡erossi@stanford.edu. http://www.stanford.edu/~erossi. (650)7241427. 579 Serra Mall, Stanford, CA, 94305-6072, USA.

1 Introduction

Knowledge is becoming cheaper to process, store and transmit. If the economic problem of society is, as Hayek (1945) argued, to use optimally the available knowledge, then such changes in the ability of individuals to process and transmit knowledge must impact both the organization of work and the rewards associated with knowledge-related skills. This paper seeks to understand this impact by studying an equilibrium model of organizations formed by heterogeneous agents who can acquire and transmit the knowledge required for production. We analyze the reward structure and knowledge organization associated with this goal.

A large recent literature has studied the internal organization of hierarchies.[1] Whether the specific papers are concerned with how hierarchies process information, monitor performance, allocate resources, or facilitate knowledge acquisition, they all solve the firm optimization problem without embedding it in an equilibrium framework. In particular, previous work has mostly ignored the interactions between the hierarchy's design problem and the labor market.[2] And yet changes in the internal organization of hierarchies (e.g. task assignment, spans of control, firm size, number of layers of management) and in the labor markets (e.g. wage structure and assignment of workers to jobs) are simultaneously determined in equilibrium. Wages of individuals respond to the tasks they have been assigned, and these are in turn a consequence of their position in the hierarchy. Conversely, an individual's position in the hierarchy, and the hierarchy design, depends on the wages that individuals of different talents command. This paper presents an equilibrium model of hierarchies that integrates the internal hierarchical structure and the labor market outcomes, and studies the interdependencies between them.

Our starting point is the theory of organizations in Garicano (2000). He shows, in a model with ex-ante identical agents, that a 'knowledge-based hierarchy' is the efficient way to organize the acquisition of knowledge required for production when matching problems and solutions is costly. Agents either learn how to perform the most common tasks and

[1] Four categories of models of hierarchies, all of them exclusively concerned with the firm's optimization problem, have been proposed. A first one deals with information processing under bounded rationality, where agents in hierarchies are information processors (Radner, 1992, and Radner and van Zandt, 1992, Bolton and Dewatripont, 1994) or resource allocators (Crémer, 1980, Geanakoplos and Milgrom, 1991). A second one focuses on monitoring as the main role of managers (Calvo and Weillisz, 1978, 1979, and Qian, 1994). A third strand (Rajan and Zingales, 2001) studies the role of hierarchies in administering access to core resources. Finally, Garicano (2000) studies hierarchies' role in knowledge acquisition.

[2] The only exception is Rosen (1982), which sets up, but does not characterize, an equilibrium model of hierarchies with multiple layers where workers have heterogeneous skills.

become production workers or specialize in problem solving and transmit their knowledge about exceptional and/or difficult problems to production workers as needed. The role of the organization is to increase the utilization rate of the knowledge of experts by shielding them from problems that less knowledgeable workers can solve equally well.

In light of our position on the centrality of knowledge, this paper puts agent heterogeneity in cognitive skills squarely at the core of the analysis. In our model, agents with different cognitive skill must acquire knowledge and choose the job they want to perform (production or management/problem-solving at one of several possible layers). Workers who choose management jobs must choose the size of their teams and the tasks that team members must learn. The economy-wide problem is an assignment problem in which workers must be matched to each other and assigned to a particular layer of one of many possible hierarchical structures. We solve this problem and obtain the earnings schedule that supports this assignment of agents to hierarchies.[3]

The equilibrium displays three empirically interesting features that are worth anticipating here. First, it displays positive sorting, in the sense that higher ability agents share their knowledge with higher ability subordinates (production workers or lower level managers). The reason is that the expertise of highly skilled managers must be shielded from easy questions. As a result, more able agents have larger teams (since team members are smarter, they require less guidance) and are likely to be in higher layers, further leveraging their talent.

Second, individuals in the economy are segmented by cognitive skill in the sense that the original continuous agent skill set will be partitioned in layers in a systematic way. In particular, the lowest skill workers are production workers, the next workers are self-employed entrepreneurs, the next are first level problem solvers in larger hierarchies, the next are entrepreneurs in two layer hierarchies, etc. Thus a unidimensional skill set, and a common technology available to all workers, generate an economy formed by a universe of firms with different number of layers, different sizes, and different skill sets.

Third, the model is consistent with, and provides a rationale for, a key unexplained stylized fact in the empirical literature on firms and wages: the large wage-size effect, whereby larger firms pay larger wages, even after accounting for differences in working conditions, union status, etc. (Brown and Medoff, 1989). We show that this is a natural

[3]Our focus on individual skill heterogeneity as the basis for organizational heterogeneity is consistent with recent empirical work on the structure of wages. In particular, recent work using matched firm-worker data, Abowd, Kramarz and Margolis (1999) find that 'virtually all of the interindustry wage differentials is accounted for by variation in average individual heterogeneity between sectors.'

consequence of positive sorting: in equilibrium, better agents are matched with better subordinates, who ask them questions less often. This allows better managers to have larger teams. This effect is amplified as the most skilled agents are assigned to higher ranked management roles. More layers and larger teams imply larger firms. This matching explains why the wage-size puzzle persists in the data even after controlling for worker fixed effects. A given worker will earn more in a large firm, as he will be matched with higher quality workers.

We then proceed to use the model to shed some light on the impact of information technology (IT) on wages and organization. The analysis distinguishes two aspects of IT: its ability to improve access to stored knowledge through the increase in raw processing power and its impact on the cost of communication among agents.[4] Cheaper access to knowledge, for example in the form of cheaper access to databases, increases information available at a given cost. Communication technology, on the other hand, facilitates communication by agents of their knowledge.

The analysis of the impact of these technological changes takes place in two steps. First, we analyze the partial equilibrium comparative statics of the firm's choices. This implies holding wages, assignments, and the number of layers constant. Then we study the general equilibrium problem in which wages, assignment, and the internal organizational structure are determined simultaneously.

Our results show that IT improvements have different impact on wage inequality and organization, depending on whether they mainly reduce the cost of communication or the cost of accessing knowledge. Decreases in the cost of communication lead teams to rely more on managers, reducing the decentralization of the organization. As a result, wage inequality among workers decreases, but wage inequality among top managers, and between them and workers, increases. Spans of control increase, as communication is cheaper, and the number of layers of managers increases. Intuitively, improvements in communication technology generate a type of 'superstar' (Rosen (1981)) effect, whereby each top level manager can better leverage their knowledge, and efficiently combine their high level knowledge with the knowledge of their employees.

In contrast, the effect of reductions in the cost of accessing knowledge, due to a decrease in the cost of processing power, is unambiguously to increase the number of tasks performed

[4] There may be other relevant aspects of information technology that we are ignoring here, such as its ability to improve labeling of categories or to improve the monitoring of agents work. This model concentrates on the two aspects above, which we consider of the greatest importance.

at all organizational layers, even if wages and the marginal price of skills go up. Thus this technological change results in increasing decentralization. As a consequence, wage inequality within a given (managerial or productive) layer always increases. Moreover, it results in increases in the spans of managers, who may deal with more subordinates as each of them posses less problems. This reduction has an ambiguous effect on the number of layers. If communication costs are reasonably high, the improvement in access to knowledge will lead to fewer layers, while if communication costs are low, lower cost of information will lead to more layers. The reason is that improvements in the access to knowledge decrease the value of leveraging knowledge, but also the cost of extra layers since knowledge is cumulative.

Empirical research has identified a broad pattern of changes in the labor market and in the organization of firms over the last two decades that we seek to address with our theory. Several studies have documented the substantial increase in wage inequality in the 80's and early 90's which has been found to be mostly due to an increase in the demand for skill,[5] and, more specifically, for cognitive skills,[6] and which appears to be correlated with the increasing use of information technology.[7] Concerning the internal structure of organizations, recent work has documented, for the same period, an increase in the autonomy and responsibility of workers, a reduction of the number of layers of management, and an increase in the managerial span of control. Again, these changes have also been associated with the use of information technology.[8] A different pattern is observed for the second part

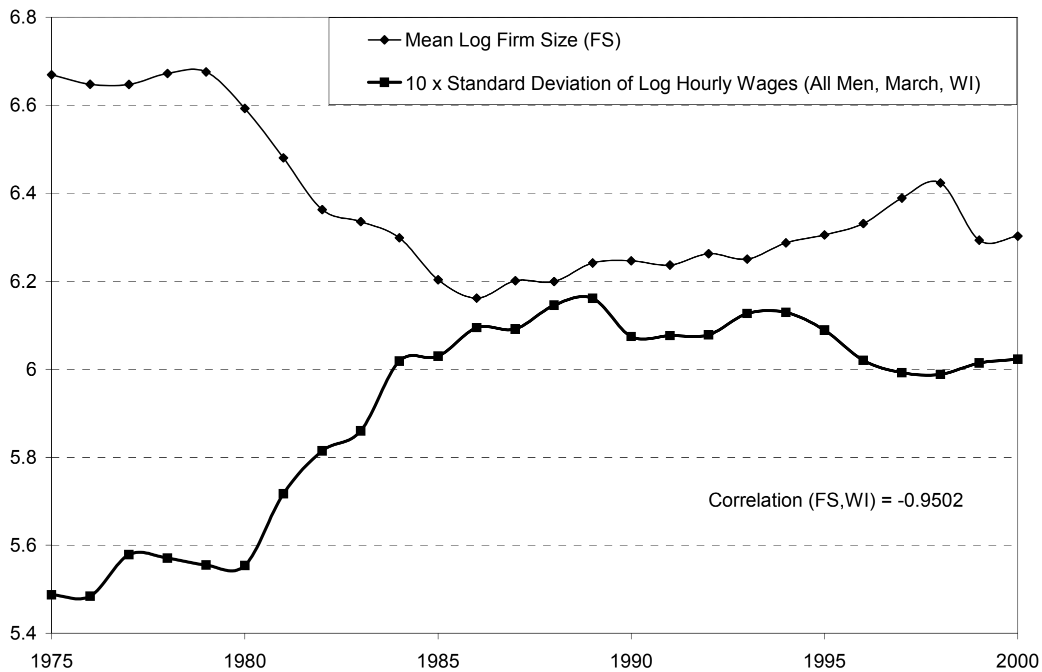
[5]Katz and Murphy (1992) are the first paper to show that increases in inequality are consistent with skill biased technological change. Later, Autor, Katz, and Krueger (1998) and Dunne, Haltiwanger, and Troske (1996) have shown that the composition of the labor force within industry and establishments continues to shift towards the more educated workers and more skilled occupations, in spite of raises in returns to skills. Evidence that a similar patter is seen in other countries is provided by Berman, Bound, and Machin (1998).

[6]Murnane, Willet and Levy, (1995) find a higher correlation between earnings and test scores for a more recent panel of graduates than for an earlier one. This finding is particularly clear for high test scores in math.

[7]It was Krueger (1993) who first documented a substantial premium associated with computer use, of up to half of the growth in the education premium since the eighties. This finding was confirmed later by Autor, Katz, and Krueger (1997). Brynjolffson and Hitt (1997) and Lehr and Lichtenberg (1999) have found that larger computer purchases and skill are complementary.

[8]Bresnahan, Brynjolfsson and Hitt (2002) find, using firm-level data, that greater use of computers is associated with the employment of more-educated workers, greater investments in training, broader job responsibilities for line workers, and more decentralized decision-making. Caroli and Reenen (2001) find evidence of organizational change complementary with increases in demand for skills. In particular, they find evidence of decentralization of authority and a widening of the range of tasks performed by workers. Rajan and Wulf (2002), in a recent paper, present evidence that from 1986 to 1995 firms have become flatter, with less layers of management and that managerial span of control has increased. The evidence from 1995 to 1999 is weaker but suggests, as in our theory, that the flattening stopped and managerial span kept increasing.

of the 90's. The evidence shows that a clear slowdown, or even a drop depending on the measure, in the increase in wage inequality has taken place in the late 90's. The theory provides empirical implications on decentralization, numbers of layers and spans of control for the second part of the 90's, that may guide future empirical work.



Source: Mean Log Firm Size :Compustat, SD of Log Hourly Wages: March CPS using the methodology in Card and DiNardo (2002).

Figure 1: Firm Size and Wage Inequality

Figure 1 illustrates these recent changes in organization and inequality. It presents the standard deviation of hourly wages together with the mean log firm size from 1976 to 2000 in the US. The correlation between the two series is -0.95.[9] This remarkable negative correlation supports our view that there is a common source of variation driving changes in both of them. Our interpretation is that the underlying source of variation are the costs of acquiring and communicating information. Our model is able to qualitatively reproduce these changes with exogenous improvements in the cost of accessing knowledge, mostly concentrated in the 80's and early 90's but present throughout the period, followed by improvements in communication technology (e.g. e-mail, cell phones, and wireless technology)

[9] The correlation of first differences is -0.48.

in the late 90's. Some evidence of these improvements in information technology, and the particular timing described, is presented in Figure 2.[10]

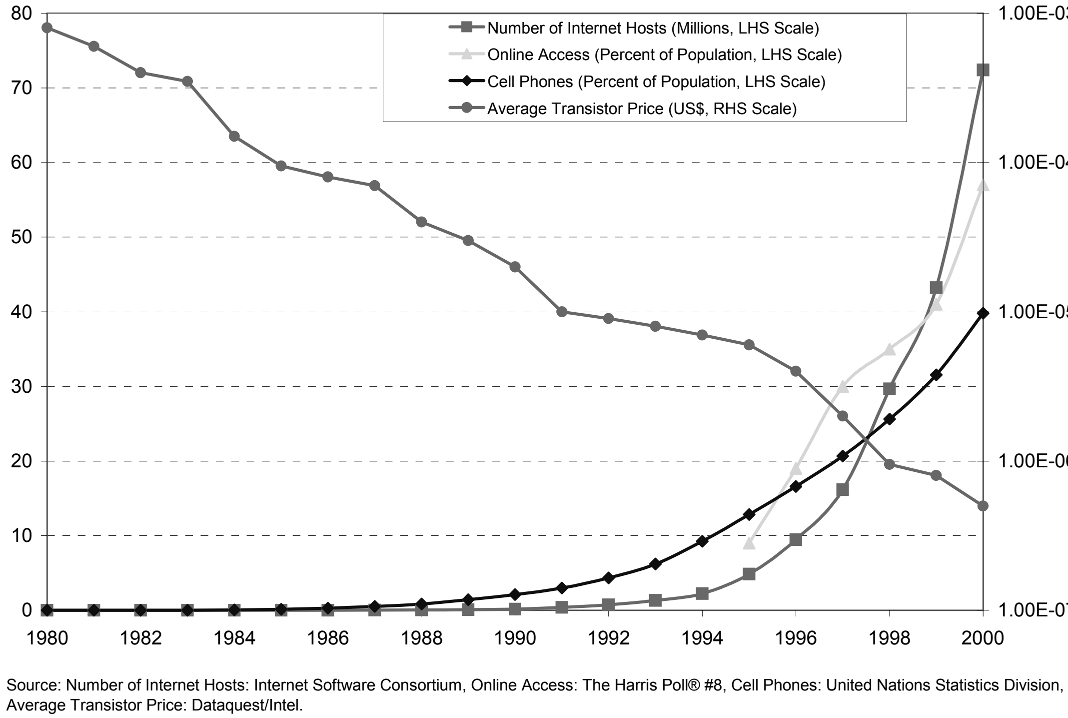
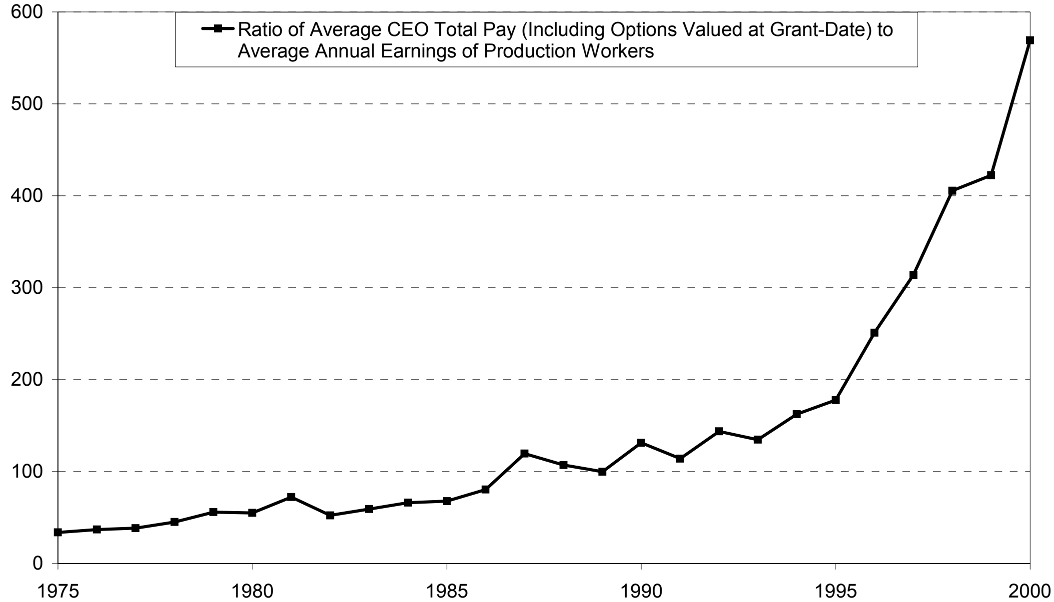


Figure 2: Changes in Information and Access to Information Technology

Even though our theory is consistent with the changes in total wage inequality illustrated in Figure 1, it implies that the earnings of top entrepreneurs, and managers, relative to those of workers should have increased throughout the 80's and all the 90's. That is, improvements in both types of technology lead to a higher CEO premium. We present evidence of this fact in Figure 3. It shows the evolution of the average CEO total pay as a ratio of average worker pay from 1980 to 2000.[11]

[10] We are interpreting internet access, and the number of internet hosts, to result most importantly in improvements in communication technology given the importance of E-mail. However, cheaper access to the internet involves also improvements in the technology to access information.

