Organization of Knowledge and Taxation

Marek Kapička and Ctirad Slavík[*]

January 13, 2021

Abstract

This paper studies how labor income taxation interacts with the organization of knowledge and production, and ultimately the distribution of wages in the economy. A more progressive tax system reduces the time that managers allocate to work. This makes the organization of production less efficient and reduces wages at both tails of the distribution, which increases lower tail wage inequality and decreases upper tail wage inequality. The optimal tax system is only modestly more progressive than the current one in the United States. However, the optimal tax progressivity is substantially smaller than if the wage structure was exogeneous.

J.E.L Codes: E6, H2, D8, L23

Keywords: inequality, wages, knowledge based hierarchies, income taxation

[*]Marek Kapička, CERGE-EI. Email: marek.kapicka@cerge-ei.cz. Ctirad Slavík, CERGE-EI. Email: ctirad.slavik@cerge-ei.cz. CERGE-EI is joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Czech Academy of Sciences. Politických vězňů 7, Praha 1, 111 21, Czech Republic. We would like to thank Pawel Doligalski, Asier Mariscal, John Moore, B. Ravikumar and Florian Scheuer, as well as audiences at various seminars, workshops and conferences for useful comments. We also thank Faisal Sohail for help with the CPS data. The authors acknowledgde support from the Czech Science Foundation, grant 20-28979S.

1 Introduction

United States and other developed economies have recently experienced substantial changes in wage inequality. In particular, after 1986, the upper tail wage inequality (90/50 percentile ratio) has significantly increased, while the lower tail wage inequality (50/10 percentile wage ratio) has decreased (many empirical studies confirm this fact, see for example Garicano and Rossi-Hansberg (2011) with data until 2015). Standard models that study optimal taxation either assume that the wage distribution is exogenous, or that it can be partially modified by human capital investment. This is true for papers that use mechanism design techniques such as Mirrlees (1971) and many followers, as well as for papers that use parametric tax functions, such as Heathcote et al. (2017) and others. Neither of the approaches can explain the observed changes in the upper and lower wage inequality without artificially engineering "correct" changes in the underlying exogenous distributions of wages or abilities. In addition, the interaction between changes in wage inequality and changes in taxes is nonexistent or limited.

In this paper we analyze the interaction between taxation and wage inequality using a theory that studies how society organizes and uses knowledge in production through knowledge based hierarchies (Garicano (2000), Garicano and Rossi-Hansberg (2006)). The theory of knowledge based hierarchies can explain simultaneous changes in the upper tail and the lower tail wage inequality as follows. It takes time to coordinate and communicate knowledge among managers and production workers. A decrease in the required communication time makes good managers relatively more useful in solving tasks, which increases wages of good managers relative to less able managers (an increase in upper tail inequality). At the same time, a decrease in the communication time allows mangers to supervise more production workers. Even the production workers at the bottom of the distribution benefit from being matched with better managers, which reduces lower tail wage inequality.

We augment the theory by generalizing the communication cost function and by endogenizing hours worked by the agents. If managers change their hours worked, time available for supervision of production workers changes. This has an effect similar to changes in the required communication time. It thus produces changes in the upper and lower wage inequality. Since distortive taxation changes equilibrium hours worked, it changes the wage distribution too, in addition to the standard effects on earnings. Furthermore, the theory is able to generate a number of additional predictions about the interaction of taxation and organization of firms, span of control, firm growth and other variables.

We assume that the income-tax function that the government uses has the constant-rate-of progressivity form as in Benabou (2002). This allows for closed form solutions (see also Heathcote et al. (2017) and Kapička (2020)). We calibrate a version of the model with one layer of management to match the moments of the U.S. wage distribution and compare the status quo with two scenarios. First, we calculate the optimal tax progressivity in a world in which the wage distribution is exogeneous. In this case, the optimal tax progressivity is substantially higher than in the status quo (0.412 vs. 0181). However, when wages are endogenously determined by the knowledge based hierarchies, the optimal tax progressivity is $\tau = 0.197$, only modestly higher than in the status quo. Lower tax progressivity increases hours worked of everyone (analogously to a decrease in the communication time), which increases upper tail wage inequality, but decreases lower tail wage inequality. Ignoring endogenous wage changes implies that the planner ignores these effects leading to potentially large welfare losses.

