American Economic Journal: Macroeconomics 2019, 11(1): 49–88
https://doi.org/10.1257/mac.20170390

Learning, Career Paths, and the Distribution of Wages†

By Santiago Caicedo, Robert E. Lucas Jr., and Esteban Rossi-Hansberg*

We develop a theory of career paths and earnings where agents organize in production hierarchies. Agents climb these hierarchies as they learn stochastically from others. Earnings grow as agents acquire knowledge and occupy positions with more subordinates. We contrast these and other implications with US census data for the period 1990 to 2010, matching the Lorenz curve of earnings and the observed mean experience-earnings profiles. We show the increase in wage inequality over this period can be rationalized with a shift in the level of the complexity and profitability of technologies relative to the distribution of knowledge in the population. (*JEL* D83, E24, J24, J31)

This paper develops a new model of an economy that generates sustained productivity growth. One distinctive feature of the model is that all knowledge in the economy is held by the individual people who comprise it: there is no abstract technology hovering above them in the ether. A second feature, necessarily involving heterogeneous labor, is a kind of complementarity in production involving people with different skill levels: the marginal product of any one person is contingent on the people he works with. A third feature, closely related to the second, is that improvements over time in individual skill levels depend on imitation or stimulation or inspiration from other people in the economy. All growth is taken to arise from this force.

Mathematically we will study the behavior of a continuum of agents, each with different skill levels. Production at any date will depend on the way people are organized, on which tasks are assigned to which people at different skill levels. Production increases by any one person will depend on the knowledge of the people with whom he interacts. We add to these features a cohort structure with a stationary birth and death process. The resulting model then has the familiar ingredients of human capital theory and so invites comparisons to census cross-section evidence and age-earnings profiles, to panel evidence on individual careers over time.

*Caicedo: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637 (email: scaicedosoler@uchicago.edu); Lucas: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637 (email: relucas@uchicago.edu); Rossi-Hansberg: Department of Economics, Princeton University, 289 Julis Romo Rabinowitz Building, Princeton, NJ 08544 (email: erossi@princeton.edu). We thank Ezra Oberfield, Mariana Laverde, Chien-Yu Lai, Jeremy Lise, and seminar participants at various institutions and conferences for helpful comments. All authors declare they have no relevant or material financial interests that relate to the research described in this paper.

† Go to https://doi.org/10.1257/mac.20170390 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

After working out the logic of the model, then, we address some of these empirical challenges.

We begin in Section I with a static model built on the hierarchical structures pioneered by Garicano (2000) and Garicano and Rossi-Hansberg (2004, 2006). These models describe a very specific complementarity among agents of different productivity types. Agents can contribute to production as workers or else as managers to whom a worker can turn to for help with a problem that is too hard for him to solve on his own. Each manager can review several unsolved problems that are passed up to him. Those problems that are within the manager's ability are then solved by the worker with his advice. Unsolved problems at this higher layer can be passed up to managers with still higher abilities, and so on.

One can think of successive *layers* of managers. Caliendo, Monte, and Rossi-Hansberg (2015) identifies different occupations within a dataset of French firms with these theoretical layers, and shows that changes in firms' actual occupational structure are well correlated with theoretical changes. It seems natural to think of these complementarities as organized into firms, as in this example, but one can also think of these managers as independent consultants, and workers as buyers of consultant services. What matters for overall production is not the corporate structure in place but the assignment of people of diverse skill levels to different roles in the production process.

For our purposes, the census evidence covers the entire economy and we need a theory at a general equilibrium level. Since our focus is on the labor market, we want to determine the prices—earnings or wages—of people at every skill level. We view every agent as endowed with a given type, which we identify as the goods he could produce on his own.[1] In the hierarchical model we use here, given types are central too, but until we know the occupational structure, we cannot determine earnings. A general equilibrium must be solved to connect types and earnings.

The rest of the paper is structured as follows. We begin with the static problem of solving for equilibrium wages and productivity. In Section I, we set out the notation for this model, work out the assortative matching linking workers and managers, and define the hierarchy that we work with. In Section II, we formulate and study the problem of an idealized planner who wants to maximize total goods production. The planner assigns some agents to be workers, others to be managers, and defines the matching function connecting them. Then in Section III, we decentralize this planning problem to obtain a competitive equilibrium in which the earnings of all managers and workers are determined. In this modeling of assigning agents to tasks, only agents' skill levels matter for earnings. The time and effort that may have led to those levels—years of schooling, age, luck, or any other qualities—are immaterial.

The framework we propose has two key differences relative to the previous work by Garicano and Rossi-Hansberg (2004, 2006). First, it features distributions of skills and problems with unbounded support. Second, and perhaps more important, we assume that problems that are harder to solve also yield higher output. These two features are essential to introduce dynamics and growth into our theory since

[1] We refer to the type of the agent interchangeably as his skill or knowledge level.

they facilitate the incorporation of knowledge diffusion in a setup that features a balanced-growth path. We do this in Section IV where we introduce some dynamics into this model. We introduce a fixed demography, learning, and a cohort structure. Then we interpret the static model developed in Sections I-III as cross sections evolving along a balanced growth path. With these elaborations, we connect the distribution of types studied in Sections II and III to age-earnings profiles.

As in Lucas (2009), we interpret age-earnings profiles as outcomes of a balanced growth path with a constant growth rate driven by a process of learning from others. An agent is identified by his type and at any date there is a distribution of types. Agents learn—improve their types—by interacting with people, some of whom have higher types than their own. The type distribution has a Pareto tail, ensuring that every agent will always find someone who knows more than he does. In the balanced growth path, an agent's career path starts at the bottom of the hierarchy and evolves over time taking him to higher management positions, with more subordinates, and higher wages. These transitions depend on his learning, which is stochastic, and on the distribution of knowledge in the economy. The learning process we use implies that the distribution of knowledge shifts over time, but maintains its shape and the thickness of its tail.

In this paper, however, the type distribution does not, by itself, define production in the economy or its age-earnings profiles. For this we need a second distribution, describing the difficulty of the problems that agents are asked to solve. If these problems were held fixed over time while agent's skills kept improving, all agents would eventually be able to solve virtually all the problems that confront them and there would be no need for a hierarchy. In the same way, a problem distribution that moved too fast to the right would eventually eliminate production. Using the language of Goldin and Katz (2007), the race of skills and technology would be won by one side or the other. In order to maintain a balanced growth, we need to rule out these possibilities. Throughout, though, we are silent about the source of the evolution of the problem distribution and just assume that it evolves in a manner that guarantees the existence of a balanced growth path.

In Section V, we show how we calibrate the few parameters in our parsimonious model to fit the fraction of managers and the distribution of earnings in 1990 and discuss several comparative statics exercises. In Section VI, we compare the balanced growth properties of the calibrated model to age-earnings profiles, as well as the dispersion of the wage distribution across age classes. The model is able to fit the observed age-earnings profile quite well, although it cannot account for the increase in dispersion of wages for older agents. This pattern was also hard to fit in models without a hierarchical structure, such as Lucas (2009).

We finish our analysis in Section VII where we show that, in the calibrated model, shifts in the distribution of the complexity of problems relative to the distribution of knowledge can account well, and parsimoniously, for the increase in wage inequality between 1990 and 2010. The results suggest a cause for the evolution of wages during this period: new technologies that only the most knowledgeable experts can utilize but that also pay more than in the past. The evidence seems to suggest that this technological change was not accompanied by a similar shift in the distribution of knowledge in the population. This technological change leads to relatively large changes in the wage distribution but, in our calibration, to modest declines in de-trended output. Section VIII offers some final remarks and the Appendix presents all the proofs and additional material.

I. Planning Problems: Examples

We consider a closed economy in which people of different types are employed to produce a single consumption good. Everyone is endowed with one unit of work time. The labor force is defined by a cdf $F(z)$ (density f) on $(0, \infty)$, where z indexes a person's skill level. Call F the *population distribution*. An idealized planner organizes these people so as to maximize total production, given a production technology. This technology is modeled closely on earlier work by Garicano and Rossi-Hansberg (2004, 2006) but some of the details are new, and we will provide a self-contained description.

First, consider production under autarky. Every agent z uses his unit of time to generate a production possibility that involves taking a draw from a distribution of problems, defined by a cdf $G(z)$ (density g), which we call the *problem distribution*. If he draws $y \leq z$, then he produces y units of the good, which is to say he solves the problem given to him. If he draws $y > z$, he produces nothing. We think of many people at each z level who produce the average $\int_0^z yg(y)\,dy$. Total production at all levels z is then

$$\int_0^\infty \int_0^z yg(y)\,dy f(z)\,dz.$$

The remaining problems,

$$\int_0^\infty \int_z^\infty yg(y)\,dy f(z)\,dz,$$

are simply discarded. In this example, the planner does nothing except collect all the goods that are individually produced. Clearly, knowledge is not optimally employed; more knowledgeable agents solve the same set of problems than less knowledgeable ones.

Garicano and Rossi-Hansberg (2004, 2006) propose a second role for agents, which we adopt here. Call the agents who generate problems and produce on their own, as described above, *workers*. Other agents, whom we call *managers*, can be used as advisers, using their time to help workers to solve some of the problems that they themselves could not solve. We assume that each manager can review $\kappa \geq 1$ unsolved problems with his endowed unit of time. A problem y that is reviewed by a manager z will be solved if $y \leq z$ and otherwise will remain unsolved. In this more complicated situation, the planner must determine which agents are workers and which are managers, and he needs to assign κ unsolved problems to each manager. We emphasize that no one in this economy—worker, manager, or planner—knows how difficult an individual problem is unless he either generates the problem himself and can solve it or he reviews the problem himself and knows how to enable the worker who drew the problem to solve it. Understanding the difficulty of a problem is thus equivalent to knowing how to solve it.

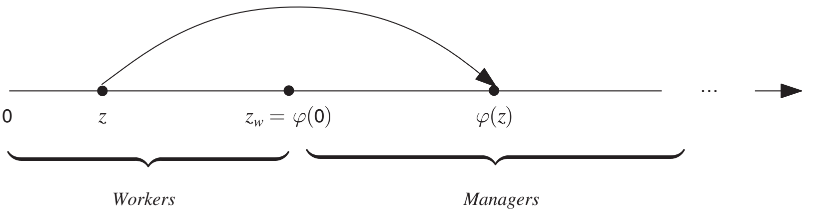

Figure 1. Assignment with One Layer of Managers

Consider then a second example in which agents $z \in [0, z_w]$ are assigned to be workers and the rest, $z \in (z_w, \infty)$, are assigned to be managers in an economy. Figure 1 illustrates this assignment. Each one of the workers z must be assigned to a particular manager $\varphi(z)$.

Suppose the planner chooses a cutoff z_w and a continuous, increasing matching function $\varphi : [0, \infty) \to (z_w, \infty)$ that together satisfy

$$(1) \qquad \kappa\big[F(\varphi(y)) - F(z_w)\big] = \int_0^y \big[1 - G(z)\big] f(z)\, dz, \quad \text{all } y \in [0, z_w].$$

The right side is the number of problems that people in $[0, y)$ cannot solve on their own. The left side is the number of problems that the assigned managers in $(\varphi(0) = z_w, \varphi(y)]$ can review. That is, equation (1) guarantees that there are enough managers to deal with all the problems generated by workers: a sort of market clearing of problems. Clearly, generating problems that will not be reviewed, or allocating unemployed management time, is not optimal. An optimal allocation necessarily satisfies (1).

Now suppose that z_w is chosen so that every unsolved problem in $[0, z_w]$ is reviewed by exactly one manager in (z_w, ∞) and that all managers are so employed. Then equation (1) would imply

$$(2) \qquad 1 - F(z_w) = \frac{1}{\kappa} \int_0^{z_w} \big[1 - G(z)\big] f(z)\, dz.$$

In this case, total production would become

$$(3) \qquad \int_0^{z_w} \int_0^z y g(y)\, dy\, f(z)\, dz + \int_0^{z_w} \int_z^{\varphi(z)} y g(y)\, dy\, f(z)\, dz$$
$$= \int_0^{z_w} \int_0^{\varphi(z)} y g(y)\, dy f(z)\, dz.$$

This is a feasible plan, but is it optimal?

The reason this plan might not be optimal is that all problems above $\varphi(z)$ for $z \in (0, z_w]$ are discarded. Could it be the case that the planner prefers to have some of the more knowledgeable agents deal with the problems that the less knowledgeable

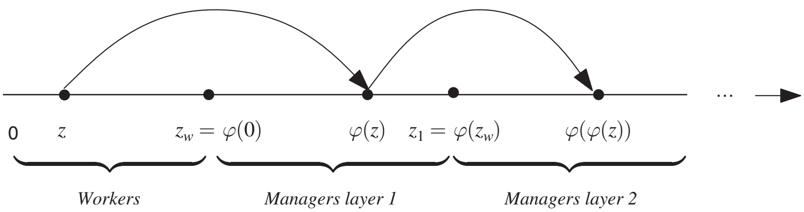

Figure 2. Assignment with One Layer of Managers

managers discard? Maybe it is optimal to have managers solve other manager's discarded problems. To allow for this possibility, let $z_1 = \varphi(z_w)$ and impose the following condition for managers performing this role,

$$F(\varphi(y)) - F(z_1) = \int_{z_w}^{y} \frac{1 - G(z)}{1 - G(\varphi^{-1}(z))} f(z)\, dz, \quad \text{for all } y \in (z_w, \varphi(z_w)].$$

The right side is the number of problems that managers in (z_w, y) cannot solve (conditional on workers below them not being able to solve them either, $1 - G(\varphi^{-1}(z))$) and thus are passed up to managers in the second layer, (z_1, y). Again, if this condition is not satisfied, a manager's time is clearly not allocated efficiently. As we illustrate in Figure 2, the resulting assignment matches each workers in $z \in [0, z_w]$ to a layer-one manager in $\varphi(z) \in [\varphi(0), \varphi(z_w)]$, and in turn, each one of the managers in layer one to a layer-two manager in $\varphi(\varphi(z)) \in [\varphi(\varphi(0)), \infty)$.

In general, a matching function that includes agents that solve the problems discarded by $i + 1$ groups of agents will have to satisfy

$$(4) \quad F(\varphi(y)) - F(z_{i+1}) = \int_{z_i}^{y} \frac{1 - G(z)}{1 - G(\varphi^{-1}(z))} f(z)\, dz, \quad \text{for all } y \in [z_i, z_{i+1}],$$

where

$$(5) \qquad z_{i+1} = \varphi(z_i), \quad \text{for } i = 1, 2, \ldots.$$

Following the practice of Garicano and Rossi-Hansberg (2004, 2006), we refer to the intervals (z_i, z_{i+1}) with $z_0 = z_w$ as managerial layers, denoted $L_1, L_2, \ldots$.