[11] In spite of the fact that this theory provides a consistent explanation for the data, it does not pretend to explain all the variation in the different variables we incorporate in the analysis. However, our theory does provide a structural framework that could be potentially used to test different hypothesis of the effect of information technology. This type of test is, nevertheless, beyond the scope of this paper.



Source: CEO sample is based on all CEOs included in the S&P 500, using data from Forbes and ExecuComp. CEO total pay includes cash pay, restricted stock, payouts from long-term pay programs, and the value of stock options granted using ExecuComp's modified Black-Scholes approach. (Total pay prior to 1978 excludes option grants, while total pay between 1978 and 1991 is computed using the amounts realized from exercising stock options during the year, rather than grant-date values.) Worker pay represents 52 times the average weekly hours of production workers multiplied by the average hourly earnings, based on data from the Current Employment Statistics, Bureau of Labor Statistics. We thank Kevin Murphy for this data.

Figure 3: Worker-Manager Wage Inequality

An important qualification is that only the boundary of the hierarchy, and not the boundary of the firm, is determined in our model. These boundaries may or may not coincide.[12] In fact, we present a decentralization of the model where all agents are self-employed either as consultants or workers, and the transactions of knowledge ('consulting services') take place in the market. In this case the size of the hierarchy *is* different than the size of the firm. We view this ability of the model to predict transactions of knowledge in hierarchies within firms or within consulting service markets as a richness of the model, since this duality is also a feature of the real world. Although we choose not to settle formally this indeterminacy, previous literature has advanced several reasons for transactions to take place within firms rather than in the market that could potentially be incorporated in our framework.[13]

[12] The data in Figure 1 is on firms, and not on hierarchies. As long as the determinants of the boundaries of the firm are not affected significantly by changes in information technology, changes in the size of hierarchies result in corresponding changes in the size of firms.

[13] First, knowledge may be proprietary, and fear of expropriation may limit the access to markets (Rajan and Zingales, 2001). Second, knowledge may be specific to the particular production process, and so in

A recent theoretical literature has studied the impact of technology on the labor market, motivated by the changes in wage inequality illustrated in Figure 1. Within this literature, Saint Paul (2000) has focused on the role of knowledge on the structure of earnings, studying team formation and selection when there are spillovers between worker's skills.[14] Other theoretical contributions studying the link between technology and the structure of earnings have clarified issues such as the possibility of reverse causation from skills to technology (Acemoglu, 1998), the impact on within wage inequality of within group ability distinctions (Galor and Moav, 2000), the role of worker heterogeneity in the ability to implement new technologies (Galor and Tsiddon, 1997), or the impact of market size on specialization (Mobius, 2000). None of these papers delivers implications for organizational structure or for the occupational distribution of the labor force.

Thus none of the existing analysis of the impact of technology on the labor market contains a model of hierarchies permitting the analysis of the internal, within firm, implications of technological change. To achieve this, a model of the hierarchical structure of the firm needs to be embedded in an equilibrium model. Lucas (1978) presents the first equilibrium model of occupational choice, although without hierarchies. Only Rosen (1982) develops an equilibrium model of hierarchies. His hierarchical model, which assumes a generic multiplicative technology between managerial and worker skill, is, however, not suitable for the study of the impact of IT on the organization of production and the wage structure. The challenge then for the researcher trying to build an equilibrium theory of wages and positions is to embed a full fledged model of a hierarchy in an equilibrium framework. We choose to do this using Garicano (2000) as it is the only previous paper that explicitly models both knowledge access and communication.

The paper is structured as follows. Section 2 presents the model, starting with the two layer case to facilitate its comprehension and then building on to the multiple layer case. Section 3 analyzes the partial equilibrium effect of technology (i.e. with wages, layers and assignments held constant) on organization and demand for skills. Section 4 constructs the equilibrium when wages, assignments and layers adjust, shows that it involves

a world of incomplete contracts hold-up problems may require the integration of the hierarchy in a firm (Grossman and Hart, 1986, and Hart and Moore, 1990). Finally, knowledge transactions may require some specialized language or code that has been developed to deal with contingencies that occur in a particular business (Crémer, Garicano and Prat, 2003). In our model, we choose not to make any of these possible assumptions, and leave the two possible interpretations of the knowledge exchanges open throughout.

[14] He considers the role of information technology as increasing the size of the population over which a worker's ideas can be spread. Then, inequality may actually fall with improvments in IT, as creativity has also positive effects over production workers.

positive sorting, proves its existence, and characterizes the equilibrium. It also shows that an alternative, decentralized, version of the economy, where individuals shop for consulting services, is equivalent to the firm formulation. Finally, Section 5 studies the equilibrium impact of IT.

2 The Model

We study a production process which requires time and knowledge of the tasks that must be performed. Agents of heterogeneous ability learn to solve tasks and choose an occupation and a team to join. The available occupations are different layers of problem solving and production. All agents supply a unit of labor. The equilibrium allocation in the economy determines the wages for all agents, the tasks they perform, and the composition and structure of teams.

2.1 Production and Knowledge

A task is drawn per unit of time per worker. Production takes place when either the worker knows how to perform the task, or when she can ask someone else who knows. In the second case, workers incur communication or 'helping' time cost h.[15] Some tasks are known a priori to be more common than others. We rank the tasks by the likelihood they will be confronted, so that the Z task is associated with density $f(Z)$, where $f(Z)$ is decreasing.

A worker i can learn to perform tasks at a cost c^i. Workers learn the most frequent tasks first, and then move on to the less common ones, so that knowledge of more knowledgeable workers always encompasses the knowledge of the less knowledgeable ones. That is, knowledge is cumulative. We say that a worker has acquired knowledge z^i when she has learned to perform all tasks in $(0, z^i)$.

The analysis in Garicano (2000) shows that, when matching problems and solutions is hard, the optimal organization of the acquisition of knowledge involves less knowledgeable workers learning about the most common problems and spending all their time in production, and 'managers' specializing in solving problems and learning about increasingly more exceptional problems. This organization has a pyramidal structure, with each layer having a smaller number of workers than the previous layer.

[15] Similarly to the information processing literature (e.g. Radner and Van Zandt (1992), Bolton and Dewatripont (1994)we aggregate all communication costs on the worker who is asked to solve the problem. With Garicano (2000) we assume this cost is incurred regardless of whether the worker asked for help knows the answer, since she must figure out anyway if she might know the answer and communicate with the worker who asked.

Following this analysis, we take as our starting point an economy formed by individuals of heterogeneous cognitive skill that may join in teams (composed of managers and production workers). Managers are specialized in acquiring knowledge about tasks and in problem solving, and workers specialize in production and acquire knowledge only about the most common tasks. In this subsection we restrict our attention, for simplicity, to two layer organizations. After we explain the fundamental features of the production process we will generalize the analysis to an arbitrary number of layers. The number of layers in the economy will be determined in equilibrium.

All agents are endowed with one unit of time. Output produced by a team with n production workers, who spend each 1 unit of time in production and acquire knowledge about tasks z^p at a unit learning cost c^p, and a manager who acquires knowledge $z^m > z^p$ at a cost c^m as:

$$y = F(z^m)n - c^p n z^p - c^m z^m. \tag{1}$$

Managers spend a fraction of time h helping each one of n workers in the team whenever they cannot perform a task, which happens with probability $1 - F(z^p)$. Workers do not spend time asking questions. The constraint is then:[16]

$$hn(1 - F(z^p)) = 1.$$

Managers spend time answering questions even if they do not know the answer. The time constraint implies that the span of the manager is limited by the knowledge of production workers; if production workers acquire more knowledge, they will require help less often, and managers will be able to supervise larger teams.

2.2 Occupational Choice and the Skills Supply

Agents are heterogeneous in their cognitive ability. But ability does not have a natural scale. We define (and scale) from now on ability as follows. Measure the cost for each i worker of learning to perform an interval of tasks of unit length $z = 1$ in units of foregone output, and call this cost c^i. Call the ability of the worker with the highest measured cost $\alpha = 1$ and the ability of the worker with the lowest cost $\alpha = 0$. We can then draw a straight line between these points with intercepts t and slope k, and assign ability α^i from

[16] This constraint is at equality since the manager specializes in solving problems. Note that we ignore throughout integer constraints, and we solve the problem deterministically, i.e we ignore the possibility that two problems come up at the same time.

this line to the rest of the workers depending on their cost of acquiring knowledge. This method results, without loss of generality, in an implicit definition of ability in this world so that a doubling in the ability of a worker results in a proportionate reduction in the cost of learning a given interval of problems. We then write the cost function as a function of ability:

$$c(\alpha; t) = t - k\alpha. \tag{2}$$

There are two technological parameters in this function: t and k. Both of them represent the state of information technology, i.e. a technology that decreases the cost of learning to find the solution to a given problem. In what follows, we do not assume a priori that technological change favors more skilled workers, and center the analysis on changes in t.

We can induce, following the procedure above, from the exogenous distribution of learning costs in the population, a distribution of cognitive abilities with support $[0, 1]$. We assume that this distribution can be described by a continuous density function, $\alpha \sim \phi(\alpha)$. Finally, we assume that no substantial heterogeneity exists in communication skills among workers.[17]

Agents are income maximizers. The agent's problem is to choose her occupation to maximize income, given the earnings schedule. Namely, the problem of an agent endowed with skill α is given by:

$$U(\alpha) = \max\{\Pi(\alpha), w(\alpha)\} \tag{3}$$

The second term in (3), $w(\alpha)$, is the market wage for an agent of skill α hired as a worker. The first term $\Pi(\alpha)$ is the managerial income earned by the agent if she chooses to go into management. The calculation of this income requires solving the manager's problem, obtaining the market wage for all workers, and finding out which workers are assigned to him in equilibrium. This is the object of the next sections.

2.3 Managers and the Demand for Skills

Consider now the problem of a worker of skill α^m who chooses to go into management and hires a team of n production workers of skill α^p. The manager chooses the number of workers hired, the interval of tasks he learns, and the production worker's knowledge, given

[17] Although clearly differences in the ability of workers to communicate may be relevant, most of the empirical literature on skill differentials has found cognitive skills to be a very good predictor of wage slopes (see Murnane, Willett and Levy, 1995).

wages, to maximize his residual (after labor costs) rent:[18]

$$R(\alpha^m, \alpha^p, w) = \max_{z^m, z^p, n} F(z^m)n - c(\alpha^p; t)nz^p - c(\alpha^m; t)z^m - wn \tag{4}$$

subject to the time constraint of the manager:

$$hn(1 - F(z^p)) = 1. \tag{5}$$

It will simplify the discussion to define the proportion of tasks that a worker or manager can perform as $Q^i = F(z^i)$. Let then $z^i = g(Q^i)$, i.e. $g = F^{-1}$ which implies, $g' > 0, g'' > 0$. Replacing the constraint into the objective of the manager (4), and using this new notation, the problem becomes:

$$R(\alpha^m, \alpha^p, w) = \max_{Q^m, n} Q^m n - c(\alpha^p; t)g(1 - \frac{1}{hn})n - c(\alpha^m; t)g(Q^m) - wn \tag{6}$$

2.4 The model with multiple layers

We want to use the model to give us predictions about the number of layers in a firm, and to analyze how the structure of firms affects wages and earnings. For this purpose, we now generalize the production process introduced in the previous subsections to allow agents to join into team with multiple layers of managers. Agents will choose between becoming entrepreneurs (top level managers), managers of intermediate layers, or workers. We analyze the problem of an entrepreneur that maximizes rents, given wages and assignments, and can hire different layers of managers and workers. Consider the problem of an entrepreneur of layer $\ell + 1$ with skill $\alpha^{\ell+1}$ managing a firm of $\ell + 1$ layers, with managers of ability α^l ($l = 1, ..., \ell$) and workers of ability α^0, given wages w^l ($l = 0, ..., \ell$). The entrepreneur chooses the number of managers/workers at all layers and their knowledge to solve the following problem:

$$\max_{\{Q^l\}_{l=0}^{\ell+1}, \{n^l\}_{l=0}^{\ell}} Q^{\ell+1}n^0 - c(\alpha^{\ell+1}; t)g(Q^{\ell+1}) - \sum_{l=0}^{\ell} \left[c(\alpha^l; t)n^l g(1 - \frac{n^{l+1}}{hn^0}) + w^l n^l \right] \tag{7}$$

[18] We write here R rather than Π to emphasize that he is solving the problems for a given arbitrary assignment and wage level. Thus from the perspective of a manager matched with a set of workers, the wage is a given number. There is no loss of generality, in assuming a manager is matched with a given worker quality as the skill functions are continuous: a given measure of managers can be matched with any measure of workers.

subject to

$$\begin{aligned} hn^0(1-Q^\ell) &= n^{\ell+1}=1 \\ hn^0(1-Q^{\ell-1}) &= n^\ell \\ &\vdots \\ hn^0(1-Q^0) &= n^1, \end{aligned}$$

and

$$1 \geq Q^{\ell+1} \geq Q^\ell \geq ... \geq Q^0 \geq 0.$$

The first set of constraints are time constraints for the different layers of managers and the top level entrepreneurs. The last constraint says that higher layer managers must have more knowledge than lower layer managers. This is a convention in our notation. In the rest of this section we will transforms this problem to rewrite it in a convenient recursive form.