The rest of this paper is organized as follows. In the next section, we discuss the related literature. Section 3 lays out the general model with multiple layers of management. For simplicity and without much loss of generality, we omit the possibility of self-employment. Section 4 characterizes the equilibrium. In section 5 we consider a simple example with a closed form solution to provide economic intuition for our results. Section 6 characterizes analytically the optimal tax progressivity. Section 7 discusses the calibration of a one-management-layer version of the model and section 8 presents the main quantitative results. Section 9 concludes.

2 Related Literature

There are several strands of literature that are related. A large volume of research provides a connection between the wage distribution and taxes through general equilibrium effects. Meh (2005), Boháček and Zubrický (2012) and Bruggemann (ming) follow Quadrini (2000) and Cagetti and De Nardi (2006), and consider tax reforms in Bewley-Aiyagari economies with entrepreneurial activity. Taxes affect workers' wages first through changes in capital accumulation and, second, through endogenous occupational choice. They do not consider optimal taxation, however. Optimal taxation in models with entrepreneurship is considered by Albanesi (2011) and Shourideh (2012) who, however, do not model workers and, hence, there is no occupational choice. Ales and Sleet (2016) study optimal taxation of top CEOs. They assume that higher effort of top earners (CEOs) positively affects the productivity and profits of the firm. However, workers are not explicitly modelled either, and, therefore, there is no direct channel trough which taxation of top earners would influence the wage schedule of regular workers.

Several recent papers study optimal taxation in models with heterogeneous occupations and endogeneous wages. Rothschild and Scheuer (2013) and Scheuer (2014) study optimal taxation in an occupational choice Roy model. Slavík and Yazici (2014) study optimal taxation in a growth model with skilled and unskilled labor and capital-skill complementarity, and Ales et al. (2015) study optimal taxation in a task-to-talent assignment model of the labor market. As in the aforementioned papers, the interaction between different occupations (or workers and entrepreneurs) in these papers is only through general equilibrium effects.

Models in which managers, or enterpreneurs, interact with workers and thus affect their wages directly, are less frequent. Saez et al. (2014) consider a model in which workers' wages are the result of bargaining between workers and CEOs. If top marginal rates are lower, then the CEOs will bargain more aggressively for higher compensation which increases wage dispersion. As a result, endogeneous wages (which are a result of compensation bargaining) lead to higher optimal wage progressivity, in contrast to the present paper.

Ales et al. (2017) and Scheuer and Werning (2017) study models similar to our model. Ales et al. (2017) build upon Rosen (1982)'s assignment model of talent allocation within a firm and focus on the optimal taxation of top labor incomes. In contrast to our model, the potential impact of taxes on workers' wages is limited. Workers are ex-ante identical, receive the same consumption and their assignment to different managers is indeterminate. We relax the assumptions leading to the assignment indeterminacy, and study the relationship between taxes and wage inequality at both tails of the wage distribution. We are thus able to model both lower-tail and upper-tail income inequality, a key aspect of our paper. Scheuer and Werning (2017) also focus on optimal taxation of top-income individuals. In an extension to their basic environment, they consider an assignment model with workers, self-employed and one layer of management. They show that the usual Mirrleesian tax formulas apply, but the convexity of the wage function implies higher elasticities at the top leading to lower optimal top marginal rates. We find that in such a model, it is impossible to match the bottom and top income inequality, which is a necessary condition for our quantitative analysis. Therefore, we focus on a richer model with a general communication cost function.

Finally, Lopez and Torres (2020) considers a framework very similar to ours, but with inelastic labor supply. They do not study labor income taxation, but instead focus on the role of firm-size-dependent policies.