All of these task assignments of the given labor force $F(z)$ are feasible, as are many other assignments not yet discussed.[2] But which, if any, of these can maximize total production? This question is answered in Section II.

[2] For example, in all the examples above we assumed that agents generating problems are the least knowledgeable ones and that the interval of workers is connected. Clearly, there exists other feasible allocations that do not satisfy these properties.

II. Planning Problems: Optimality

To consider all possible assignments, we need a better notation. Let $\mathcal{Z}_w \subset [0, \infty)$ denote the set of agents that the planner assigns to the role of generating problems (workers). Let $\mathcal{Z}_m \subset [0, \infty)$ denote the set of managers with subordinates and a manager above them. Let $\mathcal{Z}_e \subset [0, \infty)$ denote the set of top managers without a manager above them (whom we call entrepreneurs). Note that $\mathcal{Z}_w \cap \mathcal{Z}_m = \varnothing$ and $\mathcal{Z}_m \cap \mathcal{Z}_e = \varnothing$ but $\mathcal{Z}_w \cap \mathcal{Z}_e$ could be non-empty since some workers might be self-employed. Furthermore, $\mathcal{Z}_w \cup \mathcal{Z}_m \cup \mathcal{Z}_e = [0, \infty)$: all agents belong to one of these sets.

As shown in Garicano and Rossi-Hansberg (2006), the production process described above implies a technology that is supermodular in the talent of an agent and its subordinates. This implies that the assignment function, $\varphi : (\mathcal{Z}_w \backslash \mathcal{Z}_e) \cup \mathcal{Z}_m \to \mathcal{Z}_m \cup \mathcal{Z}_e$, is monotone increasing. With some abuse of notation, let $\varphi(\mathcal{Z})$ denote the set of managers assigned to a set of agents $\mathcal{Z}$.

The problem of a planner is then given by

$$\max_{\{\varphi(\cdot)\}_{(\mathcal{Z}_w \backslash \mathcal{Z}_e) \cup \mathcal{Z}, \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e}} \int_{\mathcal{Z}_w} \int_0^{\infty} y g(y)\, dy\, f(z)\, dz - \kappa \int_{\mathcal{Z}_e} \int_z^{\infty} y \frac{g(y)}{1 - G(\varphi^{-1}(z))}\, dy\, f(z)\, dz, \tag{6}$$

subject to

$$\int_{\varphi(Y)} f(z)\, dz = \frac{1}{\kappa} \int_Y [1 - G(z)] f(z)\, dz, \quad \text{all } Y \subset \mathcal{Z}_w \backslash \mathcal{Z}_e, \tag{7}$$

and

$$\int_{\varphi(Y)} f(z)\, dz = \int_Y \frac{1 - G(z)}{1 - G(\varphi^{-1}(z))} f(z)\, dz, \quad \text{all } Y \subset \mathcal{Z}_m, \tag{8}$$

where the first term in the objective function (6) is the value of all the generated problems if all of them are solved, and the second term is the value of the discarded problems. The constraints just replicate the ones we introduced in equations (1), (4), and (5). Note that, given the sets $\mathcal{Z}_w$, $\mathcal{Z}_m$, and $\mathcal{Z}_e$, the two constraints (7) and (8) fully specify the assignment function φ. Hence, the problem of the planner can be reduced to choosing the sets of agents that become workers, managers, and entrepreneurs.

In this section, we report what we can establish about the solution to this planning problem. The next theorem shows that in the optimal allocation, the sets of workers, managers (if non-empty), and entrepreneurs are connected and ordered by their skill level, where workers are the least-skilled agents and entrepreneurs the most-skilled ones. The results allows us to eliminate many feasible allocations as solutions. All proofs are relegated to Appendix A.

THEOREM 1: *Suppose that* $F(\cdot)$ *and* $G(\cdot)$ *are differentiable. Then, any solution to the planning problem is characterized by a matching function* $\varphi(\cdot)$, *and a pair of thresholds* z_w *and* z_e *such that* $\mathcal{Z}_w = [0, z_w]$, $\mathcal{Z}_m = [z_w, z_e]$, and $\mathcal{Z}_e = [z_e, \infty)$. *If* $\mathcal{Z}_m \neq \varnothing$, *then* $z_w < z_e$; *otherwise*, $z_w = z_e$.

The proof of Theorem 1 relies on the next three lemmas.

LEMMA 1: *Let* $\left(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e\right)$ *solve the planner's problem. Then for* ε *sufficiently small* $[0, \varepsilon) \nsubseteq \mathcal{Z}_w \cap \mathcal{Z}_e$.

LEMMA 2: *Let* $\left(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e\right)$ *solve the planner's problem. Then if* $z_m \in \mathcal{Z}_m$ *and* $z_e \in \mathcal{Z}_e$, *we have* $z_m < z_e$.

LEMMA 3: *Let* $\left(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e\right)$ *solve the planner's problem. Then* $z \in \mathcal{Z}_w$ *implies that* $z \leq z'$ *for all* $z' \in \mathcal{Z}_m \cup \mathcal{Z}_e$.

Together, these three lemmas imply that the optimal allocation is such that $\mathcal{Z}_w$, $\mathcal{Z}_m$, and $\mathcal{Z}_e$ are ordered with $\mathcal{Z}_w \leq \mathcal{Z}_m \leq \mathcal{Z}_e$, where $\mathcal{Z}_m$ is potentially empty.[3] This implies that in any optimal allocation with a maximum number of layers L, the set of hierarchies active in the allocation has either L or $L-1$ layers. For example, in an economy where the maximum number of layers is three, there cannot be any self-employed agents working on their own, a feature also present in Garicano and Rossi-Hansberg (2006).

Theorem 1 implies that the problem of the planner can then be rewritten as

$$\max_{\{\varphi(\cdot)\}_{[0, z_e]}, (z_w, z_e)} \int_0^{z_w} \int_0^{\infty} y g(y)\, dy\, f(z)\, dz - \kappa \int_{z_e}^{\infty} \int_z^{\infty} y \frac{g(y)}{1 - G(\varphi^{-1}(z))}\, dy\, f(z)\, dz, \tag{9}$$

subject to

$$F\big(\varphi(y)\big) - F\big(\min\{z_w, z_e\}\big) = \frac{1}{\kappa} \int_0^y \big[1 - G(z)\big] f(z)\, dz, \tag{10}$$

$$\text{all } y \in \big[0, \min\{z_w, z_e\}\big],$$

and

$$F\big(\varphi(y)\big) - F\big(\varphi(\min\{z_w, z_e\})\big) = \int_{\min\{z_w, z_e\}}^{y} \frac{1 - G(z)}{1 - G(\varphi^{-1}(z))}\, f(z)\, dz, \tag{11}$$

$$\text{all } y \in \big[\min\{z_w, z_e\}, z_e\big].$$

[3] We use $\mathcal{Z}_w \leq \mathcal{Z}_m \leq \mathcal{Z}_e$ to mean that for any $z_w \in \mathcal{Z}_w$, any $z_m \in \mathcal{Z}_m$, and any $z_e \in \mathcal{Z}_e$, $z_w \leq z_m \leq z_e$.

A solution to the planning problem is thus fully described by a matching function $\varphi(\cdot) : [0, z_e] \rightarrow [z_w, \infty)$ and two cutoffs (z_w, z_e).

As we mentioned above, for any pair of cutoffs, we can construct the matching function φ using (10) and (11). That is, provided that $F(\cdot)$ is strictly increasing, we have

$$\varphi(y) = F^{-1}\Big(F(\min\{z_w, z_e\}) + \frac{1}{\kappa}\int_0^y [1 - G(z)] f(z)\, dz\Big), \quad \text{all } y \in [0, \min\{z_w, z_e\}]$$

and

$$\varphi(y) = F^{-1}\left(F\big(\varphi(\min\{z_w, z_e\})\big) + \int_{\min\{z_w, z_e\}}^y \frac{1 - G(z)}{1 - G(\varphi^{-1}(z))} f(z)\, dz\right),$$

$$\text{all } y \in [\min\{z_w, z_e\}, z_e].$$

Moreover, if we assume that $F(\cdot)$ is differentiable, then $\varphi(\cdot)$ is continuous and differentiable except possibly at $\min\{z_w, z_e\}$, where there is a kink in the matching function if $\kappa \neq 1$.

In light of these facts, our task is to find the cutoffs (z_w, z_e) that solve the planning problem. We first consider the situation where there are no intermediate managers, or where $\mathcal{Z}_m = \varnothing$. In this case, $z_w = z_e$ and only one cutoff needs to be determined. Note further that in this case $\varphi(z_w) = \infty$ and so (10) implies, for $y = z_w$, that equation (2) holds. That is,

$$1 - F(z_w) = \frac{1}{\kappa}\int_0^{z_w} [1 - G(z)] f(z)\, dz.$$

Clearly, violating this constraint is not optimal since it implies that either too many problems are generated, or there is excess managerial capacity to review problems. Lemma 4 in the Appendix makes this argument precise. Hence, when $z_w = z_e$, equation (2) determines the value of z_w that maximizes output. In turn, $\varphi(\cdot)$ is determined by (10) for $y < z_w$.

Focus next on the case where $\mathcal{Z}_m \neq \varnothing$ and so $z_w < z_e$. Then we can restate the problem of the planner as

$$\max_{(z_w, z_e)} \int_0^{z_w}\int_0^{\infty} yg(y)\, dy\, f(z)\, dz - \kappa \int_{z_e}^{\infty}\int_z^{\infty} yg(y)\, dy \frac{f(z)}{1 - G(\varphi^{-1}(z, z_w))}\, dz \tag{12}$$

subject to

$$F\big(\varphi(y)\big) - F(z_w) = \frac{1}{\kappa}\int_0^y [1 - G(z)] f(z)\, dz, \quad \text{all } y \in [0, z_w] \tag{13}$$

and

$$F\big(\varphi(y)\big) - F\big(\varphi(z_w)\big) = \int_{z_w}^y \frac{1 - G(z)}{1 - G(\varphi^{-1}(z, z_w))} f(z)\, dz, \quad \text{all } y \in [z_w, z_e]. \tag{14}$$

We have shown in Lemmas 1–4 that an optimal plan must satisfy (12)–(14). We now add the following result.

THEOREM 2: *Suppose that $F(\cdot)$ and $G(\cdot)$ are differentiable. Then there is a pair (z_w, z_e) that satisfies* (12)–(14).

This argument is not constructive and does not address uniqueness, so in Appendix D we describe the numerical algorithm we use to calculate maxima. In particular, we compare solutions in which all problems are eventually solved and so $z_e = \infty$ and solutions in which z_e is finite and so some problems are optimally discarded. We also solve the case with $z_w = z_e$. We can then compare total production in these allocations and easily compute the maximum.

III. Decentralized Equilibrium

The planning problem studied in Section II was a step toward obtaining a competitive equilibrium. To this end, we next describe a decentralization of the planner's allocation $(\varphi(\,\cdot\,), z_w, z_e)$, given F, G, and κ. We are looking for a wage function that gives incentives to firms to build the same production teams as the planner. Given that agents are income maximizers, they just choose the job that is willing to pay more for their services, conditional on their skill.

We consider organizations defined by a single worker $z \in [0, z_w]$ that hires managers at levels

$$\varphi(z), \quad \varphi\big(\varphi(z)\big) = \varphi_2(z), \ldots, \varphi_L(z),$$

where L denotes the maximum number of layers. This organization generates one problem with an expected return of

$$\int_0^{\varphi_{L_z}(z)} y g(y)\, dy,$$

where $\varphi_L(z)$ is the top manager matched to worker z. We need to price the labor time of these managers $w(\varphi_i(z)), i = 1, \ldots, L$, and the residual value $w(z)$ of worker z. The units of manager-of-layer-i's time that the firm of worker z hires, $n_i(z)$, is equal to the fraction of problems that subordinates one layer below cannot solve, relative to the number of problems a manager can handle, κ. That is,

$$n_i(z) = \frac{1 - G(\varphi_{i-1}(z))}{\kappa}, \quad \text{all } i = 1, \ldots, L, \tag{15}$$

where $\varphi_0(z) \equiv z$ is the identity function. Let L be the maximum number of managerial layers consistent with the planner's solution $(\varphi(\,\cdot\,), z_w, z_e)$. Firms in such an economy can then have either L or $L - 1$ layers.

The profit of an organization with worker $z \in [0, z_w]$ and L_z layers is given by

$$\Pi(z) = \int_0^{\varphi_{L_z}(z)} y g(y)\, dy - \sum_{i=1}^{L_z} \frac{1 - G(\varphi_{i-1}(z))}{\kappa} w(\varphi_i(z)) - w(z).$$

The problem of this organization is

$$\max_{\varphi(z), L_z} \Pi(z). \tag{16}$$

The associated first-order conditions are

$$\begin{aligned} w'(z) &= \frac{g(z)}{\kappa} w(\varphi(z)), \\ w'(\varphi_i(z)) &= \frac{g(\varphi_i(z))}{1 - G(\varphi_{i-1}(z))} w(\varphi_{i+1}(z)) \quad \text{for } i = 1, \ldots, L_z - 1, \\ w'(\varphi_{L_z}(z)) &= \frac{\kappa}{1 - G(\varphi_{L_z - 1}(z))} \varphi_{L_z}(z)\, g(\varphi_{L_z}(z)). \end{aligned} \tag{17}$$

The equilibrium wage schedule $w(z)$ is therefore calculated by solving the following differential equations, starting from the last layer in the economy, L,

$$\begin{aligned} w'(z) &= \frac{g(z)}{\kappa} w(\varphi(z)) && \text{for } z \in [0, z_w], \\ w'(z) &= \frac{g(z)}{(1 - G(\varphi^{-1}(z)))} w(\varphi(z)) && \text{for } z \in [z_w, z_e], \\ w'(z) &= \frac{\kappa g(z)\, z}{1 - G(\varphi^{-1}(z))} && \text{for } z \in [z_e, \infty). \end{aligned} \tag{18}$$

The matching function and the cutoffs $(\varphi(\cdot), z_w, z_e)$ are given by the solution to the planner's problem. The choice of L_z is a discrete choice problem that also coincides with the hierarchies present in the planner's allocation, namely $L - 1$ or L.