The problem above can be broken down in two parts. The first part is the problem of the entrepreneur, given that all lower layers of management behave optimally. The second part, that we discuss next, is the problem of managers. Managers of layer l maximize the value per worker of an l-layer branch of the firm given their own knowledge, wages (including their wage) and assignments. Let $\vec{\alpha}^\ell \equiv (\alpha^\ell, ..., \alpha^0)$ and $\vec{w}^\ell \equiv (w^\ell, ..., w^0)$, then we can denote the value per worker of an $\ell-$layer branch recursively by

$$\begin{aligned} & p^\ell(Q^\ell; \vec{\alpha}^\ell, \vec{w}^\ell) \qquad (8) \\ = & \max_{Q^{\ell-1}} \left[Q^\ell - Q^{\ell-1}\right] - \left[c(\alpha^\ell;t)g(Q^\ell) + w^\ell\right] h(1-Q^{\ell-1}) + p^{\ell-1}(Q^{\ell-1}; \vec{\alpha}^{\ell-1}, \vec{w}^{\ell-1}), \end{aligned}$$

subject to

$$Q^\ell \geq Q^{\ell-1} \geq 0$$

for $\ell > 1$ and

$$p^0(Q^0; \alpha^0, w^0) = Q^0 - c(\alpha^0;t)g(Q^0) - w^0$$

for $\ell = 0$. That is, the top manager of the branch is choosing the knowledge (and therefore the number of managers per worker) of all managers and workers in the branch, given her knowledge, wages, and assignments. Or, equivalently, he is choosing the knowledge of the manager/worker one layer below him and he is letting all lower level managers choose the knowledge of their subordinates. Notice that the number of workers in the firm is not chosen by intermediate managers, they *only* choose the number of lower layer managers per worker by deciding the knowledge of their immediate subordinates.

The first order condition for the problem above is given by

$$h\left[c(\alpha^\ell;t)g(Q^\ell)+w^\ell\right]+\frac{\partial p^{\ell-1}(Q_M^{\ell-1};\cdot)}{\partial Q^{\ell-1}}\gtreqless 1, \tag{9}$$

with equality iff $Q^\ell \geq Q_M^{\ell-1} \geq 0$. The first weak inequality ($\geq$) holds if $Q^\ell = Q_M^{\ell-1}$, the second ($\leq$) if $Q_M^{\ell-1}=0$. The first order condition equates the marginal cost and marginal benefit of increasing the knowledge of the lower layer management. The marginal cost is the decrease in the contribution to output of the branch top manager. The marginal benefits is the decrease in the knowledge and wage costs (monetary costs of time) spent on solving problems at the top layer of the branch, and the increase in the value of layer $\ell-1$ branches.

The derivative of the value of the branch with respect to the knowledge of its highest layer manager (the envelope condition) is given by

$$\frac{\partial p^\ell(Q^\ell;\cdot)}{\partial Q^\ell}\leq 1-c(\alpha^\ell;t)g'(Q^\ell)h(1-Q_M^{\ell-1}), \tag{10}$$

for an internal solution ($Q^\ell > Q_M^{\ell-1}$). That is, the change in the value per worker of a branch when we change the knowledge of its top manager has two components. On one hand, the gain in branch's output resulting from the ability of its top managers to solve more problems. On the other, the increase in the learning cost per worker of the top manager.

Combining, the first order and envelop condition we obtain,

$$c(\alpha^\ell;t)g(Q^\ell)+w^\ell-c(\alpha^{\ell-1};t)g'(Q_M^{\ell-1})(1-Q_M^{\ell-2})\gtreqless 0, \tag{11}$$

with equality iff $Q^\ell \geq Q_M^{\ell-1} > Q_M^{\ell-2}$. The first weak inequality ($\geq$) holds if $Q^\ell = Q_M^{\ell-1}$, the second ($\leq$) if $Q_M^{\ell-1}=Q_M^{\ell-2}$. That is, managers choose the knowledge of lower level managers so as to balance the monetary gains of their savings in time with the learning costs of lower management. The equations above implicitly define a function $Q_M^{\ell-1}(Q^\ell,p^\ell)$. It is easy to show that this function is strictly increasing in the ability of the manager of layer ℓ (see Proposition 2).

The original problem in (7) can then be rewritten as

$$\begin{aligned}&R^{\ell+1}(\alpha^{\ell+1};p^\ell(\cdot;\vec{\alpha}^\ell,\vec{w}^\ell))\\ =\;&\max_{1\geq Q^{\ell+1}\geq Q^\ell\geq 0}\left[Q^{\ell+1}-Q^\ell+p^\ell(Q^\ell;\vec{\alpha}^\ell,\vec{w}^\ell)\right]\frac{1}{h(1-Q^\ell)}-c(\alpha^{\ell+1};t)g(Q^{\ell+1}),\end{aligned} \tag{12}$$

for $\ell \geq 0$ and

$$R^0(\alpha^0)=\max_{Q^0} Q^0-c(\alpha^0;t)g(Q^0).$$

Summarizing, we can interpret the multilayer problem as follows: An entrepreneur of layer $\ell+1$ with skill $\alpha^{\ell+1}$ hires managers of layer ℓ with knowledge Q^ℓ at wage w^ℓ per managed worker. Entrepreneurs decide on their own knowledge and the knowledge of managers/workers immediately below them. The managers of the ℓth layer in turn manage the remaining ℓ-layers of the firm given their knowledge Q. These ℓ-layers are worth $p^\ell(Q;\vec{\alpha}^\ell,\vec{w}^\ell)$ per worker managed for the firm.

The first order conditions of problem (12) are given by

$$\frac{1}{h(1-Q_E^\ell)} - c(\alpha^{\ell+1};t)g'(Q_E^{\ell+1}) \gtreqless 0, \tag{13}$$

and

$$\frac{\left[Q_E^{\ell+1} - Q_E^\ell + p^\ell(Q_E^\ell;\cdot)\right]}{1-Q_E^\ell} + \frac{\partial p^\ell(Q_E^\ell;\cdot)}{\partial Q^\ell} - 1 \lesseqgtr 0, \tag{14}$$

where both equations hold with equality iff $1 \geq Q_E^{\ell+1} \geq Q_E^\ell \geq 0$. The first weak inequality ($\geq$) in equation (13) holds if $1 = Q_E^{\ell+1}$, the second ($\leq$) if $Q_E^{\ell+1} \geq Q_E^\ell$. In equation (14), the first weak inequality ($\leq$) in equation (13) holds if $Q_E^{\ell+1} = Q_E^\ell$, the second ($\geq$) if $Q_E^\ell \geq 0$. The interpretation of (14) is similar to the one of condition (9), entrepreneurs choose the knowledge of lower layer management to balance the loss in output associated with problems that they solve, with the gains in the value of branches of the firm and the monetary costs of the time they save. One could combine (14) with the envelope condition in (10) to obtain a condition that equates the gain associated with the time saved by the entrepreneur and the marginal learning cost of increasing the knowledge of lower management. Condition (13) equates the marginal gain in output of increasing the entrepreneur knowledge with the marginal costs per worker of increasing knowledge. The system of first order conditions implicitly defines two functions, $Q_E^{\ell+1}(\alpha^{\ell+1};p)$ and $Q_E^\ell(\alpha^{\ell+1};p)$.

In this section we have shown that the problem of an entrepreneur can be solved sequentially. First layers managers choose the knowledge of workers, second layer managers choose the knowledge of first layer managers, etc. Entrepreneurs choose the knowledge of the managers one layer below and their own knowledge, thereby implicitly choosing the number of workers in the firm. Although the original problem and the sequential problem are equivalent, the second formulation simplifies considerably the analysis of the model.

3 Comparative Statics

We characterize here the partial equilibrium impact of access to information and communication technology on knowledge, task assignment, and managerial spans. We identify an organization as more decentralized if managers are needed 'less often' in the production process.

The following proposition shows that the knowledge of entrepreneur and next layer managers/workers increases with improvements in communication and access to information technology. Consider first the effect of an increase in the information at the disposal of agents decreasing the cost of learning to perform a task. Because of the decrease in costs, entrepreneurs and managers learn more. Intuitively, the organization becomes more decentralized, as agents in the lower level are able to deal with more tasks on their own and require less assistance of entrepreneurs.

Now consider the case of a reduction in communication costs. Lower communication costs imply an increase in the number of problems entrepreneurs can solve, and therefore a higher value of knowledge that leads to an increase in the knowledge acquired by entrepreneurs. The increase in the knowledge of entrepreneurs in turn increases the value of knowledge of lower level manager or workers, which in turn leads to an increase in the knowledge acquired by lower level managers or workers. Notice that there is another effect that we have not emphasized. As communication costs go down and entrepreneurs can solve more problems, the value of agent's knowledge, which has the sole value of reducing the number of questions to the entrepreneur, goes down. However, the increase in knowledge acquired by the entrepreneur increases the value of the knowledge of managers/workers enough so that the second effect never dominates the first one. The formal result is presented in the next proposition. The proofs of all propositions in this section are in the appendix.

Proposition 1 *Given wages, assignments, and the equilibrium maximum number of layers the knowledge of entrepreneurs and next layer managers/workers increases with improvements in communication and access to information technology. That is,* Q_E^{l+1} *and* Q_E^l *are decreasing in* t *and* h *for all layers* $l = 0, ..., \ell - 1$.

A similar proposition cannot be proven for the case of intermediate managers and workers. The effect on their knowledge of improvements in the access to information and communication technology is uncertain. Given the knowledge of their boss, as the availability of information increases the monetary costs of their boss's time decreases, since acquiring knowledge is less costly for the boss. The value of the workers knowledge, which is the result of saving the manager time, also decreases. This implies that, given the boss's knowledge, the knowledge acquired by the intermediate manager/worker decreases. On top of this effect, we know that the boss will learn how to solve more problems, which in turn increases intermediate manager/worker productivity. These two effects go in opposite direction and it is not possible to prove in general that one will dominate the other.

As communication technology improves, entrepreneurs and managers can solve problems for more agents, which results in a decrease in the value of their time given their knowledge. This implies a decrease in the number of problems learned by middle management and workers, given the knowledge of their boss. Again we have the effect of more knowledge learned by entrepreneurs and next layer managers that implies an increase in the productivity of the knowledge of these agents. We prove the presence these two effects separately in the next two propositions.

Proposition 2 *Given wages, assignments, and the equilibrium maximum number of layers the knowledge of intermediate managers and workers increases with the knowledge of managers/entrepreneurs directly above them. That is,* Q_M^l *is a strictly increasing function of* Q^{l+1} *for all layers* $l = 0, ..., \ell - 2$.

Proposition 3 *Given wages, assignments, and the equilibrium maximum number of layers the knowledge of intermediate managers and workers decreases with improvements in the access to information and communication technology. That is,* Q_M^l *is a decreasing function of* t *and* h *given* Q^{l+1} *for all layers* $l = 0, ..., \ell - 2$.

We now turn to the effect of access to information and communication technology on the span of control. In order to analyze the effect of access to information technology on the spans of control throughout the hierarchy we need to know how all levels of knowledge change relative to each other. However, as we mentioned above, the sign of the effect of changes in access to information and communication technology on the knowledge of intermediate managers and workers is ambiguous. Given this, we can not make general statements about the span of control of each layer of management, namely statements about

$$\frac{n^{l-1}}{n^l} = \frac{1 - Q^{l-2}}{1 - Q^{l-1}}.$$

It is possible, however, to determine how the average span of control of entrepreneurs (measured by the number of workers in the firm) will change with improvements in the access to information and communication technology. To understand the claim that the number of workers has an interpretation as the average span of control, notice that

$$n^0 = \frac{n^\ell}{1} \frac{n^{\ell-1}}{n^\ell} ... \frac{n^0}{n^1},$$

and so the number of workers is the geometric mean of the span of control at all layers. An increase in the availability of information will lead to an increase in the knowledge of entrepreneurs and next layer managers as we showed above. This implies that the entrepreneur will have to answer less problems and so will be able to manage a larger firm. The same is true for decreases in communication costs, since the new technology will imply more knowledge of next layer managers/workers. However, in the case of communication technology we have another effect, since the improvements has a direct effect on the number of problems that entrepreneurs can solve. Both effects increase the number of workers in the firm. We present these results formally in the following proposition.

Proposition 4 *Given wages, assignments, and the equilibrium maximum number of layers the average span of control of entrepreneurs (or firm size) increases with improvements in the access to information and communication technology. That is, in a firm with $l+1$ layers,*

$$n^0 = \frac{1}{h(1-Q^{l-1})}$$

is increasing in t and h.

4 Equilibrium

The previous analysis has allowed us to obtain, for a given hierarchy, the proportion of tasks each agent should learn to perform, and team sizes.[19] This section solves for the assignment of workers to jobs, the wage schedule that supports this assignment, and the number of layers of each hierarchy.[20]

Let $Q^l(\alpha^l)$ be the equilibrium knowledge of an agent with ability α^l, $w^l\left(\alpha^l\right)$ her wage, and let $a^l(\alpha^l)$ be the equilibrium ability of a entrepreneur/manager of layer $l+1$ that manages an agent with ability α^l. That is, the function a^l assigns the ability of a manager/worker

[19] We will talk about workers, managers and entrepreneurs for simplicity; clearly, these are measure 0 atoms in a continuous distribution. The only appropriate way to think about them are masses of workers and managers at given intervals. The reader should thus not read from the existence of the assignment function that teams are homogeneous. Instead, any mass of workers may be matched with managers, depending on the shape of the skill distribution.

[20] Note that the problem we confront is different from Sattinger (1993) and Teulings (1995) (and all other) canonical ('Ricardian') assignment problems. First, rather than matching one worker and one machine, we match here one manager and any number of workers. Second, our economy must not assign given machines to given workers; we have instead to determine which workers are going chose to be managers and which one to be production worker. Third, the interaction between manager skill and worker skill is not direct, but takes place, as we shall see, through the team size and the knowledge acquired. In short, having smarter workers allows a manager to have a larger team, as they ask less often; but in turn, this requires a smarter manager, as the questions smarter workers ask are less common.

to the entrepreneur/manager one layer above. Fix a layer l and define the vector of assignments of all lower level abilities as a function of the ability of the entrepreneur/manager of layer l ($l = 1, ..., \ell$) by

$$\vec{a}^l(\alpha^l) = \left[\alpha^l, (a^{l-1})^{-1}(\alpha^l), ..., (a^0)^{-1}(...((a^{l-1})^{-1}(\alpha^l)))\right].$$

Similarly, denote the vector of wages as a function of the ability of the layer l manager by

$$\vec{w}^l(\alpha^l) = \left[w(\alpha^l), w^{l-1}((a^{l-1})^{-1}(\alpha^l)), ..., w^0((a^0)^{-1}(...((a^{l-1})^{-1}(\alpha^l))))\right].$$

The two vectors described above will help us define the value of a branch and the rents of the firm as a function of the ability of the entrepreneur. So define the equilibrium value of a branch per worker by

$$p^{\ell *}(\alpha^\ell, \alpha^{\ell-1}) \equiv p^\ell(Q^\ell(\alpha^\ell); \alpha^\ell, \vec{a}^{\ell-1}(\alpha^{\ell-1}), w^\ell(\alpha^\ell), \vec{w}^{\ell-1}(\alpha^{\ell-1})), \tag{15}$$

for $\ell \geq 1$ and

$$p^{0*}(\alpha^0) \equiv p^\ell(Q^0(\alpha^0); \alpha^0).$$

We can also define the rents of a hierarchy by

$$R^{\ell **}(\alpha^\ell, \alpha^{\ell-1}) = R(\alpha^\ell; p^{\ell *}(\alpha^{\ell-1}, (a^{\ell-2})^{-1}(\alpha^{\ell-1})), \tag{16}$$

for $\ell \geq 1$ and $R^{0**}(\alpha^0) \equiv R^0(\alpha^0)$.

In order for the function a^l to be well defined and invertible, we need the assignment to be singled valued and monotone. We prove that this is in fact the case at the end of this section.