3 Setup

There is a measure one of agents. Agents like to consume, and dislike to work. Their preferences are represented by an additively separable utility function

$$U(c) - V(\ell),$$

where $c \geq 0$ is consumption, $\ell \geq 0$ is time spent at work, U is increasing, concave and differentiable, and V is increasing, convex and differentiable. We assume that the utility function takes the form

$$U(c) = \log c, \quad V(\ell) = \kappa \frac{1}{1+\eta} \ell^{1+\eta}$$

for $\eta > 0$ being the inverse of Frisch elasticity of labor, and $\kappa > 0$.

The technology is similar to Garicano (2000), Garicano and Rossi-Hansberg (2006) or Geerolf (2017). Agents differ in their knowledge, $z \in [\underline{z}, \overline{z}]$, exogenously given. The distribution of knowledge is $G(z)$, with $G(\underline{z}) = 0$, $G(\overline{z}) = 1$ and has density function $g(z)$. There is a continuum of tasks per period distributed according to $F(z)$ defined on $[0, \overline{z}]$, with a density function $f(z)$. An agent with knowledge z can solve all tasks in $[0, z]$ and produce $F(z)$ per unit of time. An agent working ℓ units of time thus can produce $\ell F(z)$. If $\underline{z} > 0$ then there is a mass of problems $F(\underline{z}) > 0$ that every agent can solve, namely $[0, \underline{z}]$.

Rather that producing on their own, agents form teams, where only some agents (workers) solve tasks while other agents (managers) specialize in explaining harder tasks to the others. There can be more than one layer of management. We make two assumptions about communication between workers and managers. First, agents do not know whether they can solve a problem when it arrives. The agents thus first try to solve a particular problem by themselves and, if they cannot, ask the manager for help. A worker with knowledge x_0 asks for help with tasks that he cannot solve, that is with tasks $z \geq x_0$. The manager in the first layer helps them understand how to solve the problem, if he can. If the manager has knowledge x_1 then he helps the worker with tasks $z \in [x_0, x_1]$. Tasks harder than x_1 are passed on to the managers in the second layer, where the process is repeated. If there are I layers of management and the top layer manager has knowledge x_I, then the management will ultimately be able to explain all tasks weakly easier than x_I. Tasks harder than x_I will be unsolved by the organization. In any case, the problem itself is solved by the worker; managers do not solve problems themselves.

Second, we assume that managers spend time communicating over the delegated problems (all the communication costs are incurred by the managers). The way to think about this assumption is that a worker approaches a manager with a problem that he/she cannot solve. The worker explains the problem to the manager at which point the manager incurs the time costs. After the problem has been communicated, the manager helps the worker solve the problem, if he can. A problem needs to be explained to the worker only once; once it has been explained, the worker can solve it whenever it arrives.

Organizations have a team consisting of production workers and I layers of management. The set $[\underline{z}, \overline{z}]$ is partitioned into $I+1$ connected subsets separated by I thresholds $z_1, z_2, \ldots, z_I$. For easier notation, we set $z_0 = \underline{z}$ and $z_{I+1} = \overline{z}$ to be the lower bound and upper bound on knowledge. Agents with knowledge in $[z_0, z_1]$ are production workers. Agents with knowledge in $(z_1, z_2]$ are first level managers, agents with knowledge in $(z_i, z_{i+1}]$ are managers of level i. Managers of level I, who are at the top of the hierarchy, have knowledge $(z_I, z_{I+1}]$.

We denote the knowledge of the production worker by x_0 and the knowledge of the manager in layer i by x_i. Managers in layer i are able to advise with tasks easier than z_{i+1}. After receiving advice, workers produce output. The production of the team is

$\ell_0 F(x_I)n_0$, where n_0 is the number of production workers, ℓ_0 are hours worked by the production workers, and x_I is the knowledge of the layer I manager.[1]