Workers with knowledge given by $\varphi_i(0)$, $i = 1, \ldots, L$, should be indifferent between working in layer $i - 1$ or layer i, since otherwise some agents close to them would like to work for a different employer, which would imply that $\varphi(\cdot)$ cannot be continuous. Since the planner chooses a continuous matching function $\varphi(\cdot)$, in equilibrium the wage function needs to be continuous as well. Hence,

$$\lim_{z \uparrow \varphi_i(0)} w(z) = \lim_{z \downarrow \varphi_i(0)} w(z) \quad \text{all } i = 1, \ldots, L. \tag{19}$$

Furthermore, the profits of the least productive worker/organization $\Pi(0)$ have to be equal to zero since workers are the residual claimants of the firm. Namely,

$$\int_0^{\varphi_L(0)} y g(y)\, dy = \sum_{i=1}^{L} \frac{1 - G(\varphi_{i-1}(0))}{\kappa} w(\varphi_i(0)) + w(0). \tag{20}$$

The conditions in (19) and (20) provide the L initial conditions that we need to solve the system of first-order differential equations in (18).

Note that the zero profit condition together with (16) and (17) imply that the profits for any organization $z \in [0, z_w]$ are zero, since for $i \geq 1$,

$$\begin{aligned}\frac{\partial \Pi(z)}{\partial z} &= \left(\varphi_L(z)\, g(\varphi_L(z)) - \frac{(1 - G(\varphi_{L-1}(z)))}{\kappa} w'(\varphi_L(z))\right) \frac{\partial \varphi_L(z)}{\partial z} \\ &\quad + \cdots \\ &\quad + \left(\frac{g(\varphi_i(z))\, w(\varphi_{i+1}(z))}{\kappa} - \frac{(1 - G(\varphi_{i-1}(z)))}{\kappa} w'(\varphi_i(z))\right) \frac{\partial \varphi_i(z)}{\partial z} \\ &\quad + \cdots \\ &\quad + \left(\frac{g(z)\, w(\varphi(z))}{\kappa} - w'(z)\right) = 0.\end{aligned}$$

It should be clear that what we have called an "organization" in this section is only one of many possibilities. Our choice to consider organizations defined by a single worker $z \in [0, z_w]$ was adopted only as a convenience. This is a constant returns competitive equilibrium model in which profits are zero. We have developed a production technology involving a specific kind of complementarity about workers with different skill levels, not a theory of the firm. In equilibrium, individual earnings depend jointly on an agent's type and on the way he contributes to the productivity of others. This is all internalized—there are no external effects—but there are no distinctions between firms of different sizes or between employees and consultants.

IV. Age-Earnings Profiles: Theory

We have developed a Garicano/Rossi-Hansberg type model that generates a model of individual earnings for the economy as a whole, $w(z)$. Now we merge these ideas with a model that admits stratification by age and other attributes. The knowledge distribution F now evolves over time, so we write $F(z, t)$, and we introduce a family of cohort-type distributions $H(z, s, t_b)$, interpreted as the probability that a person of age s born at date $t_b = t - s$ will have a type no larger than z.

The distributions F and H and, as we will see, the problem distribution $G(z, t)$, will set the stage for a model of growth. We suppose that age is continuously distributed on $s \in [0, \infty]$ with stationary distribution $\pi(s)$. In particular, we assume a continuous arrival model of diffusion of ideas, in which agents have $\alpha\pi(s)\, ds$ meetings per unit of time that take the form of draws from individuals of all ages. An agent z who engages with an agent of type z' emerges from the match with type $\max(z, z')$.

If h is a small time interval, then the probability that a person of age $s + h$ born at t_b has type less than or equal to z is[4]

$$
\begin{aligned}
H(z, s+h, t_b) &= H(z, s, t_b)\,\Pr\{\text{no draw at } t_b + s \text{ better than } z\} \\
&= H(z, s, t_b) \prod_0^{\infty} H(z, \tilde{s}, t_b + s - \tilde{s})^{\alpha h \pi(\tilde{s}) d\tilde{s}}.
\end{aligned}
$$

Taking logs on both sides of the equation, dividing by h, and letting $h \to 0$,

$$
\frac{\partial \log H(z, s, t_b)}{\partial s} = \alpha \int_0^{\infty} \log H(z, \tilde{s}, t_b + s - \tilde{s})\, \pi(\tilde{s})\, d\tilde{s}.
$$

The right side is just α times an average of the cdfs $H(z, \tilde{s}, t_b + s - \tilde{s})$, weighted by the densities $\pi(\tilde{s})$ on $(0, \infty)$. We assume that initially the agent knows nothing, $\log H(z, 0, t_b) = 0$. Then integrating this last expression from 0 to s, we get

$$
\log H(z, s, t_b) = \alpha \int_0^{s} \left(\int_0^{\infty} \log H(z, \tilde{s}, t_b + \tau - \tilde{s})\, \pi(\tilde{s})\, d\tilde{s} \right) d\tau,
$$

which we can write in calendar time t as

$$
\log H(z, s, t - s) = \alpha \int_0^{s} \left(\int_0^{\infty} \log H(z, \tilde{s}, t - s + \tau - \tilde{s})\, \pi(\tilde{s})\, d\tilde{s} \right) d\tau. \tag{21}
$$

For the individuals of the cohort born at $t - s$, the inner integral denotes the best match at age τ and the outer integral the best match up to age s.

We would like to find a family of distributions that solves equation (21) on a balanced growth path. More precisely, assume that all quantiles of the knowledge distributions $H(z, s, t - s)$ grow at the same rate γ and define the invariant distribution $\Gamma(\,\cdot\,, s)$ by $H(z, s, t - s) = \Gamma(z e^{-\gamma t}, s)$. Let $x = z e^{-\gamma t}$ denote the de-trended knowledge and rewrite equation (21) in terms of the balanced growth path (BGP) distribution

$$
\log \Gamma(x, s) = \alpha \int_0^{s} \left(\int_0^{\infty} \log \Gamma(x e^{\gamma(s - \tau)}, \tilde{s})\, \pi(\tilde{s})\, d\tilde{s} \right) d\tau. \tag{22}
$$

It is convenient for our purposes to assume that the distributions Γ are Fréchet, so that

$$
\log \Gamma(x, s) = -\mu(s)\, x^{-1/\theta}.
$$

[4] $\prod_0^{\infty}$ denotes the "product integral" defined as follows: let $f(x)$ be any function,

$$
\prod_a^b f(x)^{dx} = \prod_{\Delta x \to 0} f(x_i)^{\Delta x} := \exp\left(\int_a^b \log(f(x))\, dx \right).
$$

See Slavík (2007) for a summary and history of product integration.

The Fréchet distributions have the property that they will be preserved under the operations we use. As long as the parameter θ is positive, individual careers will involve learning from others, and on average, older workers will have higher z values than younger ones. Then (22) implies

$$-\mu(s)\, x^{-1/\theta} \;=\; \alpha \int_0^s \left(\int_0^\infty -\mu(\tilde{s})\, x^{-1/\theta}\, e^{-(\gamma/\theta)(s-\tau)}\, \pi(\tilde{s})\, d\tilde{s} \right) d\tau,$$

and canceling gives

$$\mu(s) \;=\; \left(\int_0^\infty \mu(\tilde{s})\, \pi(\tilde{s})\, d\tilde{s} \right) \frac{\alpha\theta}{\gamma} \left(1 - e^{-(\gamma/\theta)\, s}\right). \tag{23}$$

Note that if $\mu(\,\cdot\,)$ solves (23), any proportionally scaled version of $\mu(\,\cdot\,)$ solves the functional equation as well. Hence, let us normalize $\mu(\,\cdot\,)$ such that $\int_0^\infty \mu(\tilde{s})\, \pi(\tilde{s})\, d\tilde{s} \;=\; 1$. Integrating both sides of equation (23) with density $\pi(s)$ results in

$$\int_0^\infty \mu(s)\, \pi(s)\, ds \;=\; \left(\int_0^\infty \mu(\tilde{s})\, \pi(\tilde{s})\, d\tilde{s} \right) \int_0^\infty \frac{\alpha\theta}{\gamma} \left(1 - e^{-(\gamma/\theta)\, s}\right) \pi(s)\, ds.$$

Simplifying yields an expression that determines implicitly the growth rate, γ, given by

$$\frac{\gamma}{\theta} \;=\; \alpha \int_0^\infty \left(1 - e^{-(\gamma/\theta)\, s}\right) \pi(s)\, ds. \tag{24}$$

Summarizing, we have shown that on a BGP, a solution to the system of functional equations defined by (21) is obtained by choosing H to be a Fréchet distribution

$$\log H(z, s, t - s) \;=\; -\mu(s)\, e^{(\gamma/\theta)\, t}\, z^{-1/\theta}$$

with location parameter $\mu(s) \;=\; \left(\alpha\theta/\gamma\right)\left(1 - e^{-(\gamma/\theta)\, s}\right)$, where γ solves equation (24). Using this distribution, we can compute the knowledge distribution for individuals of *any* age that we used in our static model,

$$\begin{aligned} F(z, t) \;&=\; \int_0^\infty H(z, s, t - s)\, \pi(s)\, ds \\ &=\; \int_0^\infty \exp\left(-\mu(s) \left(z e^{-\gamma t}\right)^{-1/\theta} \right) \pi(s)\, ds \\ &=\; \Phi(z e^{-\gamma t}). \end{aligned} \tag{25}$$

So $F(z, t)$ is the average of Fréchet distributions, with the invariant distribution of types in the balanced growth path $\Phi(\,\cdot\,)$ given by

$$\Phi(x) \;=\; \int_0^\infty \exp\left(-\mu(s)\, x^{-1/\theta} \right) \pi(s)\, ds.$$

Given this invariant distribution of types in the balanced growth path, we will assume that the problem distribution is also an average of Fréchet distributions with potentially a different location parameter, K, but with the same tail parameter θ. The value of K measures the location of the distribution of problems relative to the distribution knowledge in the population (given our normalization of $\int_0^{\bar{s}} \mu(\tilde{s})\,\pi(\tilde{s})\,d\tilde{s} = 1$). Namely,

$$G(z,t) = \int_0^{\infty} \exp\Big(-K\mu(s)\left(z e^{-\gamma t}\right)^{-1/\theta}\Big)\,\pi(s)\,ds \tag{26}$$

$$= \Phi(K^{-\theta} z e^{-\gamma t})$$

$$= F\big(K^{-\theta} z, t\big).$$

Along a BGP, the date t is immaterial, and we can compute the mean and the variance of log wages as

$$W(s) = C + \int_0^{\infty} \log w(z)\; de^{-\mu(s) z^{-1/\theta}} \tag{27}$$

and

$$V(s) = \int_0^{\infty} \big(\log w(z) + C - W(s)\big)^2\; de^{-\mu(s) z^{-1/\theta}}, \tag{28}$$

where C is an arbitrary constant to be determined by the level of log wages observed in the data.

V. Calibration and Comparative Statics

The implications of the decentralized model of Section III involve only types z and the implied earnings $w(z)$. They do not depend on age, schooling levels, or other individual characteristics. Given information on the population F, the problems G that agents must solve in order to produce, and the ability κ of managers to advise in solving harder problems, we can determine the occupations and earnings of everyone in the economy. Here, we use the $F(z,t)$ and $G(z,t)$ that satisfy the balanced growth conditions discussed in the previous section. So let

$$F(\lambda(t)\, z, t) = \Phi(z)$$

and

$$G(K(t)\, z, t) = \Phi(z).$$

That is, as discussed above, we assume that the shape parameter θ of the underlying cohort distributions is the same for the knowledge and problem distribution.

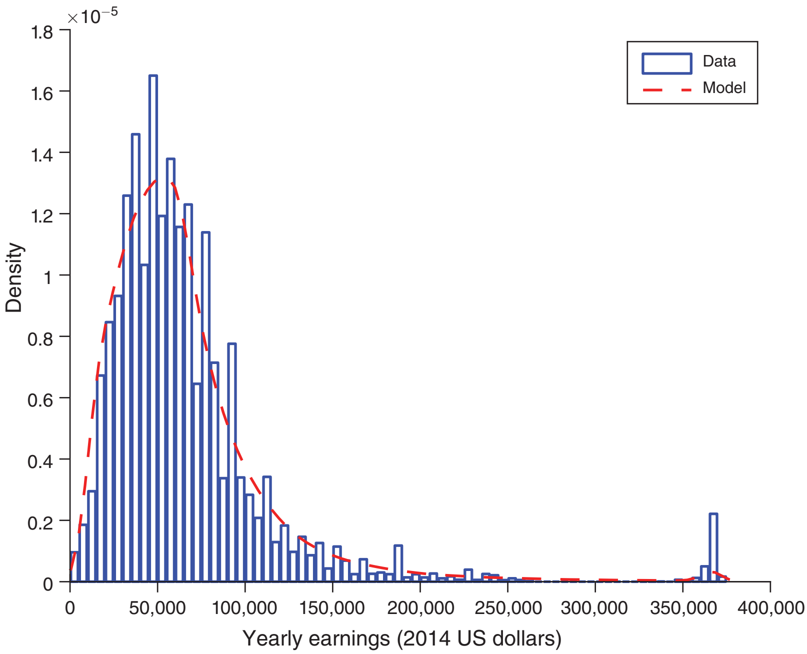

Figure 3. 1990 Distribution of Earnings and Model Fit

This facilitates some of the intuition of the comparative statics and given that we are not addressing the economic underpinnings of the distribution of problems G, any other choice would have been arbitrary as well. We also assume an economy on a balanced growth path, with $\lambda(t) = e^{\gamma t}$ and $K(t) = K^{\theta} e^{\gamma t}$. Hence, $F(z,t) = \Phi\left(ze^{-\gamma t}\right)$ and $G(z,t) = \Phi\left(K^{-\theta} ze^{-\gamma t}\right)$. On the balanced growth path, then, the distribution of knowledge and problems grow at the same rate, γ. That is, on a balanced growth path, the race between knowledge and technology is always tied.

We calibrate the model by setting $K = 0.59, \theta = 0.28$, and $\kappa = 1.2$. These parameters, together with C and α, are chosen to match the degree of inequality as measured by the United States. We match the Lorenz curve in 1990, the observed number of first layer managers in the data, and the mean of the wage distribution across experience levels. We discuss these statistics and the earnings data in the next section. In Figure 3, we present the resulting distribution of wages implied by the model as well as a histogram of the US earnings data in 1990.[5] The fit is, we believe, quite good. The model implied distribution is somewhat smoother than the data, although this is probably partly the result of the bins chosen to draw the empirical distribution. Of course, the model can also not capture fully the bump in the empirical histogram that results from top coding of the highest incomes. In terms of statistics, the model captures the 90/50 percentile ratio of 2 exactly. The 50/10

[5] We use 1990 and 2010 census data from IPUMS-USA; see Ruggles et al. (2000).