The definition of equilibrium in this setup is given by:

Definition 1 *A competitive equilibrium is*

- *an integer maximum number of layers* ℓ*,*
- *a collection of sets* $\left\{A^l = A^{lM} \cup A^{lE}\right\}_{l=0}^{\ell}$ *such that for* $\alpha \in A^{lM}$ *agents choose to be members of management layer* l*,* $l = 1, ..., \ell$*, workers for* $\alpha \in A^{0M}$*, and entrepreneurs of layer* l *for* $\alpha \in A^{lE}$*,* $l = 0, ..., \ell$*,*
- *a wage* $w^l(\alpha^l) : A^{lM} \to \mathbb{R}_+$*, for managers of layers* $l = 1, ..., \ell$ *and workers,* $l = 0$*,*
- *an earnings function of entrepreneurs* $R^{l*}(\alpha^l) : A^{lE} \to \mathbb{R}_+$ *,* $l = 0, ..., \ell$*,*

- *a collection of assignment functions* $a^l(\alpha^l) : A^{lM} \to [0,1]$, $l = 0, ..., \ell - 1$,
- *and a vector of firm's choices* $(Q^\ell, ..., Q^0, n^\ell, ..., n^0)$,

such that:

(i) Agents occupational choice is optimal given wages and managerial income, that is, agents's utility is given by

$$U(\alpha) = \max\left\{\max_{l=0,...,\ell-1}[w^l(\alpha), R^{l*}(\alpha)], R^{\ell*}(\alpha)\right\}. \tag{17}$$

(ii) Managers' choices solve the objective function (12) for all layers and the number of managers and workers satisfies

$$n^l = hn^0(1 - Q^{l-1}) \text{ all } l = 1, ..., \ell. \tag{18}$$

(iii) The wage schedule is such that the assignment $\alpha^l = a^{l-1}(\alpha^{l-1})$ *is optimal for the manager or entrepreneur of layer* l*, i.e., the value of the branch per worker managed defined in (8) and (15), or of the firm defined in (16), satisfies, for* $\alpha^{lM} \in A^{lM}$ *and for* $\alpha^{lE} \in A^{lE}$,

$$\frac{\partial p^{l^*}(\alpha^{lM}, \alpha^{l-1})}{\partial \alpha^{l-1}} = \frac{\partial R^{l^{**}}(\alpha^{lE}, \alpha^{l-1})}{\partial \alpha^{l-1}} = 0, \tag{19}$$

and the second order condition is satisfied.

(iv) The market for production workers and managers at all layers clears: at every point $\alpha^l \in A^l$ *the number of workers* $(l = 0)$ *or managers* $(\ell > l > 0)$ *that are supplied is equal to the number of workers or managers demanded:*

$$\int_{[0,\alpha^l]\cap A^{lM}} \phi(\alpha)d\alpha = \int_{[a^l(0),a^l(\alpha^l)]\cap A^{l+1}} \frac{n^l(\alpha)\phi(\alpha)}{n^{l+1}(a^{l+1}(\alpha))}d\alpha, \tag{20}$$

for all $\alpha^l \in A^l$ *and* $l = 0, ..., \ell - 1$.

An important characteristic of the equilibrium, greatly simplifies its structure: the assignment involves positive sorting, in the sense that better managers are matched with better employees. That is, the assignment functions are strictly increasing. We prove this result in the next proposition.

Proposition 5 *An equilibrium of this economy involves positive sorting.*

High ability managers hire high ability agents so that they are shielded from solving easy and common problems. Hiring better workers allows managers to specialize in solving only the harder problems that lower layer agents can not solve. Proposition (5) confirms our assertion that the assignment functions were singled valued and monotone.

4.1 Assignment

We will assume that the collection of sets $\left\{A^l = A^{lM} \cup A^{lE}\right\}_{l=0}^{\ell}$ is such that there exists a set of real numbers $\left\{\alpha^{*ll}, \alpha^{*ll+1}\right\}_{l=0}^{\ell}$, with $\alpha^{*\ell\ell+1} = 1$, such that $A^{0M} = \left[0, \alpha^{*00}\right]$, $A^{lM} = \left[\alpha^{*l-1l}, \alpha^{*ll}\right]$, $A^{lE} = \left[\alpha^{*ll}, \alpha^{*ll+1}\right]$ and $A^{\ell E} = \left[\alpha^{*\ell\ell}, 1\right]$. Different collections of sets may be an equilibrium too. That is, there may be multiple layers of firms not only multiple layers in a firm. We prove below that an equilibrium with these characteristics exists.

Conditions (18) and (20) in the definition of equilibrium imply that

$$\begin{aligned} \frac{\partial a^l(\alpha^l)}{\partial \alpha^l} &= \frac{n^{l+1}(a^{l+1}(a^l\left(\alpha^l\right)))}{n^l\left(a^l(\alpha^l)\right)} \frac{\phi(\alpha^l)}{\phi(a^l(\alpha^l))} \\ &= \frac{\left(1 - Q^{*l}(a^l(\alpha^l))\right)}{\left(1 - Q^{*l-1}(\alpha^l)\right)} \frac{\phi(\alpha^l)}{\phi(a^l(\alpha^l))} \end{aligned} \tag{21}$$

for $l > 0$ and

$$\begin{aligned} \frac{\partial a^0(\alpha^0)}{\partial \alpha^0} &= \frac{n^1(a^1(a^0\left(\alpha^0\right)))}{n^0\left(a^0(\alpha^0)\right)} \frac{\phi(\alpha^0)}{\phi(a^0(\alpha^0))} \\ &= h\left(1 - Q^{*0}(a^0(\alpha^0))\right) \frac{\phi(\alpha^0)}{\phi(a^0(\alpha^0))}. \end{aligned}$$

As an example, suppose that we are assigning managers to workers. Then, the number of managers per subordinate worker that results from the firm optimization, over the ratio of available agents with the corresponding ability implied by the skill distribution, determines the slope of the assignment function. These are ordinary differential equations that determine the functions a^l given some initial values. The boundary conditions, are given by

$$\begin{aligned} a^0(0) &= \alpha^{*01}, \\ a^l(\alpha^{*ll}) &= \alpha^{*l+1l+2}, \\ a^{l+1}(\alpha^{*ll+1}) &= \alpha^{*l+1l+2}, \\ a^{\ell-1}(\alpha^{*\ell-1\ell-1}) &= 1, \end{aligned}$$

for $l = 1, ..., \ell - 1$. Naturally, the boundaries between the abilities of the different occupations of agents, $\left\{\alpha^{*ll}, \alpha^{*ll+1}\right\}_{l=0}^{\ell}$, are determined in equilibrium. We will go over the exact construction of an equilibrium below.

4.2 Wages and earnings

Condition (17) in the definition of equilibrium requires that agents choose optimally to become workers, managers or entrepreneurs, and the layer of management or entrepreneurship.

This implies that, for $l = 0, ..., \ell$,

$$w^l(\alpha^{*ll}) = R^{*l}(\alpha^{*ll}) \tag{22}$$

and, for $l = 0, ..., \ell - 1$,

$$R^{*l}(\alpha^{*ll+1}) = w^{l+1}(\alpha^{*ll+1}). \tag{23}$$

These conditions mean that at the boundaries that separate the ability sets of managers/ workers and entrepreneurs, and at the boundaries that separate the ability sets of entrepreneurs one layer below and managers, earnings have to equate. That is, the earnings function is continuous. Clearly this is the case within occupations, however, we need to argue that it is also the case at the boundaries between occupations. Suppose the earnings function was not continuous, then agents with abilities arbitrarily close to the boundary would profit from changing profession and so the allocation would not be an equilibrium by condition (17).

The definition of equilibrium requires the optimal assignment of workers and managers at different layers to be optimal. Condition (19) implies that if α^l is a manager of layer l,

$$\frac{\partial p^{l*}(\alpha^l, \alpha^{l-1})}{\partial \alpha^{l-1}} = -\left(c_\alpha g(Q^{l-1}) + \frac{\partial w^{l-1}(\alpha^{l-1})}{\partial \alpha^{l-1}}\right) h(1 - Q^{\ell-2}) = 0.$$

If α^l is an entrepreneur of layer l,

$$\frac{\partial R^{l**}(\alpha^l, \alpha^{l-1})}{\partial \alpha^{l-1}} = -\left(c_\alpha g(Q^{l-1}) + \frac{\partial w^{l-1}(\alpha^{l-1})}{\partial \alpha^{l-1}}\right) \frac{h(1 - Q^{\ell-2})}{h(1 - Q^{\ell-1})} = 0.$$

Therefore

$$\frac{\partial w^l(\alpha^l)}{\partial \alpha^l} = -c_\alpha g(Q^l) > 0. \tag{24}$$

The slope of the entrepreneurs earning function can be determined using (12). Namely,

$$\frac{\partial R^{l*}(\alpha^l)}{\partial \alpha^l} = -c_\alpha g(Q^l(\alpha^l)) > 0. \tag{25}$$

Remember that $g(Q) \equiv z$, i.e. it represents the interval of tasks that a worker must learn. In words, the condition means that the marginal return to skill is the marginal value of agents skill, which is given by the knowledge the agent must acquire. This equation contains a lot of the intuition of what follows. More inequality will result in equilibrium whenever a technology change leads agents to learn more tasks, as then the difference between more and less skilled agents becomes more pronounced.

By definition of multiple layers, $Q^{l+1}(\alpha^{*ll+1}) \geq Q^l(\alpha^{*ll+1})$, and by Proposition 7 (See Section 4.3), the equilibrium amount of knowledge is increasing in ability, all $l = 0, ..., \ell - 1$.

That is, managers and entrepreneurs in higher layers know how to solve more rare and difficult problems and their knowledge increases with ability. Hence, since g is increasing, for $\alpha \in A^{lM}$ and $\bar{\alpha} \in A^{lE}$,

$$\frac{\partial w^l(\alpha)}{\partial \alpha^l} = -c_\alpha g(Q^l(\alpha)) < -c_\alpha g(Q^l(\bar{\alpha})) = \frac{\partial R^{l*}(\bar{\alpha})}{\partial \alpha^l} \tag{26}$$

for all $l = 0, ..., \ell - 1$. Similarly, for $\alpha \in A^{lE}$ and $\bar{\alpha} \in A^{l+1M}$,

$$\frac{\partial R^{l*}(\alpha)}{\partial \alpha^l} = -c_\alpha g(Q^l(\alpha)) < -c_\alpha g(Q^{l+1}(\bar{\alpha})) = \frac{\partial w^l(\bar{\alpha})}{\partial \alpha^l} \tag{27}$$

Define the equilibrium earnings function as

$$E(\alpha) = \begin{cases} w^l(\alpha) & \text{for } \alpha \in A^{lM} \\ R^{l*}(\alpha) & \text{for } \alpha \in A^{lE} \end{cases} . \tag{28}$$

Then, the above discussion implies that, if the earnings function is convex within a class, the whole earnings function is increasing and convex in ability.[21] That means that wages grow faster than the ability of individuals. Smarter workers not only have lower cost of learning, but also learn more. Moreover, smarter workers are in higher layers and have larger teams below them.[22] We prove these properties of the earnings function in the following proposition.

Proposition 6 *The equilibrium earnings function is increasing and convex.*

4.3 Consultant services formulation

In this section we will present an alternative formulation of the economy described above. This formulation is not based on firms but on consultants. Workers produce. When they face a problem that they do not know how to solve, they may ask a consultant that specializes in solving hard problems. Some consultants are first layer consultants, their fee is low so workers first go to ask them. If they do not know the solution to the problem, workers may go to a higher layer consultant that charges another fee for trying to solve the problem. The number of layers of consultants will be determined in equilibrium. We will show at the end of this section that this formulation is equivalent to the one with firms. The formulation is useful both for interpretation purposes and to derive some of the results.

[21] The above discussion assumes that if $\alpha \in A^{lM}$ and $\bar{\alpha} \in A^{lE}$, $\bar{\alpha} > \alpha$. This assumption is not necessary for the argument but it simplifies the exposition. Proposition (10) shows that an equilibrium with this characteristic exists.

[22] Notice that convexity with respect to ability is equivalent to convexity with respect to the cost in terms of output of learning problems, since the latter is a linear function of ability.

Consider the problem of a worker of skill α^0 that hires consultants to solve problems for them. She can hire consultants sequentially if the previous consultant was not able to solve the problem. The fee for asking a consultant with knowledge Q is denoted by $f(Q)$. If we let $w^0(\alpha^0; f)$ be the earnings of a worker of skill α^0 given a function of fees f, the problem of the worker is given by:

$$\begin{aligned} w^0(\alpha^0; f) &= \max_{\ell, \{Q^l\}_{l=0}^{\ell}} Q^\ell - c(\alpha^0; t) g(Q^0) - \sum_{l=0}^{\ell-1} (1 - Q^l) f(Q^{l+1}) \\ \text{s.t. } 1 &\geq Q^{l+1} \geq Q^l \geq ... \geq Q^0 \geq 0, \end{aligned} \tag{29}$$

where ℓ is the maximum number of layers of consultants that the worker will ask for advice. This integer is determined by comparing the optimal value for the worker with or without asking another consultant. A simpler condition is that a worker will consult ℓ layers of consultants only if

$$\frac{Q^\ell - Q^{\ell-1}}{1 - Q^{\ell-1}} \geq f(Q^\ell). \tag{30}$$

For an unanswered question to be posed to the next layer, the conditional probability of finding the solution in that layer must be higher or equal than the fee. As we increase ℓ by one, on one hand the extra knowledge acquired by consultants of layer $\ell + 1$ is highly leveraged since they can promise to answer questions for many workers; many problems will be solved before they reach them. On the other hand, these consultants have the knowledge to solve all the problems that do not reach them. They, however, never use the knowledge that lower level consultants also have. In this sense, the acquisition of knowledge is duplicated and this is costly. The trade-off between the two forces determines the number of layers, both in this and the previous formulation of the economy.

The first order conditions of the problem are given by

$$f(Q^{1D}) - c(\alpha^0; t) g'(Q^{0D}) = 0, \tag{31}$$

$$f(Q^{l+1D}) - (1 - Q^{l-1D}) f'(Q^{lD}) = 0, \tag{32}$$

$$1 - (1 - Q^{\ell-1D}) f'(Q^{\ell D}) = 0, \tag{33}$$

$$l = 1, ..., \ell - 1.$$

iff $1 \geq Q^{l+1D} \geq Q^{lD} \geq ... \geq Q^{0D} \geq 0$. These equations implicitly define the demand for knowledge of layer l consultants of a worker with skill α^0, $Q^{lD}(\alpha^0)$.

The problem of a consultant is to choose a knowledge level so as to maximize the fees he receives. Hence the problem of a consultant with skills α^l is given by

$$R(\alpha; f) = \max_{Q^l} \frac{f(Q)}{h} - c(\alpha; t)g(Q). \tag{34}$$

The first order condition of the problem is then given by

$$\frac{f'(Q^S)}{h} - c(\alpha^l; t)g'(Q^S) = 0. \tag{35}$$

Hence, we denote by $Q^S(\alpha)$ the supply of knowledge of a consultant with skills α and $\left[Q^S\right]^{-1}(Q)$ the inverse function. The latter exists since $Q^S(\alpha)$ is a strictly increasing function (see Proposition (7) below). Using the inverse function, we can define an assignment function a^l that assigns consultants of layer l to workers. In particular,

$$\alpha^l = \left[Q^S\right]^{-1} [Q^{Dl}(\alpha^0)] \equiv a^l(\alpha^0). \tag{36}$$

Proposition 7 *Q^S is a strictly increasing function of α.*

Proposition 8 *Q^S is strictly decreasing in t and h.*

We are ready to define an equilibrium in this economy:

Definition 2 *A competitive equilibrium is*

- *an integer maximum number of layers ℓ,*
- *a collection of sets $\left\{A^l\right\}_{l=0}^{\ell}$ such that for $\alpha \in A^l$ agents choose to be consultants of layer l, $l = 1, ..., \ell$, and workers for $\alpha \in A^0$, a rent function R for consultants and a wage function w^0 for workers; a fee function $f(Q)$,*
- *an assignment function a^l for each layer of consultant, $l = 1, ..., \ell$,*
- *and a function of knowledge choices $(Q^\ell, ..., Q^0)$, such that:*

(i) Agents occupational choice is optimal given wages and managerial income, that is, it satisfies

$$U(\alpha) = \max\{w^0(\alpha; f), R(\alpha; f)\}. \tag{37}$$

(ii) Consultant' choices solve the objective function (34);

(iii) Workers' choices solve the objective function (29)

(v) Supply and demand of problem solutions equates for all layers of consultants, $l = 1, ..., \ell$. That is, for all $\alpha^0 \in A^0$,

$$\int_{[0,\alpha^0]\cap A^0} \phi(\alpha)d\alpha = \int_{\left[a^l(0), a^l(\alpha^0)\right]\cap A^l} \frac{1}{(1 - Q^{l-1}(\alpha))h} \phi(\alpha)d\alpha. \tag{38}$$

We now prove that both formulations of the problem are equivalent.