The managers face a time constraint that limits how many production workers they can supervise. Consider an organization with n_0 production workers with skill x_0 and n_i managers in layer i that have skill x_i. The total time supplied by managers in layer i is $n_i\ell_i$. The total time cost depends on two factors. First, more workers will pass on proportionally more tasks to be solved and explained in layer i, and so the time cost is linear in n_0. Second, the time cost per task, θ, is allowed to depend on the skill of the subordinate manager or worker, x_{i-1}. Overall, the time constraint for the managers in layer i is

$$n_0\theta\left(x_{i-1}\right) = n_i\ell_i, \quad i = 1, \ldots, I-1, \tag{1}$$

where we assume that the time cost function θ has the following properties:

Assumption 1. *θ is differentiable and decreasing in x_{i-1}.*

A positive assortative matching then implies that $x_{i+1} > x_i$, and higher level managers will have fewer problems to solve. If, in addition, hours worked are nondecreasing in type, the organization will have a pyramidal structure with fewer managers at higher levels. Since the production technology is constant returns to scale, we assume without loss of generality that there is only one manager at the top of the firm structure, that is, $n(x_I) = 1$. As a result, the time constraint of the top manager becomes

$$n_0\theta\left(x_{I-1}\right) = \ell_I. \tag{2}$$

Garicano (2000) and Garicano and Rossi-Hansberg (2006) present a special case of this time constraint with $\theta(x_{i-1}) = h[1 - F(x_{i-1})]$. The time constraint is derived as

[1]For each worker, $F(x_0)$ problems are solved by the worker himself without help of a manager, $F(x_1) - F(x_0)$ problems are explained by the manager in the first layer to the worker (and solved by the worker), $F(x_i) - F(x_{i-1})$ are problems explained by the manager in layer i to his/her subordinates, and $1 - F(x_I)$ problems remain unsolved.

follows. Since the subordinate managers solve a fraction $F(x_{i-1})$ of problems, they pass the remaining fraction of problems $1 - F(x_{i-1})$ ones to their superiors. Managers spend time dealing with a problem regardless of whether they know the answer or not. Each problem has a fixed communication cost h units of time, and so the time cost per production worker is $h[1 - F(x_{i-1})]$. Our setup generalizes Garicano (2000) and Garicano and Rossi-Hansberg (2006) in allowing for cases where the communication cost h itself depends on the skills of the subordinate.[2]

Constraints (1) and (2) show one of the key properties of the model. Having production workers with higher knowledge allows the manager to form larger teams and multiply their production. That is, there is a *skill complementarity* between the knowledge of a manager, and knowledge of a worker. Moreover, dividing both sides of the constraints by ℓ_i, one can see that it is only the ratio of the time cost θ and manager's hours ℓ_m that matters for the time constraint. A change in either one of these variables means that the manager is able to manage a team of workers of a different size. Thus,

Remark 2. *A decrease in manager's hours ℓ_i is equivalent to an upward shift in the time cost function θ.*

The overall production of the organization is linear in the multiple of the number of workers, output per hours for each worker, and hours worked of each worker. Since the organization is able to explain to the workers all the problems easier than x_I, the overall production is

$$y = n_0 F(x_I) \ell_0.$$

Let $w(x_0)$ be the hourly wage rate of a production worker with skill x_i, and $w(x_i)$ be the hourly wage rate of a level-i manager with knowledge x_i. They are determined as follows. The payoff of a top manager with knowledge x_I that employs a production

[2]An alternative specification would have θ depend on the skill of the manager x_i as well. If higher skilled managers are more efficient in communicating their knowledge then the communication cost decreases in x_i. On the other hand, if the manager is able to identify which problems he is not able to solve and does not waste time trying to solve them, then $\theta(x_{i-1}, x_i) = h[F(x_i) - F(x_{i-1})]$ increases in x_i.

worker with skill x_0 and subordinate managers with knowledge $(x_1, \ldots, x_{i-1})$ are

$$\Pi = F(x_I)n_0\ell_0 - w(x_0)n_0\ell_0 - w(x_1)n_1\ell_1 \ldots - w(x_{I-1})n_{I-1}\ell_{I-1}.$$