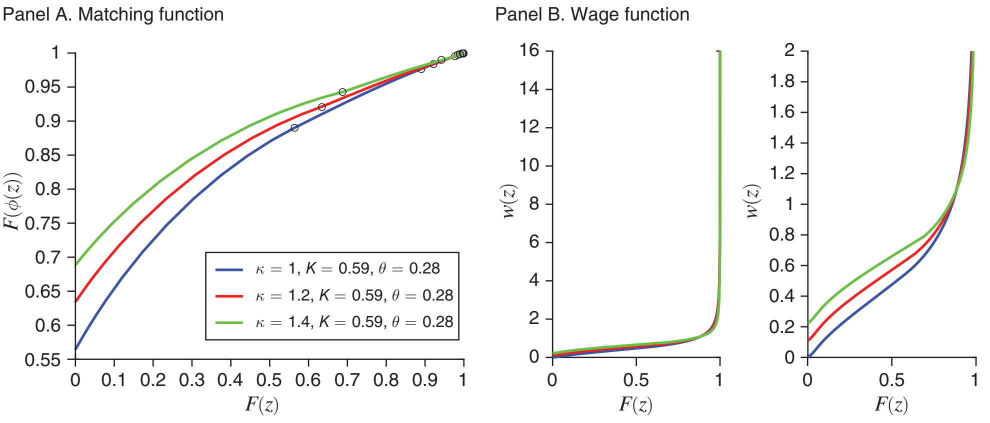

Figure 4. Comparative Statics for κ, Matching and Wage Functions

ratio is 2.3 in the data and 2.5 in the model. Appendix E shows the fit between the model and data by experience for 1990 and 2010.

Figures 4 to 6, show the resulting matching and wage functions for the chosen parameter configuration, as well as the corresponding equilibria when we vary κ, K, and θ. The matching functions show the relation of types $F(z)$ (on the horizontal axis) to the managers assigned to them, $F(\varphi(z))$. The small circles identify $F(\varphi(0)), F(\varphi(\varphi(0)))$, etc. Using our baseline parameter values implies that 63 percent of individuals are workers ($F(\varphi(0)) = 0.63$), which is approximately consistent with the evidence in Caliendo, Monte, and Rossi-Hansberg (2015) in French manufacturing, although somewhat larger than what we observe for the whole US economy (about 50 percent). Figure 4 also shows that $F(\varphi(\varphi(0))) - F(\varphi(0)) = 0.92 - 0.63 = 0.29$ is the fraction of agents assigned as layer-one managers, which matches relatively well with a share of 0.32 for the United States.[6]

In Figures 4 to 6, we also show the relation of types to earnings: the wage functions. We plot $w(z)$ against $F(z)$, and the right panel is a vertical enlargement of the lowest types. One can see how the managerial complementarities reward the best managers, making the wage function extremely steep at the top.

An increase in κ, which might be interpreted as an improvement in communication technology, allows managers to review a larger number of problems and so, in equilibrium, increases the number of workers. It also raises the earnings of all workers and many managers, although not all of them. Some of the top agents command lower wages since their talent is less scarce. Overall, an increase in κ increases output and reduces wage inequality. The Gini coefficient of the wage distribution decreases monotonically with κ. These results are presented in Figures 4 and A1 (in the Appendix).

[6] We obtain the fraction of agents employed in different layers after converting the IPUMS-USA occupation data into the French professions and socio-professional categories (PCS) classification used in Caliendo, Monte, and Rossi-Hansberg (2015). In Appendix C, we explain these calculations in more detail.

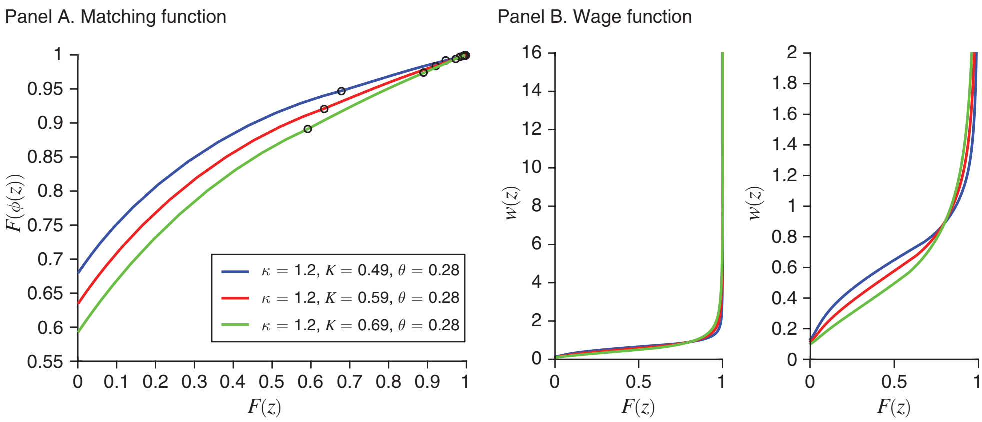

Figure 5. Comparative Statics for K, Matching and Wage Functions

In Figure 5, we present comparative statics with respect to the location of the distribution of problems, K, relative to the distribution of knowledge.[7] Higher K implies that problems are more complicated to solve, relative to the available knowledge in the economy, and also pay more. The result is that the number of workers declines and the number of managers increases. Furthermore, the wages of workers decline since the relative value of generating versus solving problems declines. Inequality at the bottom of the distribution tends to decrease, but at the very top it increases since more problems require the skill of the very best experts. These experts are the ones that obtain the big gains from the more profitable and complicated problems.

The implications of a rise in K are reminiscent of the evidence for the US labor market during the last few decades, as documented by Acemoglu and Autor (2010). An increase in K leads to an increase in output only if there are enough talented agents in the population to take care of the more complicated problems, namely, if K is relatively low or θ is high enough. Otherwise, it can lead to a decline in output since high-z problems, although more profitable, cannot be solved by a large fraction of agents and so, on average, require more managerial time. We return to changes in K in Section VII where we argue that a change in K, on its own, can go a long way toward rationalizing the observed changes in cross-sectional wages in the US economy between 1990 and 2010.

The last comparative static we analyze here involves the tail parameter of the distribution of knowledge *and* the distribution of problems. An increase in θ increases the mass in the tail of these distributions and so increases the availability of hard and profitable problems, as well as the availability of agents that can solve them. Because, by assumption, we are changing both distributions in the same manner, the matching across agents when measured in terms of their position in the distribution

[7] The means of the distribution of problems and knowledge are both growing at rate γ. A change in K generates a one time shift in the mean of the distribution of problems beyond the constant drift in the mean. Hence, with a rise in K, technology becomes more complicated and profitable permanently *relative* to knowledge in the balance growth path.

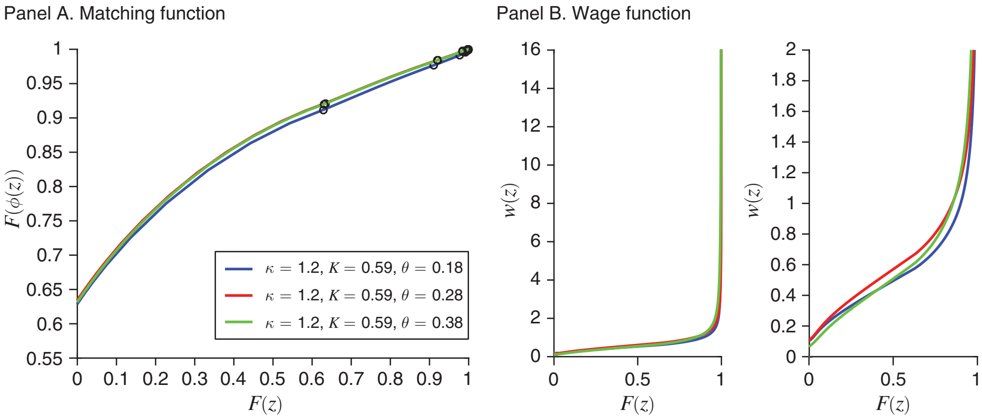

Figure 6. Comparative Statics for θ, Matching and Wage Functions

hardly changes (see Figure 6). The perceived changes are the result of increases in the optimal number of layers of management. As is evident in Figure 6, an increase in θ decreases the wages of workers and increases that of managers. A higher θ implies that the best managers face, and can solve, more complicated problems that pay more. An increase in the tail parameter of the distribution has a similar effect to an increase in K on the distribution of wages, although it has only minor effects on agent's occupational choices. It leads to lower wage inequality among workers but higher inequality across managers. Workers' time becomes less valuable since there are more knowledgeable agents that can solve the problems they cannot solve. An increase in θ also increases total output as it makes problems more profitable but not more difficult relative to the knowledge of the population.

Figures 7 and 8 present the Lorenz curves for the distribution of wages in the 1990 and 2010 US data, as well as the model-based ones for different values of K and θ, respectively.[8] We relegate to the Appendix a similar plot with comparative statics with respect to κ, since raising κ reduces wage inequality in contrast to the clearly evident increase in wage inequality during this period in the United States (Figure A1). The figures plot $\int_0^x w(z)\,dF(z)$ versus $F(x)$ for $x \in [0,\infty)$.[9] The effects discussed in the context of Figures 4 to 6 appear even more clearly. Figure 7 shows the Lorenz curves for 1990 and 2010, together with three alternative calibrations of our model when we let K vary between 0.49 and 0.69. Clearly, the calibration that matches the Lorenz curve in 1990 quite precisely is the one for $K = 0.59$, which is the value we chose for our benchmark calibration. As we argued above, increases in K can generate changes in the Lorenz curves similar to the ones observed in the data. Namely, it increases overall wage inequality, but it reduces income dispersion at the bottom of the income distribution and increases it for the very top managers. In Section VII, we return to this comparative static when we measure the change in K that can rationalize the change in wage inequality during the 1990–2010 period.

[8] We discuss the earnings data we use in detail in Section VI.
[9] We calculate the Lorenz curve using the glcurve command in STATA.

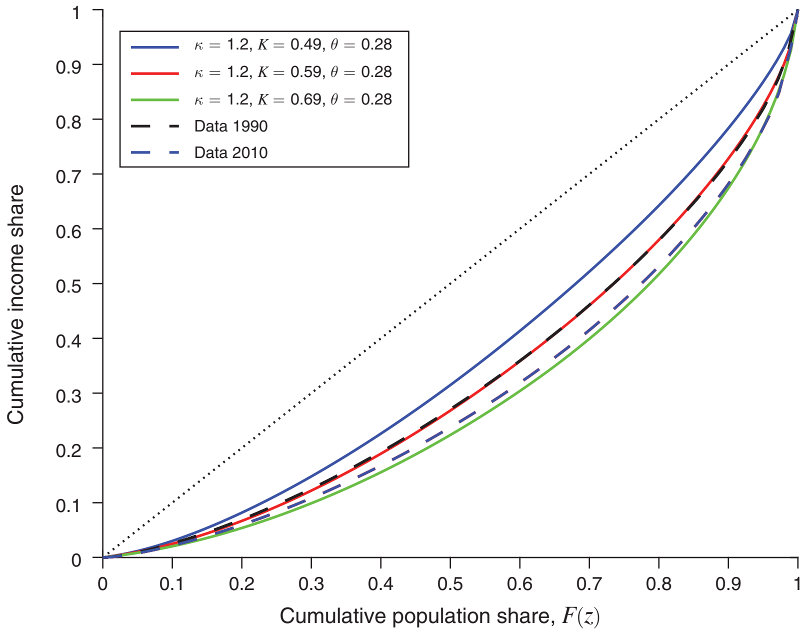

Figure 7. Data and Theory-Based Lorenz Curves for Different Values of K

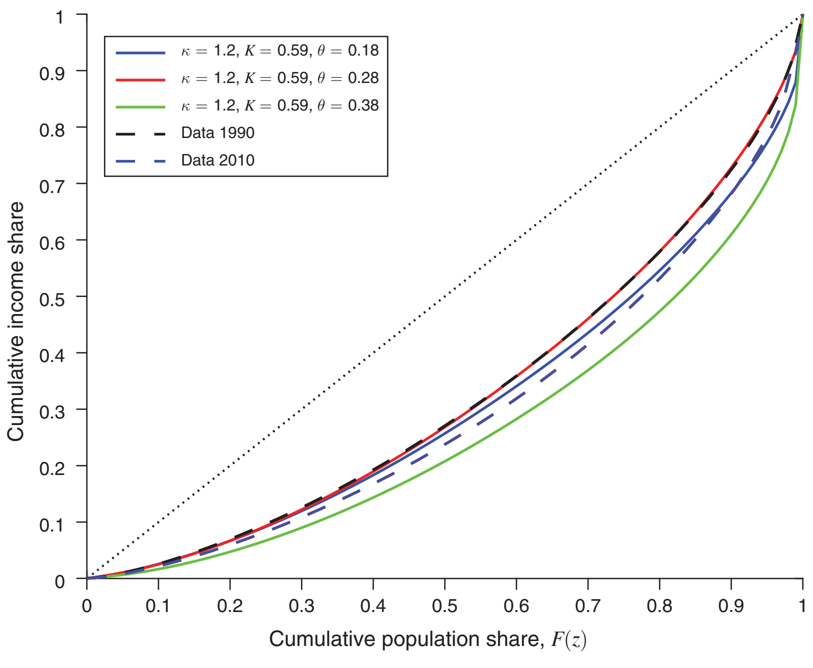

Figure 8. Data and Theory-Based Lorenz Curves for Different Values of θ

Figure 8 presents the Lorenz curves for the benchmark case of $\theta = 0.28$ as well as lower and higher values. An increase in θ yields, as with K, an increase in wage inequality. However, it distorts the Lorenz curve in a way that does not resemble the data. It increases the inequality at the top of the distribution too much. The extra mass at the tail of the distribution simultaneously creates harder and more profitable problems and the agents to solve them. These agents make gains relative to the rest of the distribution of income that look too large relative to the data. The evidence in Figures 7 and 8 shows that observed changes in wage inequality can be explained by the existence of harder and more profitable problems. However, we also need the distribution of knowledge not to catch up fully with these changes, as is the case with a change in K. We come back to this argument in Section VII.