Proposition 9 *The set of equilibria in the economy presented in Section 2.4 is identical to the set of equilibria in the economy presented in Section 4.3. That is, both formulations are equivalent.*

4.4 Existence and characterization of an equilibrium

The next proposition shows that there exists an equilibrium with the characteristics described. That is, there exist an equilibrium where agents sort according to skill. The lowest skill agents become workers and higher skill agents, that become managers or entrepreneurs, work at a higher layer the higher their skill. That is, agents in the economy are segmented by cognitive skill.

Before we present the proposition, it is useful to outline the algorithm to find an equilibrium in this economy. An equilibrium can be constructed as follows:

1. Set $\ell = 0$ and fix $\alpha^{*00} = 0$, $\alpha^{*01} = 1$ and $w^0(0) = R^{0*}(0)$.

2. We can calculate $R^{0*}(\alpha)$, all $\alpha \in [0,1]$, and, given $w^0(0)$, $R^{1**}(1,0)$. Compare $R^{0*}(1)$ with $R^{1**}(1,0)$. If
$$R^{0*}(1) > R^{1*}(1),$$
the allocation is an equilibrium and $\ell = 0$. If
$$R^{0*}(1) > R^{1*}(1),$$
$\ell > 0$, proceed to 3.

3. Set $\ell = 1$ and fix $\alpha^{*00} > 0$, α^{*01}, α^{*11}, $\alpha^{*12} = 1$ and $w^0(0) > R^{0*}(0)$.

4. Given $w^0(0)$ we can calculate $w^0(\alpha)$ for all $\alpha \in \left[0, \alpha^{*00}\right]$ and $R^{0*}(\alpha)$ for all $\alpha \in \left[\alpha^{*00}, \alpha^{*01}\right]$. Find α^{*00} such that
$$w^0(\alpha^{*00}) = R^{0*}(\alpha^{*00}).$$

5. Let the supply of workers accumulated in the interval $\left[0, \alpha^{*00}\right]$ be denoted by S^0 and the demand for workers accumulated in $\left[\alpha^{*01}, \alpha^{*12}\right]$ be denoted by D^0. (Using (21) to calculate assignments guarantees that condition (20) in the definition of equilibrium is satisfied in the interior of these intervals). Find α^{*01} such that $D^0 = S^0$.

6. Compare $R^{0*}(\alpha^{*01})$ with $R^{1*}(\alpha^{*01})$. Find $w^0(0)$ such that

$$R^{0*}(\alpha^{*01}) = R^{1*}(\alpha^{*01}).$$

Compare $R^{1*}(1)$ with $R^{2*}(1)$. If

$$R^{1*}(1) \geq R^{2*}(1),$$

$\ell = 1$ and we have found an equilibrium. If

$$R^{1*}(1) < R^{2*}(1),$$

$\ell > 1$ so proceed as in 3. but with $\ell = 2$.

Figure 4 presents a diagram of an earnings function with the different intervals and conditions that have to be met in equilibrium. At the boundaries $(0, \alpha^{*00}, \alpha^{*01}, \alpha^{*11}, ..., \alpha^{*\ell\ell}, 1)$ the wage function is not differentiable since the amount of knowledge acquired changes discontinuously. The next proposition shows that there exists an equilibrium with the characteristics described above.

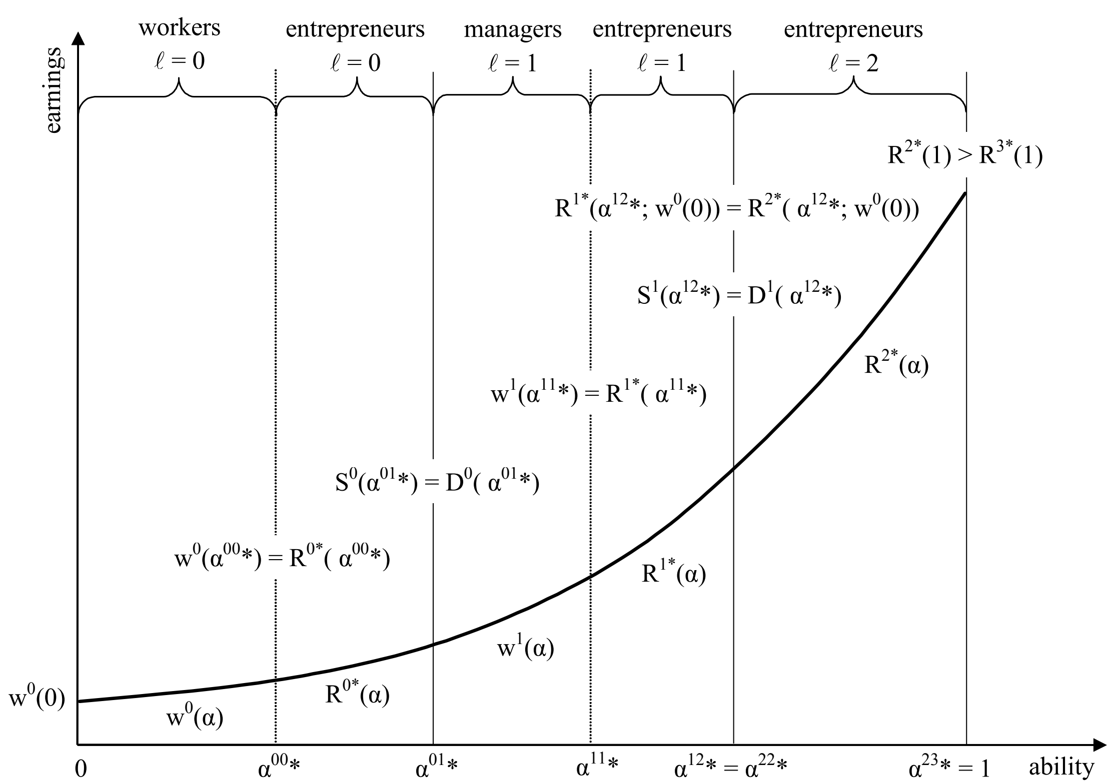


Figure 4: Construction of an equilibrium allocation

Proposition 10 *There exists an equilibrium allocation with the property that, for some vector of thresholds* $(0, \alpha^{*00}, \alpha^{*01}, \alpha^{*11}, ..., \alpha^{*\ell\ell}, 1)$, *agents with ability* $\alpha \in \left[0, \alpha^{*00}\right]$ *choose to become workers, agents with ability* $\alpha \in \left[\alpha^{*ll}, \alpha^{*ll+1}\right]$ *choose to become entrepreneurs of layer* l, $l = 0, ..., \ell$ *with* $\alpha^{*\ell\ell+1} = 1$, *and agents with ability* $\alpha \in \left[\alpha^{*l-1l}, \alpha^{*ll}\right]$ *choose to become managers of layer* l, $l = 1, ..., \ell$. *Furthermore, the last layer in the economy* (ℓ), *consists only of entrepreneurs. That is,* $\alpha^{*\ell-1\ell} = \alpha^{*\ell\ell}$.

So far we have studied the model without reference to the efficiency properties of the equilibrium allocation. Since the total time endowment of agents is allocated, and the problems that are not eventually solved would be more costly to solve than the benefits that agents may derive from solving them, there is no reason to believe that the equilibrium allocation of this economy is not efficient. However, the possibility of multiple equilibria introduces the concern that some of the equilibrium allocations may not be global optima. In the next proposition we prove that the equilibrium with the properties described in Proposition (10) is the global optimum of a constrained set of allocations.

Proposition 11 *The equilibrium allocation that is guaranteed to exist by Proposition (10) is a global optimum in the set of allocations characterized by arbitrary thresholds* (0, α^{00}, α^{01}, α^{11}, ..., $\alpha^{\ell\ell}$, 1).

We end this section with a result on the size of hierarchies, or the average span of control of entrepreneurs, and how it relates to wages. In equilibrium, higher ability agents are matched with better subordinates and they are assigned to higher ranked positions. This implies that wages in larger firms are larger than in smaller firms. Namely:

Proposition 12 *In equilibrium, larger firms, measured by number of workers, pay higher wages.*

5 The Effect of Changes in the Technology to Access and Communicate Information

This section analyzes the evidence discussed in the introduction in light of the predictions of the model. We argue that a decrease in the cost of accessing information lead to the changes in the wage structure and organization of production observed in the 80's and early 90's. These improvements in information technology also resulted in the increases in decentralization and spans of control, and in the delayering of firms, identified by the empirical literature. We also claim that further changes in the cost of accessing information, *together* with important decreases in the cost of communicating information in the mid and late 90's, lead to a stable first, and then slightly decreasing wage inequality, accompanied by a first stable, and then slightly increasing, average firm size. This latter change in communication technology has other implications on organization that, as far as we know, have not been tested in the empirical literature. In particular, according to our theory, it should have caused an increase in spans of control, constant or slightly decreasing decentralization, and constant or slightly increasing number of layers.

To illustrate these qualitative predictions of the model and advance the characterization of an equilibrium allocation, we compute several numerical examples of the model. We will use these examples to explain the general equilibrium effects of both types of technological changes in our theory.

5.1 An Example with Exponential Problem Density

We study a concrete example of the previous model with an exponential density of problems, $f(z) = e^{-\lambda z}$ and a uniform distribution of worker talent, $\alpha \sim U[0,1]$. Moreover, we let $k = 1$ and $\lambda = 2$ in all exercises. Figure 5 presents the results of this numerical simulation.

The six graphs presented in Figure 5 show the equilibrium wage and knowledge of all agents for different parameter values. The figures in the first row present the earnings curves for $h = 0.98$, and $t = 1.9$ and 1.1 respectively. The second row presents earnings when we lower the value of h to 0.8 for the same values of t, and the third row shows the same exercise for $h = 0.7$. The parameter values have been chosen to maximize visibility. Table 1 may also be useful to understand the numerical results presented in Figure 5 and described in the two next subsections.

Table 1: Numerical examples

Total wage inequality, SD			Worker/Self-employed wage inequality, SD			Entrepreneur/Manager wage inequality, SD		
$h\backslash t$	1.9	1.1	$h\backslash t$	1.9	1.1	$h\backslash t$	1.9	1.1
0.98	0.068	0.202	0.98	0.038	0.202	0.98	0.018	N/A
0.8	0.057	0.206	0.8	0	0.119	0.8	0.055	0.132
0.7	0.063	0.300	0.7	0	0.091	0.7	0.060	0.243

Span of Control ($\alpha = 1$)			Wage for $\alpha = 0$			Q of highest ability worker		
$h\backslash t$	1.9	1.1	$h\backslash t$	1.9	1.1	$h\backslash t$	1.9	1.1
0.98	1.02	0	0.98	0.01	0.12	0.98	0	0.95
0.8	1.25	6.16	0.8	0.07	0.19	0.8	0	0.80
0.7	1.43	109.89	0.7	0.11	0.26	0.7	0	0.78

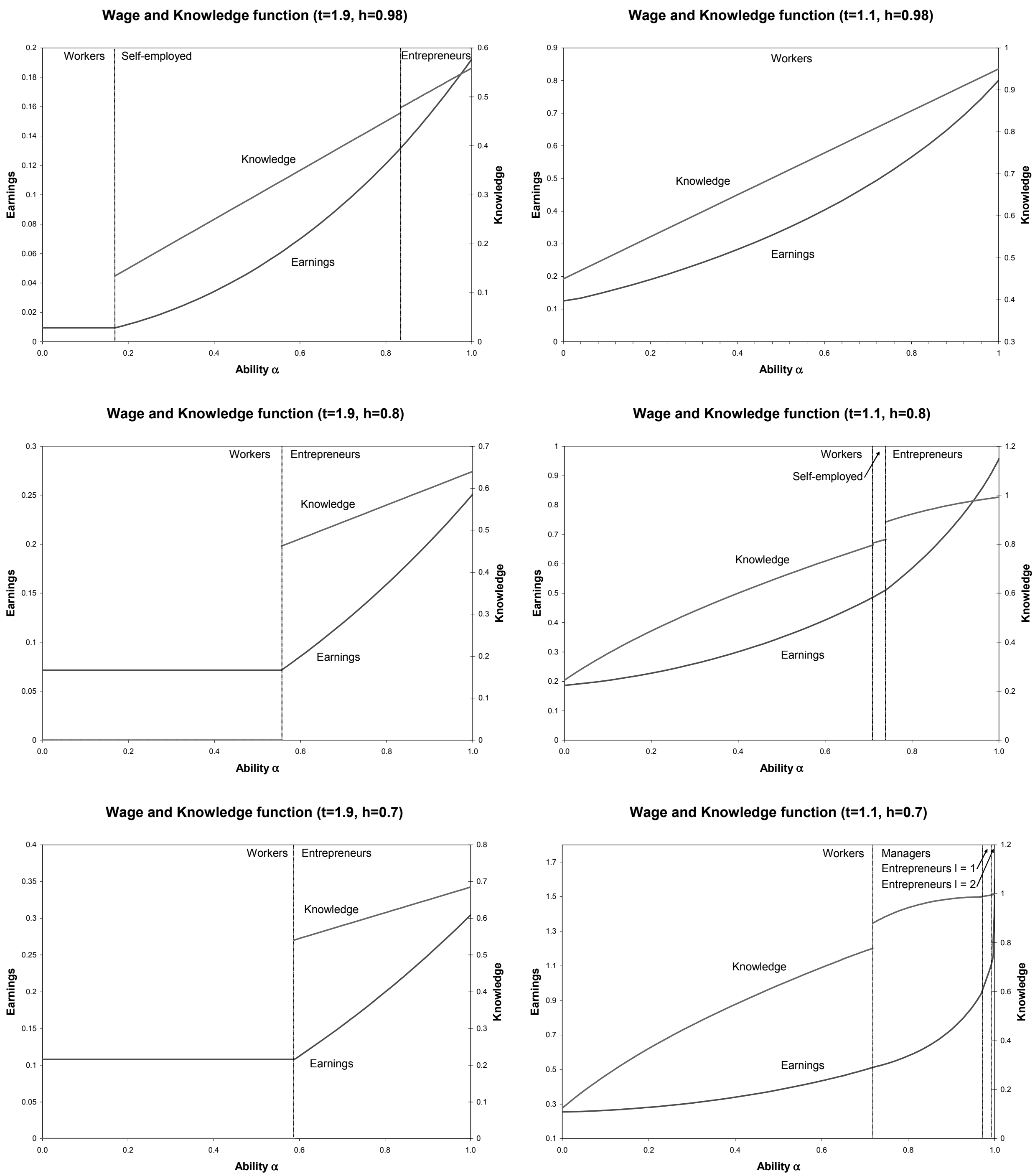


Figure 5: Numerical examples

There are several features of the equilibrium we have described above that are apparent in all the simulations presented. Wages are increasing and convex, higher ability agents learn more, and there is positive sorting. All of these are general results in our model and so should be present in all simulations. We now turn to the description of the effect of changes in communication and access to information technology.

5.1.1 Reduction in Communication Cost

In this subsection we analyze the equilibrium effect of reductions in communication costs (smaller h). As h decreases, the maximum number of layers in the economy increases. For $h = 0.98$ and $t = 1.1$ the economy has only workers that are self-employed. As we decrease h, keeping t constant, we move to an allocation with workers, a small number of self-employed agents, and managers/entrepreneurs. Even lower communication costs ($h = 0.7$) results in an allocation where the lowest ability agents are workers, higher ability agents are self-employed, and there is one layer of managers and a layer of entrepreneurs. All these results are also apparent for the same h values but a higher t. We go from an allocation with workers, self-employed agents and managers/entrepreneurs to an allocation with workers and managers/entrepreneurs. We interpret this as a weak increase in the number of layers, since for low values of h all agents work in firms.