The top manager's hourly wage rate is $w(x_I) = \frac{\Pi}{\ell_I}$. Using the time constraints to substitute away the number of production workers and of the intermediate managers yields an alternative expression for the top managers' wage:

$$w(x_I) = \frac{F(x_I) - w(x_0)}{\theta(x_{I-1})}\ell_0 - \frac{\theta(x_0)}{\theta(x_{I-1})}w(x_1) \ldots - \frac{\theta(x_{I-2})}{\theta(x_{I-1})}w(x_{I-1}). \tag{3}$$

The top manager's hourly wage rate, and so his profits, increase linearly with hours worked of the production worker. This is because the manager keeps a fraction of output $F(x_I) - w(x_0)$ from each hour that the worker spends by working. Equation (3) shows a second key complementarity in the model: there is *working time complementarity* between the hours worked of a worker, and hours worked of a top manager.

The government taxes individual earnings by a tax function $T(y)$ regardless of whether the earnings are earned by production workers or managers. We assume that the tax function $T(y)$ exhibits a constant rate of progressivity (Benabou (2002), Heathcote et al. (2014), Heathcote et al. (2017), Kapička (2020)),

$$T(y) = y - \lambda y^{1-\tau},$$

where the wedge τ determines the progressivity of the tax system and the level parameter λ of the tax function is chosen in such a way that the government budget constraint holds,

$$\mathbb{E}_y T(y) = G,$$

where G is government consumption, exogenously given. To simplify notation, we introduce the retention function $\Gamma(y) = \lambda y^{1-\tau}$ to be the after tax income as a function of pre-tax income.

3.1 The Equilibrium

We consider an equilibrium in which agents choose to become managers or production workers. There is also a version of the model in which agents have the option of becoming self-employed. In this version of the model, the self-employed agents do not form teams, they solve the problems they can on their own and produce the appropriate output. Many features of both models are similar.

Assignment. We start the description of the equilibrium conditions by characterizing the assignment of workers to managers. Agents with skills between z_0 and z_1 will become production workers. Agents with skills between z_i and z_{i+1} will become level i managers. Agents with skills above z_I will become top managers. The workers and the top managers are special: the workers because only they produce output, and the top managers because they are residual claimants. There is assortative matching, where the worst production worker is matched with the worst managers, and the best production worker is matched with the best managers. Let $m(x_i)$ for $x_i \in [z_{i-1}, z_i]$ be the knowledge of the manager at level $i+1$ that employs a subordinate of knowledge x_i (either a lower level manager, or a production worker). We extend the function on the whole space by defining $m(x_I) = x_I$ for $x_I \geq z_I$. The matching function is illustrated in Figure 1.

The equilibrium assignment matches the worst workers with the worst managers of each level:

$$m(z_i) = z_{i+1}, \quad i = 0, \ldots, I. \tag{4}$$

We do not justify here that the equilibrium assignment takes this form. The reader is referred to Garicano (2000) for such a justification.

We require the supply of subordinate workers (either production workers or managers) to be equal to the demand for subordinate workers. Equivalently, the demand for superiors by their production teams has to be equal to the supply of superiors. Let

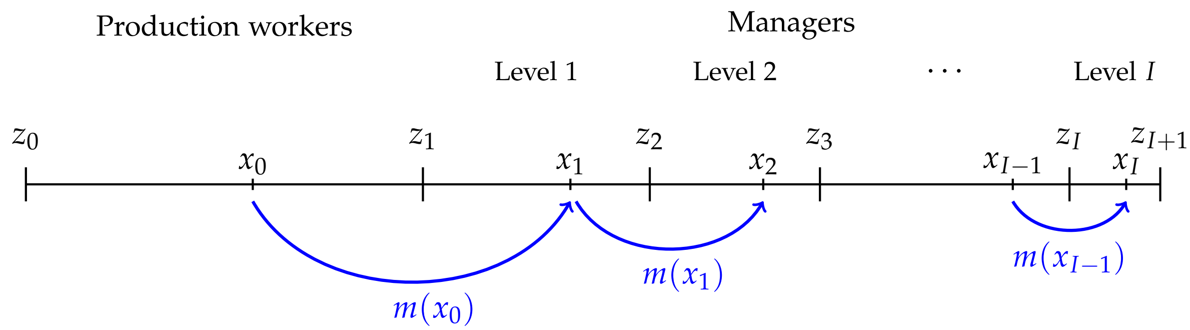


Figure 1: The equilibrium assignment $m(x)$ of subordinates to their immediate superiors. The subordinates are either workers or managers.