VI. Career Paths

The evidence on wages we use in this paper includes log means and log variances from the 1990 and 2010 IPUMS-USA (Ruggles et al. 2000).[10] We used yearly nominal wages data for white males, heads of households, that worked more than 30 weeks and 30 hours per week the previous year. We focus on full-time workers only since we have not modeled the labor supply decision of agents. We focus on white males that are heads of household in order to control for underemployment by secondary earners and minorities. In Appendix B, we show that the Lorenz curves and experience profiles that result from using an alternative sample with all full-time workers are very similar. Hence, using this alternative sample would only have a negligible effect on our calibration and no effect on any of our conclusions. The 1990 census data correspond to a sample of 5 percent of the US population, while the 2010 survey contains the compiled American Community Survey (ACS) 1 percent sample for 5 years (2006–2010).[11] Of course, good data are available for many decades prior to 1990, and there are many more categories of workers than those we use here.

In order to compare the implications of the model to the data of wages of a worker's life cycle, in this section we carry forward the function $w(z)$ that maps types into earnings, as well as the parameters $K = 0.59$, $\theta = 0.28$, and $\kappa = 1.2$ that we used in Section V. To calibrate the moments (27) and (28), we use the parameters of (23), namely,

$$\mu(s) = \frac{\alpha\theta}{\gamma}\left(1 - e^{-(\gamma/\theta)s}\right).$$

In the balanced growth path, the growth rate of the economy is given by γ. In the United States, this rate is about 0.02, so we let $\gamma = 0.02$. We can then use equation (24) together with a uniform distribution over experience levels to obtain α. That is, we let $s \in [0, 44]$ and $\pi(s) = 1/44$. Then equation (24) implies that $\alpha = 0.1027$.

[10] We have also analyzed the data for 2000 and found similar patterns.

[11] In the survey, nominal income is measured in 2010 dollars.

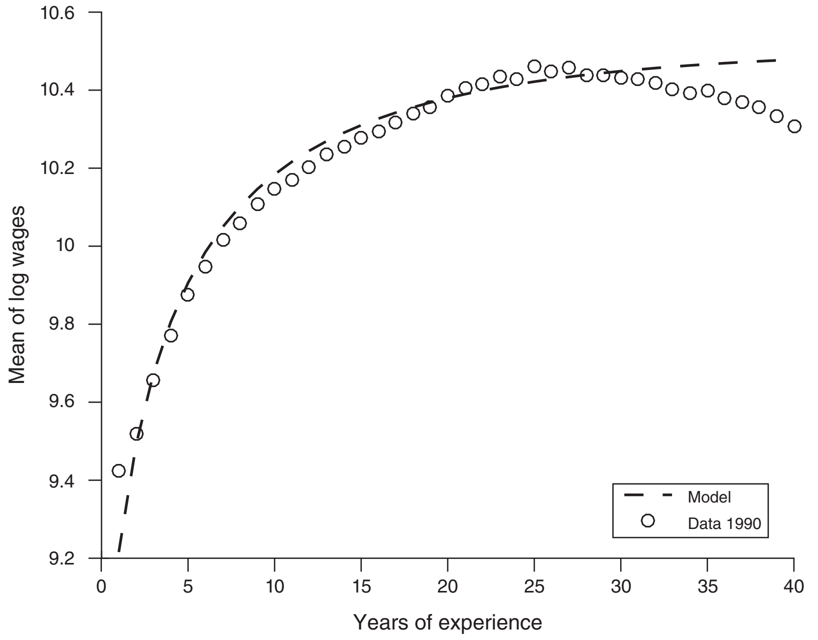

Figure 9. Mean log Wage and Experience: Model and 1990 Data

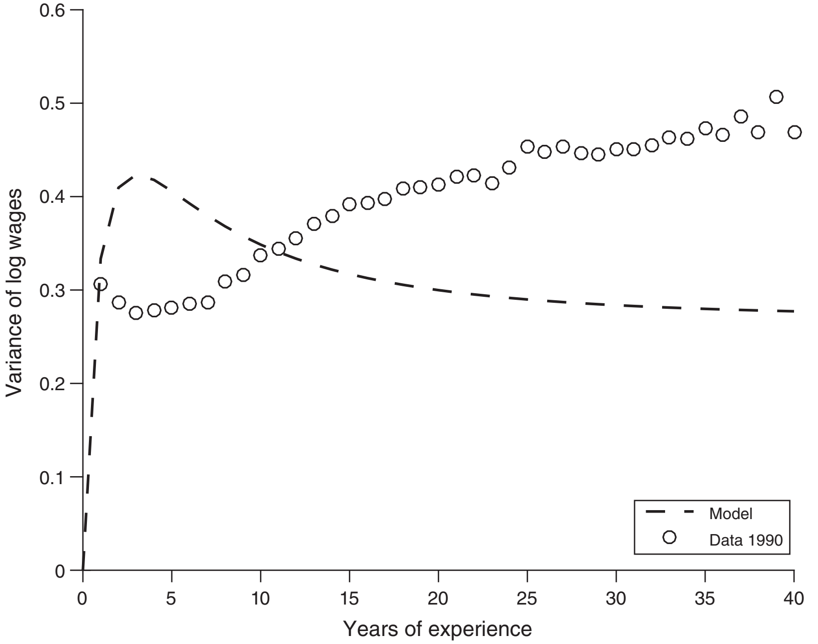

Figure 10. Variance of log Wages and Experience: Model and 1990 Data

To calculate the parameters C in (27), we minimized the mean squared difference between the predicted value of the mean log wages in the model and the actual value of the data for experience years $\{N_l, \ldots, N_u\}$. We can compute $\hat{C}$ explicitly as

$$\hat{C} = \frac{1}{N_u - N_l} \sum_{s=N_l}^{N_u} \left(\omega(s) - \int_0^{\infty} \log w(z)\, de^{-\mu(s) z^{-1/\theta}} \right),$$

where $\omega(s)$ denotes the observed mean of log wage data for individuals with s years of experience. In the simulations below, we use $N_l = 1$ years of experience and $N_u = 40$, which is to say that we are going to locate the curve that best fits the data for this range of years of experience. We obtain $\hat{C} = 10.8664$.

Figure 9 plots mean log wages as a function of experience for the 1990 census and the benchmark calibration of the model. Figure 10 plots the corresponding variances of log wages as a function of experience for the same year and for the model. On both plots, the straight dashed lines are theoretical, calibrated as described above. The dots are data. These are all displayed as experience-earnings profiles (as opposed to age-earnings profiles). The level of all the mean earnings profile is set to match up to observed levels between 1 and 40 years of experience. We have started all careers with zero productivity.

Consistent with the data, the experience-earnings profile increases over the career of individuals. In fact, the model is able to capture the rate of increase, as well as the curvature quite well. The main discrepancy is observed beyond 30 years of experience where wages decline in the data but not in the model. This may be a Ben Porath effect, or an early semiretirement effect. These are earnings per year, but the same downturn appears in the hourly earnings data provided by IPUMS. These issues require adding labor supply and retirement decisions that are not present in the model described here.[12]

The variance profile shown on Figure 10 shows that the dispersion of wages first increases and then decreases with years of experience. Everyone starts with similar prospects but over time some are lucky, learn and get promoted, while others meet less interesting people, learn less, and do not climb as high in the organizational ladder. Over time, these differences become much more pronounced and so the cross-sectional dispersion of wages increases. However, most older agents in the model do end up eventually meeting knowledgeable individuals and learning from them. Thus, even though the variance of wages remains high relative to their initial years in the labor force, it declines with the years of experience after the initial increase. The data exhibit a different pattern, with a smooth rise in wage inequality as agents go through their careers. Part of this increase might again be the result of changes in labor supply and other decisions of individuals. Our model can match relatively well the level of the variance of wages, but not its pattern over the career path of individuals. To put the magnitude of mismatch in perspective, Appendix E

[12] Our model could readily be modified to add Ben Porath-like properties as in Lucas and Moll (2014). Although perhaps interesting, such a model would imply that agents actively decide on the fraction of time they dedicate to work. However, given that we have restricted our data to a sample of full-time workers, it is unlikely that such an extension would help us match the empirical patterns we aim to explain.

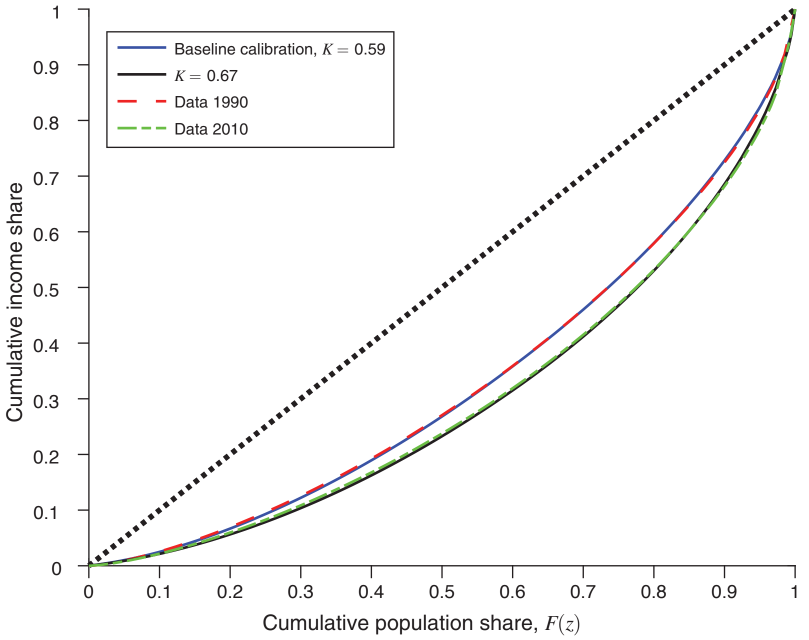

Figure 11. Lorenz Curves for 1990 and 2010 and the Model's Fit

shows that it leads to relatively small differences between the model-based distribution of earnings and the empirical histogram for different years of experience.

VII. Matching Observed Changes in Wage Inequality

During the period 1990 to 2010, and perhaps starting a couple of decades before, the US economy has experienced large changes in the distribution of earnings. These changes have been documented in a variety of studies including, for example, Katz and Murphy (1992), Goldin and Katz (2007), and Acemoglu and Autor (2010). These authors have pointed to the increase in overall inequality evident in the Lorenz curves portrayed in Figures 7 and 8, as well as to several other facts, such as the increase in the skill premium and wage polarization among skill groups. Polarization refers to the declines in the wages of agents in the middle part of the distribution relative to the ones at the top and the bottom, or, in other words, a decline in earnings inequality at the lower-half of the distribution, together with an increase at the upper-half.

We have already alluded to the fact that these changes can be replicated, at least qualitatively, as a result of an increase in K. As the distribution of problems shifts to the right relative to the distribution of knowledge, problems become more difficult but also provide a larger return if successfully solved. The result is an increase in inequality generated by an increase in the skill premium. High-skilled agents are more useful since their knowledge is relatively scarce and the problems they can solve are more valuable. Moreover, inequality among workers tends to decrease since they solve fewer problems. Their value relies more on the problems they generate (which are

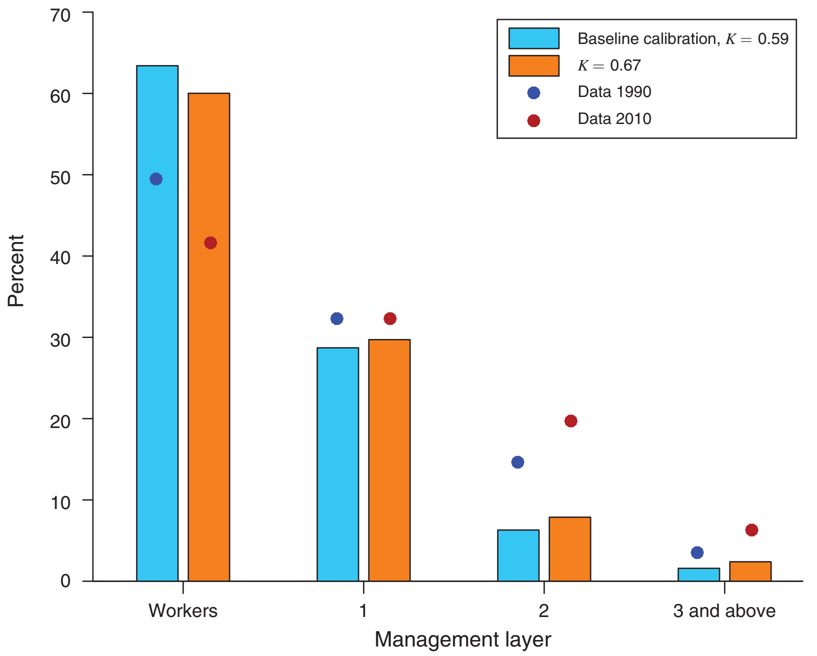

FIGURE 12. FRACTION OF WORKERS AND MANAGERS IN EACH LAYER

common to all workers) rather than the ones they solve. In contrast, inequality among managers increases since the new distribution favors the superstars.

Quantitatively we can use our baseline calibration for 1990, which uses a value of $K = 0.59$, and simply increase K to $K = 0.66$ in order to fit the 2010 Lorenz curve. We treat this parameter change as a one-time permanent shift in the problem distribution relative to the knowledge distribution, resulting in a new balanced growth path. Figure 11 shows the Lorenz curves for 1990 and 2010 together with the model-generated curves. The fit for 2010 is quite good. For example, in the data the 90/50 percentile ratio went up from 2 to 2.3 between 1990 and 2010, while the same ratio goes from 2 to 2.4 when we increase K from 0.59 to 0.66. Similarly, if we focus on the very top tail, in the data the 99.5/50 percentile ratio increased by 1.6, while it increases by 1.5 with the change in K.[13]

The larger K also generates an increase in the number of all types of managers—a feature consistent with the data during this period. Figure 12 shows the fraction of workers and managers in each layer for both calibrations, as well as the ones we compute in the data.

The rise in K also has implications on the mean experience-earnings profiles and the variance of earnings across experience levels. Figure 13 presents mean wages as a function of experience for 1990 and 2010, as well as the ones implied by our calibrations for these years. The dashed lines correspond to $K = 0.59$, should be

[13] In Appendix E, we present the fit between the distribution of earnings generated by the model and the data for 2010 overall and by experience level.