The maximum amount of knowledge acquired (the knowledge acquired by an agent with $\alpha = 1$) increases with decreases in h. In contrast, the knowledge of the highest ability worker decreases. That is, lower communication costs imply more centralization. The direct effect of h on spans of control dominates the decrease in decentralization, therefore yielding the larger span of control apparent in Table 1. The lower h, the more can knowledge be leveraged so the higher the gains from an extra layer of managers. On the other hand, the cost of duplication -given by the fact that managers need to learn how to solve the problems that their subordinates solve - does not change. Hence we obtain more layers of management. Together, the increases in span of control and number of layers imply an increase in the size of hierarchies.

Wage inequality within the working class decreases as communication technology improves, since workers learn how to solve less problems. However, the higher leverage of knowledge, resulting in higher number of layers and higher span of control, implies more wage inequality within managers/entrepreneurs. These two effect are specially clear in Table 1 for the case of $t = 1.1$, for $t = 1.9$ workers do not learn, so there is no within group wage inequality. In fact, the decrease in worker wage inequality, combined with the increase in manager/entrepreneur wage inequality may result in some agents earning less after the improvement in communication costs. This is the case in Figure 6 when $t = 1.1$ and communication costs go from 0.8 to 0.7. The example shows how there may be some losers as technology improves!

Total wage inequality, measured by the standard deviation of wages, first decreases and then increases as we decrease h. Decreases in h reduce blue collar wage inequality and increase white collar wage inequality. The first effect dominates for high levels of h and the second for low levels. Notice that the wage of the lowest ability agent increases as communication costs fall. The reason is that the higher span of control results in higher demand for workers, part of the effect is translated in higher wages for all workers and part as an increase in the ability of the best worker, thereby increasing the amount of workers.

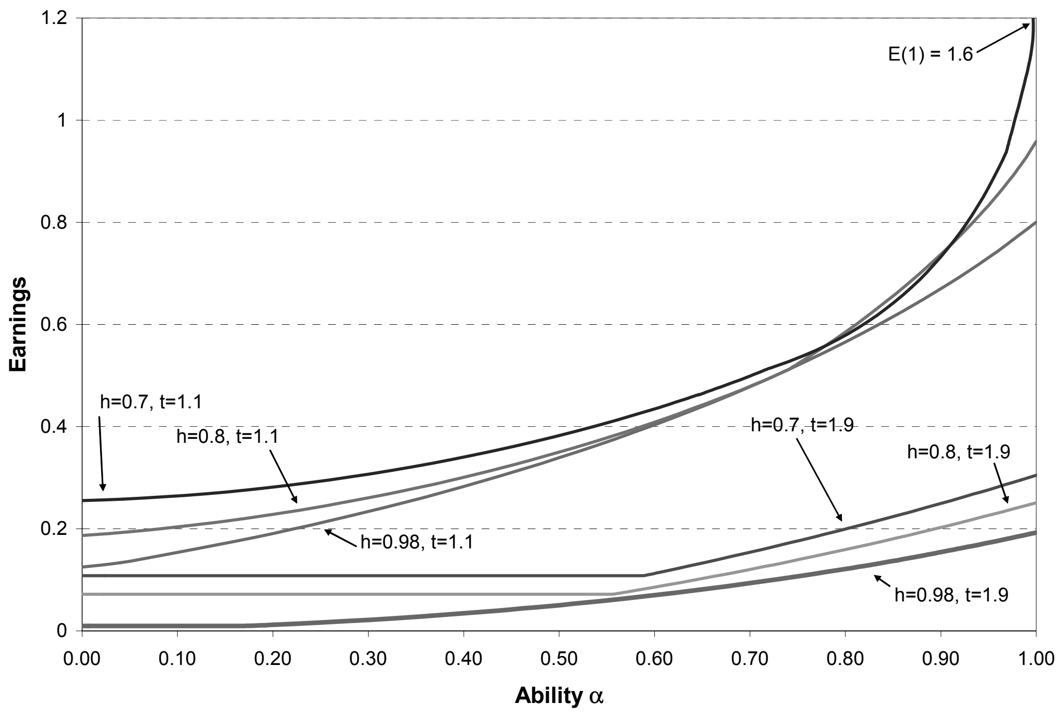


Figure 6: Equilibrium Wage Function

5.1.2 Reduction in the cost of acquiring knowledge

Consider now an increase in information availability leading to a reduction in the cost of learning tasks. As Figures 5 and 6 and Table 1 show, the general equilibrium impact of the change is a substantial increase in wage inequality within groups (measure as the standard deviation of wages). Both workers and managers acquire more knowledge since it is cheaper (more decentralization). This implies a substantial increase in workers and managers wage inequality. That is, within class wage inequality increases (See Figure 6). Since workers learn how to solve more problems, their wage increases as well. This effect is amplified since knowledgeable workers ask less questions to managers and so managers can increase their span of control (see Table 1). This in turn increases the demand for workers and therefore their wage. However, the gain from more knowledge, and more leverage of knowledge by managers, dominates the effect of higher wages of workers, and so we get an increase in total wage inequality for all three levels of communication costs.

The effect of changes in access to information technology on the number of layers is ambiguous. On one hand, lower costs of acquiring knowledge imply a lower gain from leveraging knowledge and so the value of having multiple layers decreases. On the other hand, the cost of having extra layers, overlapping knowledge, decreases too. The effect of delayering is clear for $h = 0.98$. In this case, as we decrease t, we move from an allocation with workers, managers/entrepreneurs, and self-employed agents to an allocation with workers only. In this case, improvements in the cost of acquiring information lead to smaller firms. In contrast, for low values of h, the effect of overlapping knowledge dominates and we get an allocation with 3 layers instead of 2 as we decrease t.

One of the interesting features of the model is the effect on wages of increases in managerial span. If a firm has many layers so that the highest layer entrepreneur manages a firm with a large number of workers, she can leverage her knowledge immensely which results in very high earnings. This effect can be appreciated in the numerical exercise for $h = 0.7$ and $t = 1.1$. This is the example in which we assume the best access to information and communication technology. The result, as we described above, is an economy with three layers. There are very few entrepreneurs of layer three. As can be observed in Figure 5, these entrepreneurs do not learn much more than the managers or entrepreneurs of layer two, however, they earn much more. The extend of this increase in wages can be appreciated in Figure 6, where we have plotted the equilibrium wage functions for all the examples presented above[23].

[23] As is evident in Figure 6, the wage of agents with ability $\alpha = 0$, when $t = 1.9$, increases as we increase h. These workers cannot solve any problems, the increase in their wage reflects the fact that a higher fraction of problems is solved and so the expected value of production increases. Since layer 0 workers are necessary to produce, the increase in the expected value of production is reflected in their wage. In reality, it is not clear that the gains from a higher expected value of production are assigned to unskilled workers. This is why we focus the analysis on the structure, and not the level, of the earnings function. The same is true for changes in t, although in this case workers with $\alpha = 0$ do know how to solve the simplest problems.

5.2 Empirical evidence and predictions of the theory

The empirical evidence presented in the introduction finds that in the 80's and early 90's both within and between wage inequality increased. These changes have parallel changes in the organization of the firm. Namely, more decentralization, increases in span of control and less layers of management. According to the numerical exercises in the previous section, this is only consistent with improvements in the technology to access information. In the numerical exercises, we are able to match qualitatively the results in the empirical literature when communication costs are relatively high and access to information technology improves. As we argued, for high h, decreases in t lead to delayering, which resulted in smaller hierarchies (in terms of number of employees) during this period. In particular, in our numerical example, all firms became one person firms as all workers are self-employed after the technological change. Notice that we are assuming that hierarchies and firms are equivalent for the purpose of this comparison. This does not have to be the case. However, as long as the determinants of the difference between firms and hierarchies is not changing throughout the period our conclusion applies.

We are also able to qualitatively match the empirical evidence of the mid and late 90's with improvements in communication technology. Decreases in h lead to lower within worker class wage inequality. This effect balances out with higher white collar wage inequality thereby resulting in small effects on total wage inequality (See Figure 3). Lower h also results in hierarchies with more layers and larger spans of control, which implies larger hierarchies in terms of total number of employees. Our view is that these effects were combined with further decreases in the costs of accessing information thereby dampening the effect of changes in communication technology on wage inequality and firm size. To our knowledge there are no systematic studies of the reorganization of the workplace for the late 90's to which we could contrast the predictions of our theory for this period. We will therefore offer those as testable implications. Namely, the model implies that companies became slightly more centralized in the late 90's, spans of control increased significantly, and the number of layers increased slightly. Since we are assuming that both technologies improved during this recent period, resulting changes that go in opposite directions should be small and changes that go in the same direction should be stronger.

6 Conclusion

This paper has presented a full fledged equilibrium model of organization and earnings when production requires physical inputs and knowledge and workers are heterogeneous in their cognitive skill. The equilibrium obtained is formed by a universe of knowledge-based hierarchies competing for workers and managers each with different numbers of layers, spans and skill sets. Among the salient features of the equilibrium are positive sorting between workers and managers, an increasing and convex earnings function, the stratification of worker skill into ranks by talent, and the positive relationship between wages and firm size.

The model has allowed us to provide an explanation for the patterns of organizational change and wage inequality described in the introduction. To summarize, the causal mechanism proposed for the 80's and early 90's is as follows. A decrease in the cost of access to stored knowledge leads teams to want to increase the worker's knowledge about tasks and, as a result, allows for an increase in managerial span. This change leads to an increase in the demand for skills which is dampened by the labor market supply of skilled workers. If communication costs are high, these effects lead to a decrease in the number of layers which further enhances the effect on decentralization and reduces firm size. Wage inequality within workers and within managers increases, as the marginal product of skill is proportional to the tasks learned.

Some of these effects where eliminated or reversed in the mid and late 90's by improvements in communication technology that lead to larger teams and less wage inequality. As communication technology improves, workers learn less since it is cheaper to ask their managers. These results in small differences in worker wages and therefore less wage inequality. Managers, in contrast, can deal with more subordinates therefore increasing their spans of control and earnings. Combined with taller hierarchies, these changes result in larger hierarchies or firms.

More broadly, our analysis shows that to understand the determinants of demand for skills and of wage inequality it is necessary to understand the internal structure of teams; conversely, to understand changes in the internal firm organization it is necessary to incorporate in the analysis labor market outcomes such as the wage structure. For example, the effects of communication technology improvements on the wage structure and on the assignment of workers to jobs can interact with, to the point of nullifying, effects that are unambiguous from a production function standpoint, such as the extent to which tasks are decentralized.

The model has some limitations that should be noted here and, hopefully, will lead to future work. Hierarchies do more than acquire and communicate knowledge, as the literature discussed in the introduction suggests. Future work embedding other models of hierarchy in an equilibrium framework is required in order to gauge the performance of our theory. Future research is also required to empirically test both the time series implications of the model (which, as we saw, follow from an economy wide drop in the cost of IT) and the cross sectional implications concerning the links between firm size, span, layers, and wages through the heterogeneity in worker skill. For example, carefully examining, in a matched employee-employer data set like the one used by Abowd, Kramarz and Margolis (1999), the relationship between unobservable employee effects and firm organization could provide compelling tests of the validity of our theory. Finally, we believe that the type of equilibrium model developed in this paper is a first step towards incorporating a richer organizational structure into dynamic equilibrium models. These models could then, for example, be used to study the dynamic relationship between firm size and wage inequality without relying, as in this paper, on comparative statics exercises.

REFERENCES

Acemoglu, Daron. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics*, Vol. 113, (November 1998): 1055–1089.

Autor, David H., Katz, Lawrence F. and Alan.B. Krueger (1998). "Computing Inequality: Have Computers Changed the Labor Market?." *Quarterly Journal of Economics*, Vol. 113, (November 1998): 1169-1213.

Abowd, John M., Francis Kramarz and David N. Margolis, "High Wage Workers and High Wage Firms." *Econometrica*, Vol. 67, No. 2, (March 1999): 251-333.

Berman, Eli; John Bound and Zvi Griliches, "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures," *Quarterly Journal of Economics,* Vol. 109, (May 1994): 367-397

Bolton, Patrick, and Mathias Dewatripont. "The Firm As a Communication Network." *Quarterly Journal of Economics* CIX, No. 4 (November 1994): 809-39.

Brown, Charles, and James Medoff. "The Employer Size - Wage Effect." *Journal of Political Economy,* Vol. 97, No.5, (October 1989): 1027-1059.

Bresnahan Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics*, Vol. 117, (February 2002).

Brynjolfsson, Eric, and Loren Hitt. 1997. "Computing Productivity: Are Computers Pulling Their Weight?." Working Paper, Sloan School of Management, MIT, (October 1997).

Card, D. and DiNardo, J. E. "Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *NBER Working Paper* No. 8769, (2002).

Calvo, Guillermo A. and Stanislaw Wellisz, "Supervision, Loss of Control, and the Optimum Size of the Firm." *Journal of Political Economy*, Vol. 86, No. 5, (October, 1978): 943-952.

—, "Hierarchy, Ability, and Income Distribution." *Journal of Political Economy*, Vol. 87, No. 5, Part 1, (Oct., 1979): 991-1010.

Caroli, Eve and John Van Reenen, "Skill Biased Organizational Change? Evidence from a Panel of British and French Establishments." *Quarterly Journal of Economics*, Vol. 116, (November 2001): 1449-1492.

Cremer, Jacques, "A Partial Theory of the Optimal Organization of a Bureaucracy." *The Bell Journal of Economics*, Vol. 11, No. 2, (Autumn, 1980): 683-693.

Cremer, Jacques, Luis Garicano and Andrea Prat. "Coding in organizations" Mimeo, 2003.

Dunne, Timothy, John Haltiwanger, and Kenneth R. Troske. "Technology and Jobs: Secular Changes and Cyclical Dynamics." *NBER Working Paper* No. 5656, (1996).

Garicano, Luis. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy,* Vol. 108, No. 5, (October 2000): 874-904.

Galor, Oded., Omer Moav, "Ability-Biased Technological Transition, Wage Inequality, and Economic Growth." *Quarterly Journal of Economics*, Vol. 115, (May 2000): 469-326

Geanakoplos, John and Milgrom, Paul. "A Theory of Hierarchies Based on Limited Managerial Attention." *Journal of the Japanese and International Economy*, Vol. 5, No. 3, (September 1991): 205-25.

Grossman, Sanford and Oliver Hart. "The Cost and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, Vol. 94, No. 4, (August 1986): 691-719.

Hart, Oliver and John Moore. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, Vol. 98, No. 6, (December 1990): 1119-1158.

Hayek, Friedrich A. "The Use of Knowledge in Society." *American Economic Review*, Vol. 35, (September 1945):519-30.

Katz, Lawrence F. and Kevin M. Murphy. "Changes in Relative Wages, 1963-87: Supply and Demand Factors." *Quarterly Journal of Economics,* Vol. 107, (February 1992): 35-78

Krueger, Alan B. "How Computers Changed the Wage Structure: Evidence from Micro Data." *Quarterly Journal of Economics,* Vol. 108, (February 1993): 33-60.

Frank R. Lichtenberg, and Bill Lehr, "Information Technology and Its Impact on Productivity: Firm-level Evidence from Government and Private Data Sources, 1977-1993." *Canadian Journal of Economics*, Vol. 32, No. 2, (April 1999): 335-362.

Levy, Frank, and Richard Murnane. *Teaching the New Basic Skills.* New York: Free Press. 1996b.

Lucas, Robert E., Jr. "On the Size Distribution of Business Firms." *Bell Journal of Economics,* Vol 9, No. 2, (Autumn 1978): 508-523.

Moebius, Markus M, "The Evolution of Work", Working Paper, MIT, (1999).

Murnane, Richard J., John B. Willett and Frank Levy, "The Growing Importance of Cognitive Skills in Wage Determination.", *NBER Working Paper* No. 5076, (1995).

Qian, Yingyi. "Incentives and Loss of Control in an Optimal Hierarchy." *Review of Economic Studies*, Vol. 61, No. 3, (July 1994): 527-44.