$n(x_i)$ for $x_i \in [z_{i-1}, z_i]$ be the number of direct subordinates of managers with skill $m(x_i)$. That is, $n(x_i)$ is the size of a team of workers or managers with knowledge x_i. Then the market clearing condition is

$$\int_{z_i}^{x} \frac{g(t)}{n(t)}\, dt = \int_{z_{i+1}}^{m(x)} \frac{g(t)}{n(t)}\, dt, \quad x \in [z_i, z_{i+1}] \quad i = 0, \ldots, I-1. \tag{5}$$

The left hand side is the demand for managers who have knowledge between z_{i+1} and $m(x)$ by their subordinate workers or managers with knowledge between z_i and x. The right-hand side is the supply of those managers by the organization.

Top managers. A top manager choose hours worked ℓ, but also the vector of skills of her subordinates $(x_0, x_1, \ldots, x_{I-1})$ so as to maximize her own wage rate $w(x_I)$. When making the choice, the top manager takes hours worked and wages of her subordinates as given. The wage rates of the subordinates are given by (3), and so the problem to maximize the top manager's wage rate is

$$\left[m^{-I}(x_I), \ldots, m^{-1}(x_I)\right] \in \arg \max_{x_0, \ldots, x_{I-1}} \left[\frac{F(x_I) - w(x_0)}{\theta\left(x_{I-1}\right)} \ell_0(x_0) - \sum_{i=1}^{I-1} \frac{\theta\left(x_{i-1}\right)}{\theta\left(x_{I-1}\right)} w(x_i)\right]. \tag{6}$$

The maximization problem uses the fact that, by the definition of the assignment function m, a top manager with skill x_I chooses a level-i subordinate $m^{I-i}(x^I)$.

Conditional on the wage rate $w(x_I)$, the hours worked are chosen in a standard way to maximize their utility:

$$\ell(x_I) \in \arg\max_{\ell} U\left[\Gamma(\ell w(x_I))\right] - V(\ell), \quad z_I \leq x_I \leq x_{I+1}. \tag{7}$$

Production workers and intermediate level managers. Production workers and intermediate level managers have only one choice. They face a wage rate $w(x_i)$ and choose hours worked $\ell(x_i)$ to solve

$$\ell(x_i) \in \arg\max_{\ell} U\left[\Gamma(\ell w(x_i))\right] - V(\ell), \quad z_i \leq x_i \leq z_{i+1}. \tag{8}$$

Finally, we require that the marginal agents with the threshold knowledge z_i for $i = 1, \dots, I-1$ must be indifferent between being the best at the lower level, and being the worst at the higher level. Given that the agents simply choose hours worked given the wage, they will be indifferent between both options if the wage function is continuous at the thresholds.

Aggregates. Aggregte output in the economy consists of the production of workers (recall that managers do not directly produce output):

$$Y = \int_{\underline{z}}^{z_1} \ell(t) F(m_I(t)) g(t)\, dt.$$

Aggregate consumption in the economy is the sum of total consumption of production workers, and of managers:

$$C = \int_{\underline{z}}^{\overline{z}} \Gamma\left(w(t)\ell(t)\right) g(t)\, dt.$$

By Walras Law, the requirement that the government budget constraint holds can be expressed as $C + G = Y$.