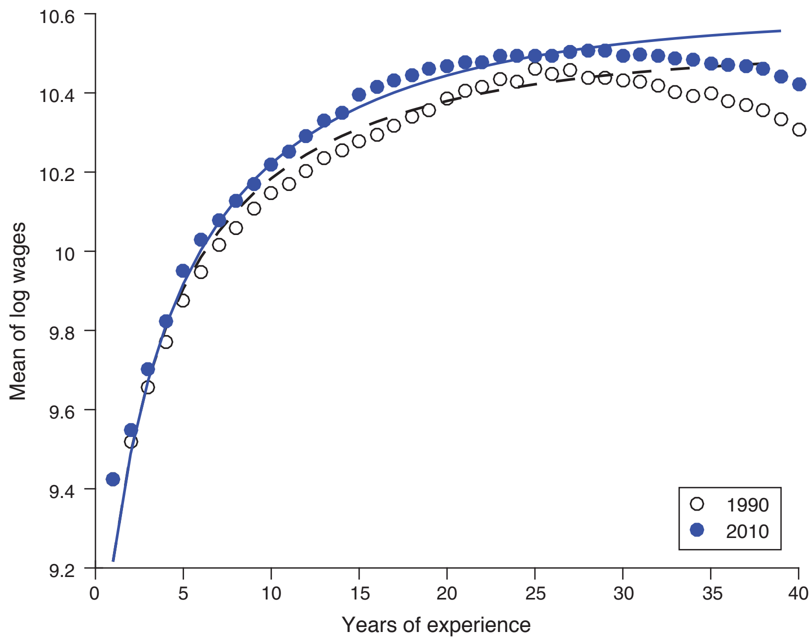

Figure 13. Experience and Change in Variance of log Earnings (1990 to 2000)

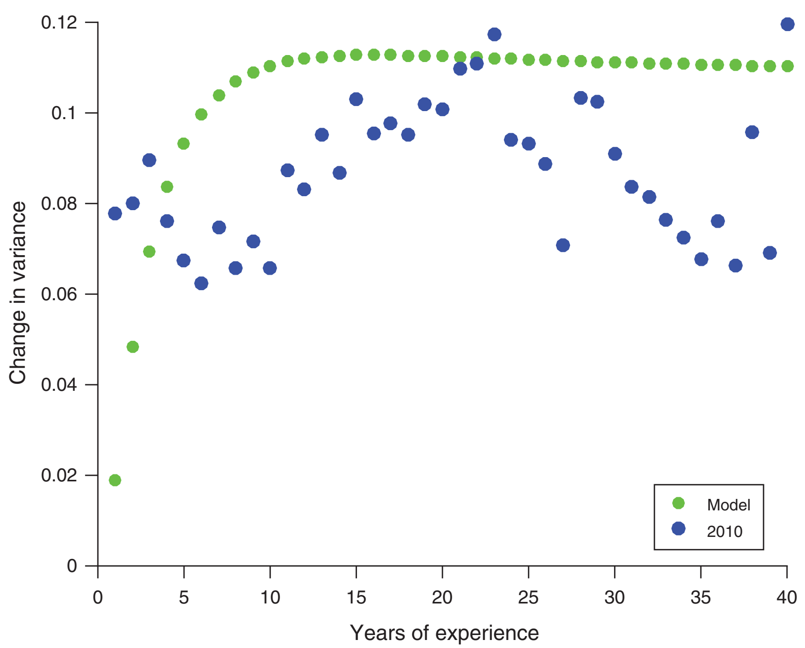

Figure 14. Fraction of Workers and Managers in Each Layer

compared to the data in 1990, and replicate Figure 9. The solid line corresponds to $K = 0.66$ and should be compared to the data in 2010. In the model, an increase in K increases the slope and level of the mean experience-earnings profiles. Figure 13 shows that these changes match the observed changes between 1990 and 2010 well. According to our model, experience became more valuable in the later period, since the acquired knowledge could be applied to a larger pool of more difficult but valuable problems.

In Figure 14, we present the change in the variance of log wages over the period 1990 to 2010, together with the implied change in the model when we increase K from 0.59 to 0.66. The model implies that the variance of log wages is always zero for agents with no experience, independently of the value of K, since all agents start with zero knowledge. The change in variance then increases rapidly with experience and stabilizes after about ten years in the labor force. These changes track the observed changes in the data reasonably well, although the data exhibit more variation. Higher K implies that knowledge is more valuable and so agents with lucky matches benefit relatively more. The magnitude of this effect generated by the model seems accurate, at least on average.

As emphasized above, we have not developed a model of the individual actions that might generate the problem distribution or its evolution over time. The results above are simply a thought experiment involving the single parameter K, one that seems to provide a good fit to changes observed in the labor market. An increase in K reflects a technological change that favors extreme outcomes, in the form of superstar ideas that are extremely well compensated. This shift can in principle increase or decrease output. The change in K that we measure during this period reduces de-trended output. Changing from $K = 0.59$ to $K = 0.66$, reduces de-trended output by approximately 2 percent.[14] We underscore that we do not explain or analyze the source of this change. We only identify it as a likely cause relative to other potential phenomena during this period. Consider, for example, an improvement in communication technology measured by an increase in the parameter κ. Such a technological change has the opposite effect on overall inequality and so does not allow us to match the 2010 Lorenz curve. Another possibility is a technological change that increases the thickness of the tail of the distribution of problems and knowledge (an increase in θ). Such a change does increase overall inequality but, as we showed in Figure 8, it generates too much inequality at the top of the distribution.

VIII. Final Remarks

Individual's earnings are determined by their skill and schooling, but also by the characteristics of the agents with whom they interact. As individuals age, acquire more experience, and learn, their career paths take them progressively to jobs with more subordinates: they climb the organizational hierarchy. Some agents climb fast because they are highly educated or because they are lucky and manage to learn

[14] In Figure A2 in Appendix B, we show that in our calibration total output is increasing in K for small values of K and decreasing for large values (inverted U-shaped pattern). The location of the maximum depends on other parameters, including θ, as we pointed out in Section V.

quickly from people they encounter during their careers. Others learn and climb the hierarchy slowly, if at all, since they meet the wrong people from whom they learn little. In this paper, we build a model of an economy in which agents go through these career paths. Formally, we have combined a model of the hierarchical organization of production similar to the one in Garicano and Rossi-Hansberg (2004, 2006) with a model of learning based on Lucas (2009). The outcome is a theory where, on the balanced growth path, individuals progress through the hierarchy according to the random set of people they encounter during their lives.

The model we propose is successful in replicating a variety of features of the US data. In particular, the model can match well the Lorenz curve of income, as well as the slope and curvature of the mean experience-earnings profile. Furthermore, the model identifies a shift in the distribution of problems relative to the distribution of knowledge as a potential source of the changes in wage inequality observed during the two decades between 1990 and 2010. Our theory suggests that, during this period, production became more difficult relative to the knowledge embedded in the population, but also resulted in higher rewards than in the past. That is, technology became more geared toward superstars, as suggested by Rosen (1981) and more recently by Gabaix et al. (2016). Can these changes be traced to the general-purpose technologies associated with computers and the internet?[15] Studying the source of this technological change is, we believe, essential to make progress in assessing the observed evolution of wage inequality and its consequences.

Appendix

A. *Proofs*

THEOREM 1: *Suppose that $F(\cdot)$ and $G(\cdot)$ are differentiable, then any solution to the planning problem is characterized by a matching function $\varphi(\cdot)$, and a pair of thresholds z_w and z_e such that $\mathcal{Z}_w = [0, z_w]$, $\mathcal{Z}_m = [z_w, z_e]$, $\mathcal{Z}_e = [z_e, \infty)$. If $\mathcal{Z}_m \neq \varnothing$, then $z_w < z_e$; otherwise, $z_w = z_e$.*

The proof is given in the next three lemmas.

LEMMA 1: *Let $(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e)$ solve the planner's problem. Then for ε sufficiently small $[0,\varepsilon) \nsubseteq \mathcal{Z}_w \cap \mathcal{Z}_e$.*

PROOF:

Toward a contradiction, suppose $[0, x_0] \subset \mathcal{Z}_w \cap \mathcal{Z}_e$. We will show that this allocation can always be improved by building a hierarchy that includes workers with z close to zero and managers with z close to x_0. Note first that the expected payoff of the worst agent is given by

$$\lim_{\varepsilon \to 0} \int_0^{\varepsilon} y g(y)\, dy = 0$$

[15] See Jovanovic and Rousseau (2005).

and for agent x_0 is

$$\lim_{\delta\to 0} = \int_0^{x_0-\delta} yg(y)\,dy > 0.$$

A manager with ability $x_0 - \delta$ can manage $\kappa/\big(1 - G(\varepsilon)\big)$ workers of ability ε and

$$\lim_{\varepsilon\to 0} \kappa/\big(1 - G(\varepsilon)\big) = \kappa > 1.$$

The total production of this team if they are all self-employed is

$$\int_0^{x_0-\delta} yg(y)\,dy + \kappa\int_0^{\varepsilon} yg(y)\,dy.$$

In contrast, if they form a team, they produce

$$\frac{\kappa}{1-G(\varepsilon)}\int_0^{x_0-\delta} yg(y)\,dy > \int_0^{x_0-\delta} yg(y)\,dy + \kappa\int_0^{\varepsilon} yg(y)\,dy$$

for $\delta \geq 0$ and $\varepsilon \geq 0$ small enough. The inequality follows since at the limit when $\delta = \varepsilon = 0$, $\kappa > 1$ implies that

$$\kappa\int_0^{x_0} yg(y)\,dy > \int_0^{x_0} yg(y)\,dy.$$

Hence, the planner would always prefer to form these hierarchies for a nonzero measure of agents than to have all of them as self-employed, which contradicts $\big(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e\big)$ being a solution to the planner's problem. ∎

LEMMA 2: *Let $\big(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e\big)$ solve the planner's problem. Then if $z_m \in \mathcal{Z}_m$ and $z_e \in \mathcal{Z}_e$, we have $z_m < z_e$.*

PROOF:

Toward a contradiction, suppose $\big(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e\big)$ solves the planner's problem above but that $z_e < z_m$ for some $z_e \in \mathcal{Z}_e$ and a $z_m \in \mathcal{Z}_m$. In this case, note that since $z_e \in \mathcal{Z}_e$, the planner fails to solve all problems that require knowledge $z > z_e$. These problems have total value

$$\big(1 - G(z_e)\big)\int_{z_e}^{\overline{z}} y\frac{g(y)}{1-G(z_e)}\,dy = \int_{z_e}^{\overline{z}} yg(y)\,dy$$

if an agent with ability $\overline{z}$ tries to solve them. In contrast, agent z_m passes along the problems he cannot solve to agent $\varphi(z_m)$. These problems have total value

$$\big(1 - G(z_m)\big)\int_{z_m}^{\varphi(z_m)} y\frac{g(y)}{1-G(z_m)}\,dy = \int_{z_m}^{\varphi(z_m)} yg(y)\,dy$$

if agent $\varphi(z_m)$ tries to solve them. But, letting $\overline{z} = \varphi(z_m)$, given $z_e < z_m$, we have

$$\int_{z_e}^{\varphi(z_m)} yg(y)\,dy > \int_{z_m}^{\varphi(z_m)} yg(y)\,dy.$$

The value of the problems discarded by agent z_e exceed the value that would have been discarded if agent z_m had been assigned to agent $\varphi(z_m)$ instead. A contradiction. ∎

LEMMA 3: *Let* $(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e)$ *solve the planner's problem. Then,* $z \in \mathcal{Z}_w$ *implies that* $z \leq z'$ *for all* $z' \in \mathcal{Z}_m \cup \mathcal{Z}_e$.

PROOF:

Toward a contradiction, suppose $(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e)$ solves the planner's problem above but does not satisfy this property. Then there exists a $z_w \in \mathcal{Z}_w$ and a $z_m \in \mathcal{Z}_m$ such that $z_w > z_m$. Note first that the value of a problem passed up to a manager $\overline{z}$ by any agent is given by $\int_0^{\overline{z}} yg(y)\,dy$, independent of the ability of the worker who generated it. (It only depends on the best agent that tries to solve it, $\overline{z}$.) In contrast, the value of a manager for the planner's objective is given by

$$\kappa \int_{\varphi^{-1}(z_m)}^{z_m} y \frac{g(y)}{1 - G(\varphi^{-1}(z_m))}\,dy - \int_0^{z_m} yg(y)\,dy.$$

Suppose the planner switches the role of z_w and z_m and makes z_m the worker and z_w the manager, holding fixed their supervisors and subordinates. Then the gain for the planner is given by

$$\begin{aligned}
&\left(\kappa \int_{\varphi^{-1}(z_m)}^{z_w} y \frac{g(y)}{1 - G(\varphi^{-1}(z_m))}\,dy - \int_0^{z_w} yg(y)\,dy\right) \\
&\quad - \left(\kappa \int_{\varphi^{-1}(z_m)}^{z_m} y \frac{g(y)}{1 - G(\varphi^{-1}(z_m))}\,dy - \int_0^{z_m} yg(y)\,dy\right) \\
&\quad = \frac{\kappa}{1 - G(\varphi^{-1}(z_m))} \int_{z_m}^{z_w} yg(y)\,dy - \int_{z_m}^{z_w} yg(y)\,dy > 0
\end{aligned}$$

since $\kappa > 1 - G(\varphi^{-1}(z_m))$. Hence, a planner would gain with the switch: a contradiction. ∎

THEOREM 2: *Suppose that* $F(\cdot)$ *and* $G(\cdot)$ *are differentiable, then there is a pair* (z_w, z_e) *that satisfies* (12)–(14).