Radner, Roy. "Hierarchy: The Economics of Managing." *Journal of Economic Literature,* Vol. 30, (September 1992): 1382-415.

—, and Timothy Van Zandt. "Information Processing in Firms and Returns to Scale." *Annales D'Economie Et De Statistique,* Vol. 25-26, (Jan.-Jun. 1992): 265-98.

Rajan, Raghuram G. and Julie Wulf. "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies.", *NBER Working Paper*, No. 9633 (2002).

—, and Luigi Zingales, "The Firm as a Dedicated Hierarchy: A Theory of the Origins and Growth of Firms." *Quarterly Journal of Economics*, Vol. 116, (August 2001): 805-851.

Rosen, Sherwin, "The Economics of Superstars." *American Economic Review*, Vol. 71, No. 5, (December, 1981): 845-858.

—. "Authority, Control and the Distribution of Earnings." *Bell Journal of Economics,* Vol. 13, No.2, (Autumn 1982): 311-23.

Saint-Paul, Gilles. "On the distribution of income and worker assignment under intrafirm spillovers, with an application to ideas and networks." *Journal of Political Economy,* Vol. 109 (1) (February 2001): 1-37.

Sattinger, Michael, "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature,* Vol. 31, (June 1993): 831-880.

Teulings, Coen N. "The Wage Distribution in a Model of the Assignment of Skills to Jobs." *Journal of Political Economy*, Vol. 103, No. 2, (April 1995): 280-315

7 Appendix

Proof of Proposition 1. Assume that both Q_E^{l+1} and Q_E^l are an internal solution of problem (12). First notice that the elements of the Hessian are given by

$$\begin{aligned}\frac{\partial^2 R^{l+1}(\alpha^{l+1};p^l)}{\partial\left(Q^{l+1}\right)^2} &= -c(\alpha^{l+1};t)g''(Q^{l+1}),\\ \frac{\partial^2 R^{l+1}(\alpha^{l+1};p^l)}{\partial Q^{l+1}Q^l} &= \frac{1}{h(1-Q^l)^2},\\ \frac{\partial^2 R^{l+1}(\alpha^{l+1};p^l)}{\partial\left(Q^l\right)^2} &= \frac{\partial^2 p^l(Q^l;\vec{\alpha}^l)}{\partial\left(Q^l\right)^2}\frac{1}{h(1-Q^l)},\end{aligned}$$

which implies that for the problem to be well defined $\partial^2 p^l(Q^l;\cdot)/\partial\left(Q^l\right)^2 < 0$. But notice that

$$\frac{\partial^2 p^{\ell-1}(Q_M^{l-1};\cdot)}{\partial\left(Q^{l-1}\right)^2} = -c(\alpha^{l-1};t)g''(Q^{l-1})h(1-Q_M^{l-2}) < 0.$$

That is, the value per worker of a firm branch must be concave in knowledge. The determinant of the Hessian is given by

$$\Delta^l = -\frac{1}{h^2(1-Q^l)^4}\left[1 + c(\alpha^{l+1};t)g''(Q^{l+1})h(1-Q^l)^3\frac{\partial^2 p^l(Q^l;\cdot)}{\partial\left(Q^l\right)^2}\right] > 0,$$

where the sign has to be positive since we are maximizing rents.

Using the first order conditions, it is easy to show that

$$\frac{\partial Q_E^{l+1}}{\partial t} = \frac{\partial^2 p^l(Q^l;\cdot)}{\partial\left(Q^l\right)^2}\frac{g'(Q^{l+1})h(1-Q^l)}{\Delta^l} < 0$$

and

$$\frac{\partial Q_E^l}{\partial t} = \frac{-g'(Q^{l+1})h}{\Delta^l} < 0.$$

Similarly, the corresponding derivatives with respect to h are given by

$$\begin{aligned}\frac{\partial Q_E^{l+1}}{\partial h} &= \frac{\partial^2 p^l(Q^l;\cdot)}{\partial\left(Q^l\right)^2}\frac{1}{h\Delta^l} < 0\\ \frac{\partial Q_E^l}{\partial h} &= -\frac{1}{h(1-Q^l)\Delta^l} < 0.\end{aligned}$$

Because of the possibility of corner solutions the proposition only holds weakly.□

Proof of Proposition 2. Deriving (9) with respect to Q^{l+1} and solving for the derivative of Q_M^l with respect Q^{l+1}, we obtain

$$\frac{\partial Q_M^l(Q^{l+1},\cdot)}{\partial Q^{l+1}} = -hc(\alpha^{l+1};t)g'(Q^{l+1})/\frac{\partial^2 p^l(Q_M^l;\cdot)}{\partial\left(Q^l\right)^2} > 0.\square$$

Proof of Proposition 3. The proof of this result can be derived directly from the first order condition in (9). Derive that expression with respect to t and h to obtain

$$\frac{\partial Q_M^l(Q^{l+1},\cdot)}{\partial t} = -hg(Q^{l+1}) / \frac{\partial^2 p^l(Q_M^l;\cdot)}{\partial\left(Q^l\right)^2} > 0,$$

and

$$\frac{\partial Q_M^l(Q^{l+1},\cdot)}{\partial h} = -\left[c(\alpha^{l+1};t)g(Q^{l+1}) + w^{l+1}\right] / \frac{\partial^2 p^l(Q_M^l;\cdot)}{\partial\left(Q^l\right)^2} > 0.$$

Again, both signs follow from $\partial^2 p^l(Q_M^l;\cdot) \diagup \partial\left(Q^l\right)^2 < 0$ (see Proposition 1).□

Proof of Proposition 4. Since t and h increase Q_E^{l+1} and Q_E^l, we know that the number of workers

$$n^0 = \frac{1}{h\left(1 - Q^{\ell-1}\right)}$$

increases with t and h. This implies, that every entrepreneur in every firm manages more workers.□

Proof of Proposition 5. Since employers can hire anyone in the economy, in equilibrium their choice has to be optimal (see Condition (19) in the definition of equilibrium). That is, for all layers l, if α^l is a manager of layer l,

$$\frac{\partial p^{l*}(\alpha^l,\alpha^{l-1})}{\partial\alpha^{l-1}} = \frac{\partial p^{l-1*}(\alpha^{l-1},\alpha^{l-2})}{\partial\alpha^{l-1}} = 0.$$

If α^l is an entrepreneur of layer l,

$$\frac{\partial R^{l**}(\alpha^l,\alpha^{l-1})}{\partial\alpha^{l-1}} = \frac{1}{h(1-Q^\ell)}\frac{\partial p^{l-1*}(\alpha^{l-1},\alpha^{l-2})}{\partial\alpha^{l-1}} = 0.$$

We will prove that the assignment exhibits positive sorting for the case of managers, the proof for the case of entrepreneurs is identical but using $R^{l*}(\alpha^l,\alpha^{l-1})$ instead of $p^{l*}(\alpha^l,\alpha^{l-1})$.

Derive the first expression with respect to α^{l-1} to get,

$$\frac{\partial\alpha^l}{\partial\alpha^{l-1}} = -\frac{\partial^2 p^{l*}}{\partial\left(\alpha^{l-1}\right)^2} / \frac{\partial^2 p^{l*}}{\partial\alpha^{l-1}\partial\alpha^l}.$$

The numerator of this expression is negative since the choice of the employer has to be optimal (See Proposition 1). The cross derivative is given by

$$\begin{aligned}
\frac{\partial^2 p^{l*}}{\partial\alpha^l\partial\alpha^{l-1}} &= \frac{\partial}{\partial\alpha^l}\left[\frac{\partial p^{l-1*}(\alpha^{l-1},\alpha^{l-2})}{\partial\alpha^{l-1}}\right] \\
&= \frac{\partial}{\partial\alpha^l}\left[-\left(c_\alpha g(Q^{l-1}) + \frac{\partial w^{l-1}(\alpha^{l-1})}{\partial\alpha^{l-1}}\right)h(1-Q^{l-1})\right] \\
&= -c_\alpha g'(Q^{l-1})\frac{\partial Q^{l-1}}{\partial\alpha^l},
\end{aligned}$$

where the second line uses the condition $\partial p^{l*}/\partial\alpha^{l-1} = 0$ for all l. We need to prove that in equilibrium $\partial Q^{l-1}/\partial\alpha^l > 0$. For this, we will use Propositions 7 and 9. Derive (9) for layer $l-1$ with respect to α^l to get

$$\begin{aligned}\frac{\partial Q^{l-1}}{\partial\alpha^l} &= \frac{-h\left[c_\alpha g(Q^l) + \frac{\partial w^l(\alpha^l)}{\partial\alpha^l}\right] - hc(\alpha^l;t)g'(Q^l)\partial Q^l/\partial\alpha^l}{\partial^2 p^{l-1}(Q^{l-1};\cdot)/\partial\left(Q^{l-1}\right)^2} \\ &= \frac{-hc(\alpha^l;t)g'(Q^l)\partial Q^l\big/\partial\alpha^l}{\partial^2 p^{l-1}(Q^{l-1};\cdot)\big/\partial\left(Q^{l-1}\right)^2}.\end{aligned}$$

Since

$$\frac{\partial^2 p^{l-1}(Q^{l-1};\cdot)}{\partial\left(Q^{l-1}\right)^2} < 0$$

(see Proposition 1) the proof is completed if we can show that $\partial Q^l/\partial\alpha^l > 0$ in equilibrium. But this is exactly what Proposition 7 below shows. Hence, given the equivalence of both formulations proved in Proposition 9, $\partial Q^l/\partial\alpha^l > 0$ and so $\partial Q^{l-1}/\partial\alpha^l > 0$.□

Proof of Proposition 6. We have shown in the text that the slope of the earning function is positive and increasing in the number of layers. To prove the result we need to show that the second derivative of the earning functions is positive within a given layer and occupation (wherever the function is differentiable). That is, we need to sign the following two expressions:

$$\frac{\partial^2 w^l(\alpha^l)}{\partial\alpha^{l2}} = -c_\alpha g^{'}(Q^l(\alpha^l))\frac{\partial Q^l(\alpha^l)}{\partial\alpha^l},$$

and

$$\frac{\partial^2 R^{l*}(\alpha^l)}{\partial\alpha^{l2}} = -c_\alpha g^{'}(Q^l(\alpha^l))\frac{\partial Q^l(\alpha^l)}{\partial\alpha^l}.$$

The term $-c_\alpha g^{'}(Q^l(\alpha^l))$ is positive by assumption. To show that $\partial Q^l(\alpha^l)/\partial\alpha^l$ is positive, we rely on Propositions (7) and (9) in the next section. □

Proof of Proposition 7. First notice that in order for the consultant maximization problem to be well defined we need

$$\frac{f''(Q^S)}{h} - c(\alpha^l;t)g''(Q^S) < 0.$$

Deriving (35) implicitly with respect to α^l we obtain that

$$\frac{\partial Q^S}{\partial\alpha} = \frac{c_\alpha g'(Q^S)}{\frac{f''(Q^S)}{h} - c(\alpha^l;t)g''(Q^S)} > 0.$$

The sign of $f''(Q)$ must be negative for the maximization problem to be well defined. □

Proof of Proposition 8. Deriving (35) implicitly with respect to t and h we obtain that

$$\begin{aligned}\frac{\partial Q^S}{\partial t} &= \frac{g'(Q^S)}{\frac{f''(Q)}{h} - c(\alpha^l;t)g''(Q^S)} < 0,\\ \frac{\partial Q^S}{\partial h} &= \frac{f''(Q)}{h^2\left[\frac{f''(Q)}{h} - c(\alpha^l;t)g''(Q^S)\right]}\\ &= \frac{1}{h\left[1 - \frac{hc(\alpha^l;t)g''(Q^S)}{f''(Q)}\right]} < 0,\end{aligned}$$

where the last inequality follows from the second order condition of the consultant problem. □

Proof of Proposition 9. In the first part of the proof we show that the problem in (7) has the same solution and value as the problem in (29) combined with the problem in (34). For this, let Q^l, n^l, R^{l*} and $\overline{w}^l$ be an equilibrium allocation according to Definition 1. Then the problem in (7) can be written as

$$R^{\ell*}(\alpha^\ell)n^\ell = Q^\ell n^0 - c(\alpha^\ell;t)g(Q^\ell)n^\ell - \sum_{l=0}^{\ell-1}\left[c(\alpha^l;t)n^l g(1-\frac{n^{l+1}}{hn^0}) + w^l(\alpha^l)n^l\right]$$

subject to

$$\begin{aligned}hn^0(1-Q^{\ell-1}) &= n^\ell\\ &\vdots\\ hn^0(1-Q^0) &= n^1.\end{aligned}$$

Notice that in the formulation in Section 2.4 we normalized $n^\ell = 1$. Here we are using a notation that allows for any normalization. In particular, normalize the problem dividing by n^0 and substitute the constraints to get

$$\begin{aligned}R^{\ell*}(\alpha^\ell)h(1-Q^{\ell-1}) &= Q^\ell - c(\alpha^\ell;t)g(Q^\ell)h(1-Q^{\ell-1}) - c(\alpha^0;t)g(Q^1) - w^0(\alpha^0)\\ &\quad -\sum_{l=1}^{\ell-1} h(1-Q^{l-1})\left[c(\alpha^l;t)g(Q^l) + w^l(\alpha^l)\right].\end{aligned}$$

But this is equivalent to

$$\begin{aligned}w^0(\alpha^0) &= Q^\ell - h(1-Q^{\ell-1})\left[c(\alpha^\ell;t)g(Q^\ell) + R^{\ell*}(\alpha^\ell)\right]\\ &\quad -\sum_{l=1}^{\ell-1} h(1-Q^{l-1})\left[c(\alpha^l;t)g(Q^l) + w^l(\alpha^l)\right].\end{aligned}$$

Define

$$f(Q^l) \equiv h\left[c(\alpha^l;t)g(Q^l) + \max\left(R^{l*}(\alpha^l), w^l(\alpha^l)\right)\right].$$

Then

$$w^0(\alpha^0) = Q^\ell - \sum_{l=1}^{\ell} h(1-Q^{l-1})f(Q^l),$$

and

$$\max\left(R^{l*}(\alpha^l), w^l(\alpha^l)\right) = \frac{f(Q^l)}{h} - c(\alpha^l;t)g(Q^l)$$

which are consistent with (29) and (34). We still need to show that the choices resulting from the two stage maximization in the "Consultants" formulation are the same as the choices in (7). Notice however that combining (31)-(33), (35) and (36) we obtain the following system of first order conditions,

$$\begin{aligned}
c(\alpha^1;t)g(Q^1) + w^1(\alpha^1) - c(\alpha^0;t)g'(Q^0) &= 0, \\
c(\alpha^l;t)g(Q^l) + w^l(\alpha^l) - (1-Q^{l-2D})c(\alpha^{l-1};t)g'(Q^{l-1}) &= 0, \\
c(\alpha^{l+1};t)g(Q^{l+1}) + R^{l*}(\alpha^l) - (1-Q^{l-1D})c(\alpha^l;t)g'(Q^l) &= 0, \\
1 - (1-Q^{lD})c(\alpha^{l+1};t)g'(Q^{l+1}) &= 0, \\
l &= 1, ..., \ell - 1.
\end{aligned}$$

This system is identical to the system of first order conditions of the problem

$$\begin{aligned}
\overline{w}^0(\alpha^0) &= \max_{\{Q^l\}_{l=0}^{\ell}} Q^\ell - h(1-Q^{\ell-1})\left[c(\alpha^\ell;t)g(Q^\ell) + R^{l*}(\alpha^l)\right] \\
&\quad - \sum_{l=1}^{\ell-1} h(1-Q^{l-1})\left[c(\alpha^l;t)g(Q^l) + w^l(\alpha^l)\right],
\end{aligned}$$

which we have shown is equivalent to the problem in (7). Hence the maximization problems in both formulations are equivalent.