Definition 1. *Given θ, F and G, the equilibrium consists of threshold values z, matching function $m : [z_0, z_{I-1}] \to [z_1, z_I]$, wage function $w : [z_0, z_I] \to \mathbb{R}_+$ and hours worked $\ell : [z_0, z_I] \to \mathbb{R}_+$ such that m satisfies (4) (5) and (6), ℓ satisfies (8) and (7), w is continous at z, and the government budget constraint holds.*

Before proceeding further and characterizing the equilibrium, we will show that the model can be substantially simplified: without loss of generality, we can normalize the skill distribution to be uniform. This normalization is based on the following proposition:

Proposition 1. *The allocation z,m,w and l constitutes the equilibrium given θ, F and G if and only if $\tilde{z} = G(z)$, $\tilde{m}(p) = G\left(m(G^{-1}(p))\right)$, $\tilde{w}(p) = w\left(G^{-1}(p)\right)$ and $\tilde{\ell}(p) = \ell\left(G^{-1}(p)\right)$ constitute an equilibrium given $\tilde{\theta}(p) = \theta\left(G^{-1}(p)\right)$, $\tilde{F}(p) = F\left(G^{-1}(p)\right)$ and $\tilde{G} = p$, where $p = G(x)$ are percentiles of the skill distribution.*

Proof. The matching function m and thresholds z satisfy (4) if and only if $\tilde{m}$ and $\tilde{z}$ satisfy (4). To show that (5) holds given $\tilde{\theta}$ and $\tilde{G}$, rewrite (5) for $i = 1, \ldots, I-1$ and $x \in [z_i, z_{i+1}]$ as follows:

$$
\begin{aligned}
0 &= \int_{z_i}^{x} \frac{g(t)}{\theta\left(m^{-1}(t)\right)}\, dt - \int_{z_{i+1}}^{m(x)} \frac{g(t)}{\theta\left(m^{-1}(t)\right)}\, dt \\
&= \int_{G(z_i)}^{G(x)} \frac{1}{\theta\left(m^{-1}(G^{-1}(q))\right)}\, dq - \int_{G(z_{i+1})}^{G(m(x))} \frac{1}{\theta\left(m^{-1}(G^{-1}(q))\right)}\, dq \\
&= \int_{\tilde{z}_i}^{p} \frac{1}{\tilde{\theta}\left(G(m^{-1}(G^{-1}(q))\right)}\, dq - \int_{\tilde{z}_{i+1}}^{\tilde{m}(p)} \frac{1}{\tilde{\theta}\left(G(m^{-1}(G^{-1}(q))\right)}\, dq \\
&= \int_{\tilde{z}_i}^{p} \frac{1}{\tilde{\theta}\left(\tilde{m}^{-1}(q)\right)}\, dq - \int_{\tilde{z}_{i+1}}^{\tilde{m}(p)} \frac{1}{\tilde{\theta}\left(\tilde{m}^{-1}(q)\right)}\, dq,
\end{aligned}
$$

where the first line changes the variable of integration from t to $q = G(t)$, the second line replaces the limits from $G(x)$ and z_i to p and $\tilde{z}_i$ and uses the definition of $\tilde{\theta}$, and the last line uses the definition of $\tilde{m}_i$. Identical arguments show that (5) holds for $i = 0$, in which case.

$$
0 = \int_{z_0}^{x} g(t)\, dt - \int_{z_1}^{m(x)} \frac{g(t)}{\theta\left(m^{-1}(t)\right)}\, dt = \int_{0}^{p} dq - \int_{\tilde{z}_1}^{\tilde{m}(p)} \frac{1}{\tilde{\theta}\left(\tilde{m}^{-1}(q)\right)}\, dq.
$$

Hence, (5) continues to hold. To see that $\tilde{z}$,$\tilde{m}$,$\tilde{w}$ and $\tilde{l}$ satisfies (6) given $\tilde{\theta}$ and $\tilde{F}$, rewrite the right-hand side of (6)

$$\frac{F(x_I) - w(x_0)}{\theta\left(x_{I-1}\right)}\ell_0(x_0) - \sum_{i=1}^{I-1}\frac{\theta\left(x_{i-1}\right)}{\theta\left(x_{I-1}\right)}w(x_i) = \frac{\tilde{F}(p_I) - \tilde{w}(p_0)}{\tilde{\theta}\left(p_{I-1}\right)}\tilde{\ell}_0(p_0) - \sum_{i=1}^{I-1}\frac{\tilde{\theta}\left(p_{i-1}\right)}{\tilde{\theta}\left(p_{I-1}\right)}\tilde{w}(p_i). \tag{9}$$