PROOF:

Let z_w^* be the highest cutoff z_w such that all problems are solved, the unique solution to

$$1 - F(z_w) = \frac{1}{\kappa} \int_0^{z_w} [1 - G(z)] f(z)\,dz + \int_{z_w}^{\infty} \frac{1 - G(z)}{1 - G(\varphi^{-1}(z))} f(z)\,dz.$$

Total production in this case will be

$$\int_0^{z_w^*}\int_0^{\infty} yg(y)\,dy\,f(z)\,dz \;=\; F(z_w^*)\int_0^{\infty} yg(y)\,dy.$$

Also note that as $z_w \to \infty$ total production converges to the autarchy level

$$\int_0^{\infty}\int_0^{z} yg(y)\,dy\,f(z)\,dz.$$

Then continuity ensures that for some (z_w, z_e) with $z_w^* \leq z_w \leq z_e$,

$$\int_0^{z_w}\int_0^{\infty} yg(y)\,dy\,f(z)\,dz - \kappa\int_{z_e}^{\infty}\int_z^{\infty} yg(y)\,dy\frac{f(z)}{1 - G(\varphi^{-1}(z, z_w))}\,dz$$
$$\geq \max\left[F(z_w^*)\int_0^{\infty} yg(y)\,dy, \int_0^{\infty}\int_0^{z} yg(y)\,dy\,f(z)\,dz\right]. \blacksquare$$

LEMMA 4: *Let* $(\varphi(\cdot), \mathcal{Z}_w, \mathcal{Z}_m, \mathcal{Z}_e)$ *solve the planner's problem. Then if* $z_w = z_e$, z_w *must satisfy*

$$\kappa[1 - F(z_w)] \;=\; \int_0^{z_w} f(z)\int_z^{\infty} g(y)\,dydz.$$

PROOF:

Toward a contradiction, suppose that $(\varphi(\cdot), \mathcal{Z}_w\,\mathcal{Z}_m, \mathcal{Z}_e)$ solves the planner's problem and that

$$\kappa[1 - F(z_w)] \;<\; \int_0^{z_w} f(z)\int_z^{\infty} g(y)\,dydz.$$

Then there are too few entrepreneurs to review all the problems unsolved by workers. Replacing z_w with $z_w - h < z_w$ in this case will increase the left side by $\kappa[1 - F(z_w - h)] - \kappa[1 - F(z_w)] \simeq \kappa h f(z_w)$, and this will allow the new managers to solve $\kappa h f(z_w)$ problems worth on average $\int_0^{z_w} yg(y)\,dy$ each (since they do not know how to solve the more difficult problems). This replacement thus adds $\kappa h f(z_w)\int_0^{z_w} yg(y)\,dy$ to production. It also reduces the right side by

$$\int_{z_w - h}^{z_w} f(z)\int_z^{\infty} g(y)\,dydz \;\simeq\; hf(z_w)\int_{z_w}^{\infty} g(y)\,dy \;=\; hf(z_w)\left[1 - G(z_w)\right],$$

workers. Since all the problems harder than z_w were thrown away, the planner valued these problems only at

$$hf(z_w)\left[1 - G(z_w)\right]\int_0^{z_w} yg(y)\,dy.$$

Since $\kappa > \left[1 - G(z_w)\right]$, the planner will not choose to do this: a contradiction.

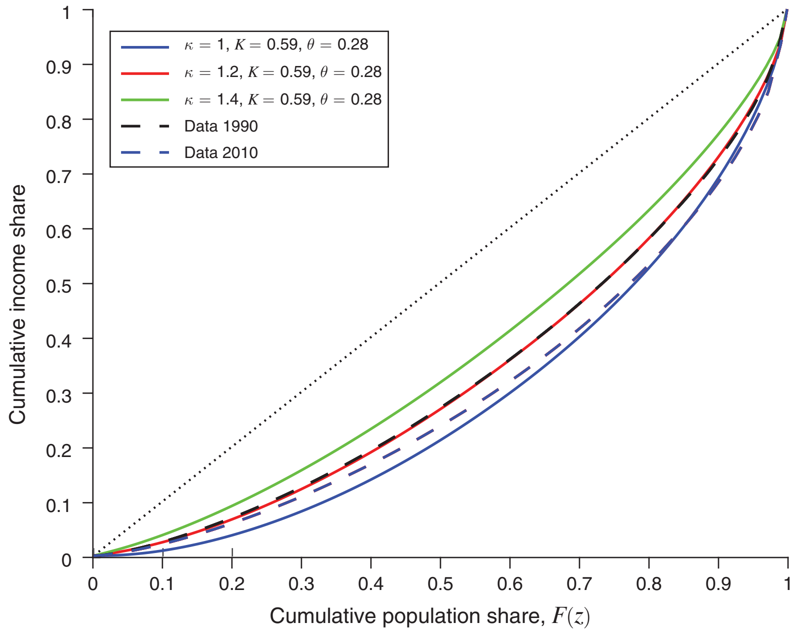

Figure A1. Data and Theory-Based Lorenz Curves for Different Values of κ

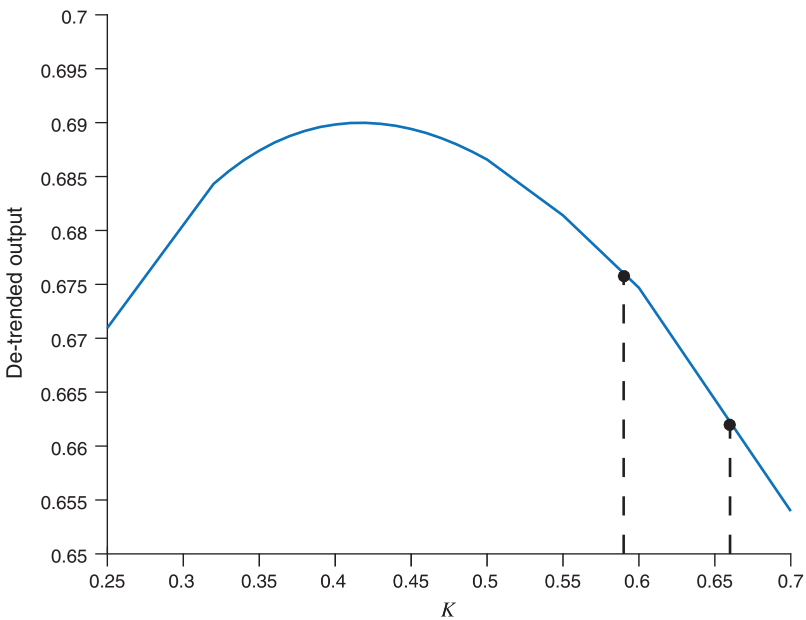

Figure A2. Fraction of Workers and Managers in Each Layer

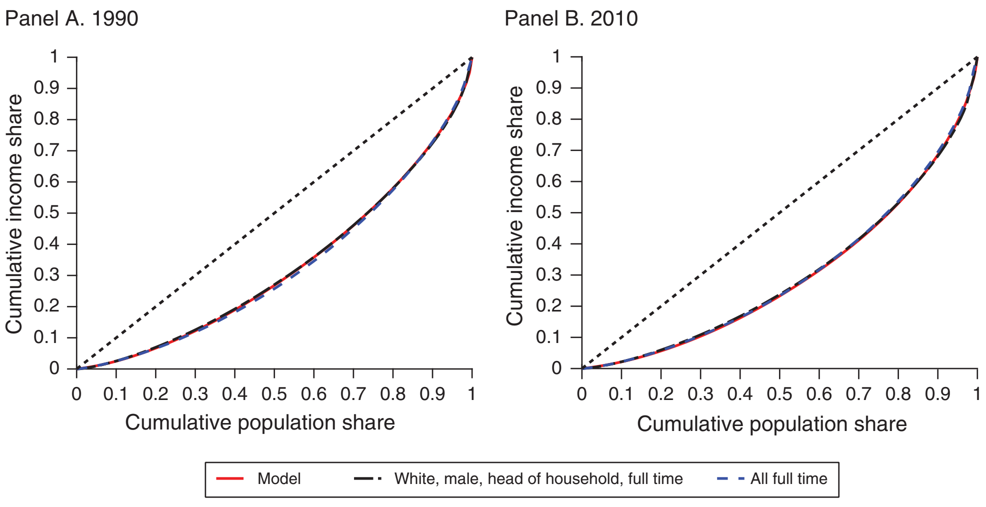

FIGURE A3. SAMPLE COMPARISON AND MODEL FIT, LORENZ CURVES, 1990 AND 2010

If, however, the left side of (10) exceeds the right, there must be entrepreneurs who advise no one. They might choose to be workers, but this would violate Lemma 3: also a contradiction. ∎

B. *Additional Plots*

Figure A1 shows the Lorenz curves for the distribution of wages in the 1990 and 2010 US data, as well as the model-based ones for different values of κ. As highlighted in Section VI, an improvement in communication technology driven by κ reduces wage inequality since it increases the span of control of knowledgeable agents, making problem generation more valuable and the talent of knowledgeable individuals less scarce.

Figure A2 shows that for our calibration total output is increasing in K for small values of K and decreasing for higher values. For a given skill distribution, if the problems are easy to solve, people would produce more output if they could draw from a distribution of more complex and profitable problems. However, if the problems are too difficult, increasing their difficulty would reduce output as a smaller portion of the population would be able to solve them.

In the main text, we use a sample of white males, heads of household, that are fully employed to control for underemployment of secondary earners and minorities. In Figures A3 and A4, we compare the results presented in the main text with results using an alternative sample of all full-time workers. As we argue in the main text, the Lorenz curves and experience profiles are very similar and so the fit of our calibrations remains quite good.

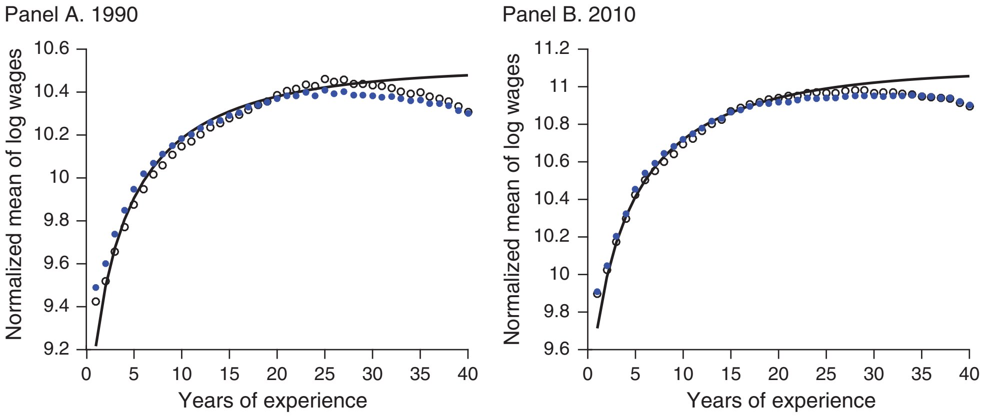

Figure A4. Sample Comparison and Model Fit, Lorenz Curves, 1990 and 2010

Table A1—Unmatched Codes as a Fraction of the Weighted Observation of the Sample

Unmatched 1990 (percentage)	Unmatched 2010 (percentage)
21.9	16.3

C. *Occupational Codes*

To obtain the fraction of agents employed in each layer, we convert the IPUMS-USA classification into the French PCS classification used in Caliendo, Monte, and Rossi-Hansberg (2015).

The occupation data in the IPUMS-USA (OCC2010) is based on the occupation coding used by the Bureau of Labor Statistics (BLS) for the American Community Survey (ACS) in 2010. We use three different crosswalks to translate the coding to the French PCS data to define the layers of our economy. First we match the OCC2010 codes to the International Standard Classification of Occupations ISCO-08, then to the ISCO-88, and finally to the PCS codes.[16]

After doing this process, there are some unmatched codes. Out of the 539 different ACS codes in our sample, 43 were unmatched (7.98 percent). In terms of the weighted observations of our sample, Table A1 shows the percentage of unmatched weighted observations for 1990 and 2010.

[16] The crosswalk between ACS codes and SOC codes can be downloaded at http://www.census.gov/people/io/methodology/, the one between SOC codes and ISCO-08 at http://www.bls.gov/soc/soccrosswalks.htm, between ISCO-08 and ISCO-88 at http://www.iolo.org/public/english/bureau/stat/isco/isco08/index.htm, and between ISCO-88 to PCS codes from the appendix of Caliendo, Monte, and Rossi-Hansberg (2015).

In Section VII, we computed the fraction of individuals in each layer using only the codes that we matched directly. The results are similar if we impute to the missing codes the value of the top (or bottom) layer implied by the immediate neighboring codes.

D. *Computations*

Here we provide a constructive method for finding a solution to the planner's problem. To do so, it is useful to describe the economy by the number of layers it has. Let L denote the number of layers, where $L = 0$ means all the individuals in the economy are workers, $L = 1$ that there are workers and one-level managers, and so on. We know from Lemma 1 that if $\kappa > 1$, it is never optimal to only have workers in the economy ($L = 0$) as it is always preferable to have managers for the lowest productivity workers. In general, we want to characterize the solution in terms of the number of layers the economy will optimally have.

The idea is to choose the worker's cutoff z_w that maximizes total output. For a given z_w, we compute total output by first calculating the matching function $\varphi(\,\cdot\,, z_w)$ and the entrepreneur's cutoff $z_e(z_w)$ such that no managerial resources are wasted. We compute the matching function by iterating on the number of layers in the economy. In each iteration, if the layer L manager of the individual that knows nothing $z = 0$ has knowledge below the entrepreneur's cutoff, we add a layer to the matching function and recompute $z_e(z_w)$. Below we describe in detail the algorithm we use to find the solution to this maximization problem.

First we introduce some notation. In an economy with L layers, each worker can be linked to L or $L - 1$ managers. Specifically, suppose that workers below some productivity level x have L managers, while workers with productivity $z \in [x, z_w]$ have $L - 1$ managers. Let $\varphi_l(z)$ denote the lth manager matched with worker $z \in [0, z_w]$. We define this mapping by applying the matching function l times, i.e., $\varphi_l(z) := \underbrace{\varphi(\varphi(\ldots\varphi}_{l\text{ times}}(z)))$. Similarly, we define the inverse matching function that maps managers to their subordinates. The function $\varphi_l^{-1}(z)$ denotes the lth subordinate of manager z, and it is defined by applying the inverse of the matching function l times, $\varphi_l^{-1}(z) := \underbrace{\varphi^{-1}(\varphi^{-1}(\ldots\varphi^{-1}}_{l\text{ times}}(z)))$.

To illustrate these definitions, consider the case where

- $(0, z_w)$ are workers of all kinds,
- $(0, x)$ are workers with access to L managers,
- (x, z_w) are workers with access to $L - 1$ managers,
- (z_w, z_e) are managers with access to other managers, and
- (z_e, ∞) are the top managers (entrepreneurs).

Figure A5 depicts this case.

Given this notation, we can describe an algorithm in which we choose the cutoff that maximizes output. As mentioned above, note that for a given set of cutoffs (z_w, z_e), we can find a function $\varphi(\,\cdot\,; z_w)$ that satisfies equations (12) and (13).

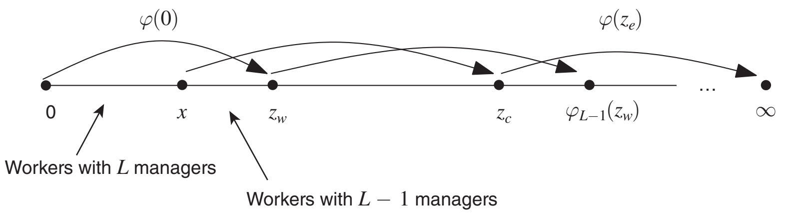

Figure A5. Partition of Workers

Moreover, as the planner's objective function (total production) is increasing in both z_w and z_e, the idea is to find the largest feasible cutoffs such that these equations are satisfied.