We still need to show that all other conditions in both definitions of equilibrium are equivalent. In particular, we need to show that (20) and (38) are equivalent and that (19) is satisfied in the second formulation. The former is trivial given (18) and the result obtained above. For the latter, first apply the Envelope Theorem to (34) to show that

$$\frac{\partial R(\alpha; f)}{\partial \alpha} = -c_\alpha g(Q).$$

Deriving $w^0(\alpha^0; f)$ with respect to α^l then yields

$$\begin{aligned}\frac{\partial w^0(\alpha^0; f)}{\partial \alpha^l} &= \left(1 - Q^l\right) \frac{\partial f(Q^l)}{\partial \alpha^l} \\ &= \left(1 - Q^l\right) \left[\frac{\partial R(\alpha; f)}{\partial \alpha} + c_\alpha g(Q)\right] \\ &= 0.\end{aligned}$$

This implies that the equilibrium assignment is optimal for the worker. Hence, an allocation of one formulation is an equilibrium if and only if it is an equilibrium of the other. □

Proof of Proposition 10. Fix a maximum number of layers, a vector of thresholds $\vec{\alpha}_\ell^{00} \equiv (\alpha^{00}, \alpha^{01}, \alpha^{11}, ..., \alpha^{\ell\ell})$[24] and an initial wage $\omega > R^0(0)$. Given this, we can solve the problems in (8) and (12) (call the solution $Q^l(\alpha; \omega, \vec{\alpha}_\ell^{00})$) and obtain wage, rent, and assignment functions for all layers using (21), (24), (25) and (23). Let $w^l(\alpha; \omega, \vec{\alpha}_\ell^{00})$, $R^{l*}(\alpha; \omega, \vec{\alpha}_\ell^{00})$ and $a^{l*}(\alpha; \omega, \vec{\alpha}_\ell^{00}))$ be the calculated wage, rent, and assignment function of layer l. Notice that the Theorem of the Maximum ensures that all functions Q^l, w^l, R^{l*} and a^{l*} are continuous in the initial wage ω, any one threshold α^j, $j = \{00, 01, ..., \ell\ell\}$, and in α. By definition

$$w^0(0; \omega, \vec{\alpha}_\ell^{01}) = \omega,$$

and the assignment functions are invertible since by (21) (justified by Proposition (5)) the assignment is singled valued and monotone. Let

$$\alpha^{00*}(\omega, \vec{\alpha}_\ell^{01}) \equiv \min\left[\left\{\alpha : w^0(\alpha; \omega, \vec{\alpha}_\ell^{01}) = R^{0*}(\alpha; \omega, \vec{\alpha}_\ell^{01})\right\}, \alpha^{01}\right].$$

Since $\omega > R^0(0)$ and the inequality in (26) holds, α^{00*} is well defined and, since the problem in (8) is continuous in w^l for all l, by the Theorem of the Maximum a continuous function of ω. Notice also that since w^0 and R^{0*} are continuous in all thresholds, α^{00*} is continuous in α^j, $j = \{01, ..., \ell\ell\}$.

Define the excess supply of workers as a function of the vector of thresholds and the wage of the lowest ability worker by

$$ES^0(\omega, \vec{\alpha}_\ell^{01}) \equiv \int_0^{\alpha^{00*}} \phi(\alpha) d\alpha - \int_{\alpha^{01}}^{\alpha^{12}} \frac{1}{h(1 - Q^1(\alpha; \omega, \vec{\alpha}_\ell^{01}))} \phi(\alpha) d\alpha, \quad (39)$$

and let $\alpha^{01*}(\omega, \vec{\alpha})$ be the threshold between layer 0 and 1, such that

$$\alpha^{01*}(\omega, \vec{\alpha}_\ell^{11}) \equiv \left\{\alpha^{01} : ES^0(\omega, \vec{\alpha}_\ell^{01}) = 0\right\}.$$

[24] We are using a notation in which the superindex of the vector of thresholds refers to the superindex of the first element of the vector. The subindex refers to the number of layers.

We need to show that α^{01*} is well defined, continuous and singled valued. First notice that (39) implies that $ES^0(\omega, \vec{\alpha}_\ell^{01})$ is continuous in α^{01}. Notice also that

$$ES^0\left(\omega, \left(0, \vec{\alpha}_\ell^{11}\right)\right) < 0$$

and

$$ES^0\left(\omega, \left(\alpha^{12}, \vec{\alpha}_\ell^{11}\right)\right) > 0.$$

Hence, by the Mean Value Theorem, α^{01*} exists. The function ES^0 may not be monotone and hence there may be several values at which it is equal to zero. On top of this, some of these crossings may disappear or the number of crossings may increase as we change the initial wage ω or any of the thresholds $\vec{\alpha}_\ell^{01}$. However, we know that at least one crossing exists by the argument above, and that there is at least one function α^{01*} that is continuous in ω and $\vec{\alpha}_\ell^{11}$. The reason is that ES^0 is continuous in both arguments. Let T be an operator that selects the minimum crossing that is continuous in ω and $\vec{\alpha}_\ell^{11}$, and redefine ES^0 as

$$\alpha^{01*}(\omega, \vec{\alpha}_\ell^{11}) \equiv T\left[\left\{\alpha^{01} : ES^0(\omega, \vec{\alpha}_\ell^{01}) = 0\right\}\right].$$

Since we are choosing the minimum continuous crossing, α^{01*} is single valued and continuous in both arguments, and by the argument above it exists.

We can proceed sequentially defining functions

$$\alpha^{ll*}\left(\omega, \vec{\alpha}_\ell^{ll+1}\right) \equiv \min\left[\left\{\alpha : w^l\left(\alpha; \omega, \vec{\alpha}_\ell^{ll+1}\right) = R^{l*}\left(\alpha; \omega, \vec{\alpha}_\ell^{ll+1}\right)\right\}, \alpha^{ll+1}\right],$$

$$ES^l(\omega, \vec{\alpha}_\ell^{ll+1}) \equiv \int_{\alpha^{l-1l*}}^{\alpha^{ll*}} \phi(\alpha) d\alpha - \int_{\alpha^{ll+1}}^{\alpha^{l+1l+2}} \frac{1 - Q^{l-1}(\left(a^{l-1}\right)^{-1}(\alpha); \omega, \vec{\alpha}_\ell^{ll+1})}{1 - Q^l(\alpha; \omega, \vec{\alpha}_\ell^{ll+1})} \phi(\alpha) d\alpha,$$

and

$$\alpha^{ll+1*}\left(\omega, \vec{\alpha}_\ell^{l+1l+1}\right) = T\left[\left\{\alpha^{ll+1} : ES^l(\omega, \vec{\alpha}_\ell^{ll+1}) = 0\right\}\right],$$

for all $l = 1, ..., \ell - 1$. Notice that we are abusing notation and ignoring the dependence of all function on thresholds for which we have obtained an equilibrium function. Clearly all of these are well defined, singled valued, and continuous by arguments that parallel the arguments above.

For each layer $l < \ell$ we need to check that

$$w^l\left(\alpha^{l-1l}; \omega, \vec{\alpha}_\ell^{ll+1}\right) > R^{l*}\left(\alpha^{l-1l}; \omega, \vec{\alpha}_\ell^{ll+1}\right). \tag{40}$$

For layer 0 we imposed this, for higher layers, since $w^l\left(\alpha^{l-1l}; \omega, \vec{\alpha}_\ell^{ll+1}\right)$ is given by (23) we need to guarantee that this condition is satisfied. That is, we need to guarantee that there are some workers/managers in each intermediate layer. If the condition is not satisfied for some l, and since

$$\lim_{\omega\to\infty} w^{l}\left(\alpha^{l-1l};\omega,\vec{\alpha}_{\ell}^{ll+1}\right)=\infty$$

by construction and

$$\lim_{\omega\to\infty} R^{l*}\left(\alpha^{l-1l};\omega,\vec{\alpha}_{\ell}^{ll+1}\right)=0$$

by (12), there exists an ω^{l} such that

$$w^{l}\left(\alpha^{l-1l};\omega^{l},\vec{\alpha}_{\ell}^{ll+1}\right)=R^{l*}\left(\alpha^{l-1l};\omega^{l},\vec{\alpha}_{\ell}^{ll+1}\right).$$

But this implies that the supply of managers of higher layers is zero and the demand is positive. Hence, set $\omega>\omega^{l}$.

If condition (40) is satisfied for all $l<\ell$, let $\alpha^{\ell\ell*}=\alpha^{\ell-1\ell*}$, and define the initial wage function ω^{*} by

$$\omega^{*}=\min\left\{\omega: R^{\ell-1*}\left(\alpha^{\ell\ell*};\omega\right)=R^{\ell*}\left(\alpha^{\ell\ell*};\omega\right)\right\}.$$

The problem in (12) and the Theorem of the Maximum imply that the functions $R^{\ell-1}(\alpha;\omega)$ are continuous in ω. As we increase the initial wage, there will be less entrepreneurs of each lower layer and there is a level of ω for which there are only entrepreneurs of layer ℓ, all other agents prefer to be workers or managers. Hence,

$$\lim_{\omega\to\infty} w^{\ell-1}\left(\alpha^{\ell\ell*};\omega\right)=R^{\ell-1*}\left(\alpha^{\ell\ell*};\omega\right)=\infty,$$

and by (12),

$$\lim_{\omega\to\infty} R^{\ell*}\left(\alpha^{\ell\ell*};\omega\right)=0.$$

We also know that

$$\lim_{\omega\downarrow\omega^{l}} w^{\ell-1}\left(\alpha^{\ell\ell*};\omega\right)<\lim_{\omega\to\infty} R^{\ell*}\left(\alpha^{\ell\ell*};\omega\right),$$

and so by continuity ω^{*} exists.

We still need to check that for the highest ability agents, $\alpha=1$, it is better to be entrepreneurs of layer ℓ than to become entrepreneurs of layer $\ell+1$. Namely,

$$R^{\ell*}(1)>R^{\ell+1*}(1), \tag{41}$$

where the entrepreneurs of layer $\ell+1$ would hire managers of layer ℓ with ability $\alpha^{\ell-1\ell*}$. If this is the case we have found an equilibrium allocation. If not, ℓ is not an equilibrium number of layers.

It remains to show that there exists an equilibrium number of layers ℓ. That is, there exist a number of layers for which (41) is satisfied. But notice that

$$\lim_{\ell\to\infty}\frac{Q^{\ell+1}-Q^{\ell}}{h\left(1-Q^{\ell}\right)}-c(\alpha^{\ell+1};t)g(Q^{\ell+1})=-c(\alpha^{\ell+1};t)g(Q^{\ell+1})<0.$$

Hence, since

$$c(1;t)g(Q^{\ell+1})>c(1;t)g(Q^{\ell}),$$

there exists an ℓ such that (41) is satisfied. $\square$

Proof of Proposition 11. The optimization problem for the restricted set of allocation is given by

$$\max_{Q(\cdot),\alpha^{j}\text{ for }j=\{00,01,...,\ell\ell\},\ell}\int\limits_{\alpha\in A^{lE},l=0,...,\ell} Q(\alpha)n^{0}(a^{-1}\left(\alpha\right))\phi\left(\alpha\right)d\alpha-\int\limits_{0}^{1}c\left(\alpha;t\right)g\left(Q\left(\alpha\right)\right)\phi\left(\alpha\right)d\alpha$$

subject to

$$\begin{gathered}
\frac{\partial a\left(\alpha\right)}{\partial\alpha}=\frac{1-Q\left(\alpha\right)}{1-Q\left(a^{-1}\left(\alpha\right)\right)}\frac{\phi\left(\alpha\right)}{\phi\left(a\left(\alpha\right)\right)},\\
n^{0}(a^{-1}\left(\alpha\right))=\frac{1}{h\left(1-Q(\alpha)\right)},\ \alpha\in A^{lE},l=0,...,\ell\\
a^{0}\left(0\right)=\alpha^{01},\\
a\left(\alpha^{ll}\right)=\alpha^{l+1l+2},\ l=0,...,\ell-2\\
a\left(\alpha^{ll+1}\right)=\alpha^{l+1l+2},\ l=0,...,\ell-2,\\
a\left(\alpha^{\ell-1\ell-1}\right)=1,\\
A^{0}=A^{0M}\cup A^{0E}=\left[0,\alpha^{00}\right]\cup\left[\alpha^{00},\alpha^{01}\right],\\
A^{l}=A^{lM}\cup A^{lE}=\left[\alpha^{l-1l},\alpha^{ll}\right]\cup\left[\alpha^{ll},\alpha^{ll+1}\right],\\
A^{\ell}=A^{\ell M}\cup A^{\ell E}=\left[\alpha^{\ell-1\ell},\alpha^{\ell\ell}\right]\cup\left[\alpha^{\ell\ell},1\right]=\left[\alpha^{\ell\ell},1\right].
\end{gathered}$$

Let $\lambda\left(\alpha\right)$ be the Lagrange multiplier associated with the first constraint. Then, the principle of optimality's necessary conditions for this problem are given by

$$\frac{1}{h\left(1-Q(\alpha)\right)}-c\left(\alpha;t\right)\frac{\partial g\left(Q\left(\alpha\right)\right)}{\partial Q}=0\text{ for }\alpha\in A^{lE},\tag{42}$$

$$c\left(\alpha;t\right)\frac{\partial g\left(Q\left(\alpha\right)\right)}{\partial Q}\left(1-Q\left(a^{-1}\left(\alpha\right)\right)\right)=\frac{\lambda\left(\alpha\right)}{\phi\left(a\left(\alpha\right)\right)}\text{ for }\alpha\in A^{lM}\tag{43}$$

and $\lambda\left(\alpha\right)$ is a continuous function that satisfies

$$\frac{\partial\lambda\left(\alpha\right)}{\partial\alpha}=\frac{\partial a\left(\alpha\right)/\partial\alpha}{\phi\left(a\left(\alpha\right)\right)}\frac{\partial\phi\left(a\left(\alpha\right)\right)}{\partial a(\alpha)}-\frac{\partial Q(a^{-1}\left(\alpha\right))/\partial\alpha}{1-Q\left(a^{-1}\left(\alpha\right)\right)}.\tag{44}$$

All other conditions are either pinned down by the constraints or by the continuity of $\lambda(\alpha)$.

To show that these conditions are equivalent to the conditions that determine the equilibrium allocation, derive (43) with respect to $a(\alpha)$ to get

$$\frac{\lambda(\alpha)}{\phi(a(\alpha))^2}\frac{\partial\phi(a(\alpha))}{\partial a(\alpha)} = c(\alpha;t)\frac{\partial g(Q(\alpha))}{\partial Q}\frac{\partial Q(a^{-1}(\alpha))}{\partial\alpha}\frac{1}{\partial a(\alpha)/\partial\alpha}.$$

Dividing by (43) and comparing the result with (44) results in

$$\frac{\partial\lambda(\alpha)}{\partial\alpha} = 0.$$

Using this result and comparing (42) and (43) with (13) and (11) we obtain that

$$\frac{\lambda(\alpha)}{\phi(a(\alpha))} = c(a(\alpha);t)\,g(Q(a(\alpha))) + w(a(\alpha)).$$

Since all other optimality conditions are then equivalent to the equilibrium conditions this implies that the equilibrium allocation is the global optimum among the constraint set of allocations. Notice that we can claim global optimality within this set since the problem is strictly concave and we are maximizing over a convex set.□

Proof of Proposition 12. The size of the team managed by an agent working in layer $\ell > 0$ is given by

$$n^0 = \frac{1}{h(1-Q^{\ell-1})},$$

or 1 if $\ell = 0$. By Proposition (7) the knowledge of this agent is increasing in his ability. By Proposition (5) a higher ability agent is matched with better agents, which implies again more knowledgeable subordinates. Hence, the size n^0 of her team increases with ability if she is a manager. Proposition (10) shows that rank also increases with ability, and since knowledge increases with rank, this makes the positive relationship between team size and ability stronger. Notice that better ability agents are also assigned to higher ability bosses, who in turn have larger teams by the same propositions and logic. Hence, higher ability agents will manage larger teams and will be employed in larger teams. Proposition (6) then yields the result, since it implies that higher ability agents will also receive larger earnings. Hence there is a positive relationship between wages and firm size. □