The algorithm will consist of two steps. In step 1, we compute the worker's cutoff such that there is exactly one layer ($L = 1$). In this case, the cutoffs are $z_w^1 = z_e^1$ defined by solution of equation (2),

$$1 - F(z_w^1) = \frac{1}{\kappa}\int_0^{z_w^1}\left[1 - G(z)\right] f(z)\,dz.$$

In step 2, we find $z_w \in [0, z_w^1]$ such that it maximizes

$$Y(z_w) := \int_0^{z_w}\left(\int_0^{\infty} yg(y)\,dy\right) f(z)\,dz - \kappa\int_{z_e(z_w)}^{\infty}\left(\int_z^{\infty} yg(y)\,dy\right)\frac{f(z)}{1 - G(\varphi^{-1}(z, z_w))}\,dz,$$

where $z_e(z_w)$ is the entrepreneur's cutoff and is implicitly defined by $\varphi(z_e, z_w) = 1$.

The Algorithm.—**Step 1**: Compute z_w^1 such that (2) is satisfied with $z_e^1 = z_w^1$.

Step 2: Choose $z_w \in [0, z_w^1]$ that maximizes the total output (using MATLAB optimization functions).

To compute $Y(z_w)$,

(i) For any $L > 1$, find z_e^L such that $\varphi(z_e^L, z_w) = 1$, i.e.,

$$1 - F(z_w) = \frac{1}{\kappa}\int_0^{z_w}(1 - G(y)) f(y)\,dy + \int_{z_w}^{z_e^L}\frac{(1 - G(y))}{(1 - G(\varphi^{-1}(y)))} f(y)\,dy.$$

(ii) Compute the matching function for $y \in \left[\varphi_{L-2}(z_w), \min\left(\varphi_{L-1}(z_w), z_e^L\right)\right]$,

$$\varphi(y) = F^{-1}\left(F(\varphi_{L-2}(z_w)) + \int_{\varphi_{L-2}(z_w)}^{y}\frac{(1 - G(y))}{(1 - G(\varphi^{-1}(y)))} f(y)\,dy\right).$$

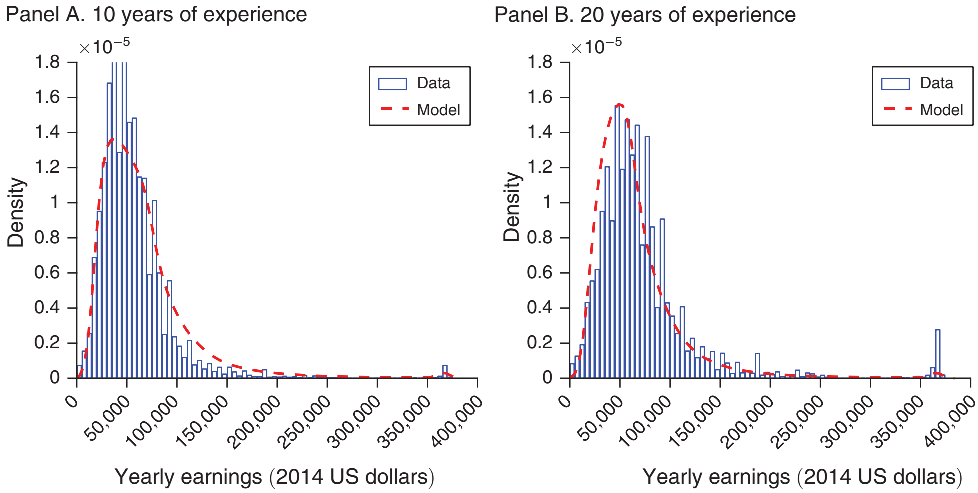

Figure A6. Wage Distribution: Data and Model Fit, 1990, 10 and 20 Years of Experience

(iii) If $\varphi_{L-1}(z_w) < z_e^L$, add another layer $L = L+1$ and go back to 1.

(iv) If $\varphi_{L-1}(z_w) \geq z_e^L$, compute total output,

$$Y(z_w) = \int_0^{z_w}\left(\int_0^{\infty} yg(y)\,dy\right)f(z)\,dz - \kappa\int_{z_e^L}^{\infty}\left(\int_z^{\infty} yg(y)\,dy\right)\frac{f(z)}{1-G(\varphi^{-1}(z,z_w))}\,dz.$$

E. *The Wage Distribution for Different Years of Experience*

Here we compare the empirical and theoretical wage distribution using the calibration of the model for 1990 and 2010 for different years of experience, $s = \{10, 20, 30, 40\}$. To make the empirical and theoretical distributions comparable, we calculate a constant for each year such that the model implied distribution has the same mean as the data.

Figures A6 to A10 plot the histogram and the fitted distribution for individuals with $s = \{10, 20, 30, 40\}$ years of experience in 1990 (Figures A6 and A7) and 2010 (Figures A9 and A10). In Figure A8, we also include the distribution for all experiences pulled (Figure 3 in the main text presents the corresponding figure for 1990).

Even though the distributions implied by the model are similar to the empirical histograms, it is noticeable that the theoretical distributions for individuals with 10 years of experience are more disperse than in the data. In contrast, for experienced individuals, the model produces a slightly thicker tail than what we observe in the data. These plots illustrate the magnitude of the differences in the variance-experience profiles between the data and our baseline calibration presented in Figure 10.

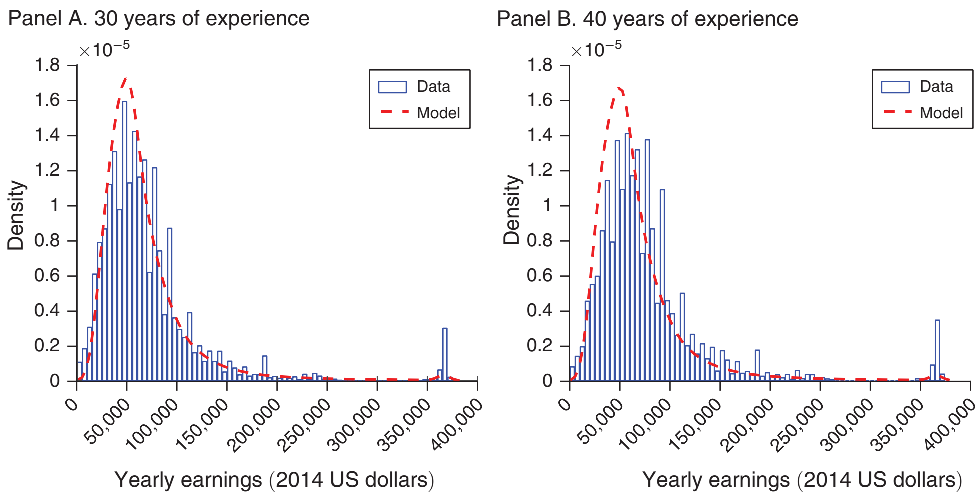

Figure A7. Wage Distribution: Data and Model Fit, 1990, 30 and 40 Years of Experience

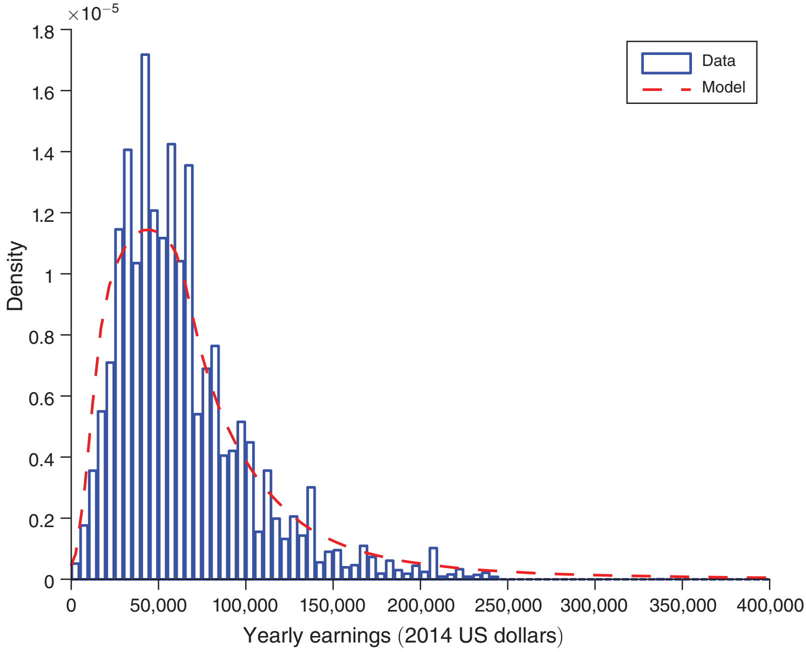

Figure A8. Wage Distribution: Data and Model Fit, 2010, Overall

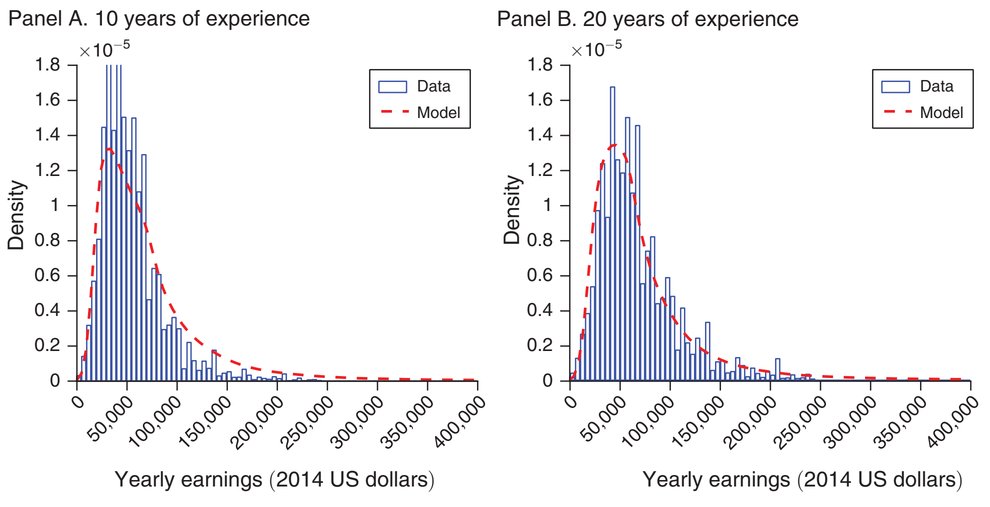

Figure A9. Wage Distribution: Data and Model Fit, 2010, 10 and 20 Years of Experience

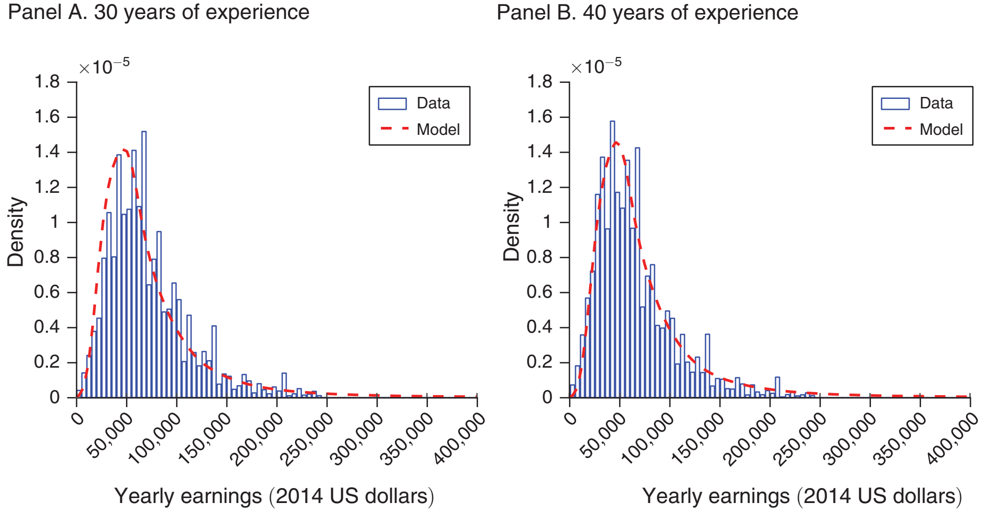

Figure A10. Wage Distribution: Data and Model Fit, 2010, 30 and 40 Years of Experience

REFERENCES

Acemoglu, Daron, and David Autor. 2010. "Skills, Tasks, and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, Vol. 4B, edited by Orley Ashenfelter and David Card, 1043–1171. Amsterdam: North-Holland.

Caicedo, Santiago, Robert E. Lucas Jr., and Esteban Rossi-Hansberg. 2019. "Learning, Career Paths, and the Distribution of Wages: Dataset." *American Economic Journal: Macroeconomics.* https://doi.org/10.1257/mac.20170390.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The Anatomy of French Production Hierarchies." *Journal of Political Economy* 123 (4): 809–52.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica* 84 (6): 2071–2111.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5): 874–904.

Garicano, Luis, and Esteban Rossi-Hansberg. 2004. "Inequality and the Organization of Knowledge." *American Economic Review* 94 (2): 197–202

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121 (4): 1383–1435.

Goldin, Claudia, and Lawrence F. Katz. 2007. "The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005." National Bureau of Economic Research (NBER) Working Paper 12984.

Jovanovic, Boyan, and Peter L. Rousseau. 2005. "General Purpose Technologies." In *Handbook of Economic Growth*, Vol. 1, Part B, edited by P. Aghion and S. N. Durlauf, 1181–24. Elsevier: North-Holland.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107 (1): 35–78.

Lucas, Robert E., Jr. 2009. "Ideas and Growth." *Economica* 76 (301): 1–19.

Lucas, Robert E., Jr., and Benjamin Moll. 2014. "Knowledge Growth and the Allocation of Time." *Journal of Political Economy* 122 (1): 1–51.

Rosen, Sherwin. 1981. "The Economics of Superstars." *American Economic Review* 71 (5): 845–58.

Ruggles, Steven, Catherine A. Fitch, Patricia Kelly Hall, and Matthew Sobek. 2000. "IPUMS-USA: Integrated Public Use Microdata Series for the United States." In *Handbook of International Historical Microdata for Population Research*, edited by Patricia Kelly Hall, Robert McCaa, and Gunnar Thorvaldsen, 259–84. Minneapolis: Minnesota Population Center.

Slavík, Antonín. 2007. *Product Integration, Its History and Applications*. Prague: Matfyzpress.