Since the left-hand side of (3.1) is maximized by $[m^{-I}(x_I), \ldots, m^{-1}(x_I)]$, the right-hand side is maximized by

$$\begin{aligned}\left[G\left(m^{-I}(x_I)\right), \ldots, G\left(m^{-1}(x_I)\right)\right] &= \left[G(m^{-I}(G^{-1}(p_I))), \ldots, G(m^{-1}(G^{-1}(p_I)))\right] \\ &= \left[\tilde{m}^{-I}(p_I), \ldots, \tilde{m}^{-1}(p_I)\right].\end{aligned}$$

Hence (6) holds as well. It is straightforward to show that $\tilde{\ell}$ satisfies (8) and (7), $\tilde{w}$ is continuous at $\tilde{z}$, and that the government budget constraint holds as well. Hence if (z, m, w, ℓ) constitutes an equilibrium given (θ, F, G), then $(\tilde{z}, \tilde{m}, \tilde{w}, \tilde{l})$ constitutes an equilibrium given $(\tilde{\theta}, \tilde{F}, \tilde{G})$. Since all operations are equivalent, the reverse implication holds as well. ■

Proposition 1 says that a change in the underlying distribution of skills can be always represented as a joint transformation of the time cost function θ, and of the task arrival distribution F. There is nothing in the model that allows us to distinguish between the two. Equivalently, we can express the problem in the percentiles of the underlying distribution G, and transform θ and F appropriately. This not only simplifies the problem technically but, as we shall see, allows us to characterize its properties more sharply. This is so because most of the properties of the equilibrium matching and wage functions might be ambiguous when expressed as functions of the underlying skills, but they gain clarity when expressed as functions of the percentiles. We will henceforth assume:

Assumption 3. *G is uniform on $[0, 1]$.*

In what follows, we will impose various assumptions on θ and F in the normalized problem. In the light of Proposition 1 , those should be understood as joint assumptions on θ and F, and G. For example, assuming that $\tilde{\theta}$ satisfies Assumption 1 is equivalent to assuming that θ is decreasing in x and g is increasing in x, or that θ is increasing in x and g is decreasing in x. Similarly, assumptions about $\tilde{F}$ translate into joint assumptions about F and G in the original problem.[3]

4 Characterizing the Equilibrium

We now characterize the equilibrium of the model. First, it is easy to show that, given that the utility is logarithmic in consumption, income and substitution effects cancel out, and the agents choose hours worked that are independent of their knowledge. Everyone's hours worked are given by $\ell(z) = \bar{\ell}(\tau)$ where

$$\bar{\ell}(\tau) = \left(\frac{1-\tau}{\kappa}\right)^{\frac{1}{1+\eta}}. \tag{10}$$

The fact that hours worked are constant across all agents allows us to substantially simplify the problem. It is only the ratio of communication costs $\theta(\cdot)$ and hours worked that matters from the wage distribution rather than $\theta(\cdot)$ and $\bar{\ell}(\tau)$ individually (or the particular values of κ, η and τ). That is, we can normalize ℓ to one for all agents and redefine the communication cost function by setting it equal to $\theta(\cdot)/\bar{\ell}(\tau)$. The wage rate and rent rate schedule satisfy the following property: $w\left(z; \ell(\tau), \theta(\cdot)\right) = w\left(z; 1, \theta(\cdot)/\bar{\ell}(\tau)\right)$, and the earnings or each agent are $\bar{\ell}(\tau)$ times wages or rents. We can then characterize the equilibrium wage and rent distribution. Any changes in hours worked due to a change in taxes will manifest themselves as a change in the communication costs θ.

[3]From a practical perspective, the normalization is perhaps less important. In our quantitative analysis of section 7 and 8 we find it convenient to normalize F to be uniform on $[0, 1]$ and calibrate a flexible distribution of skills G.

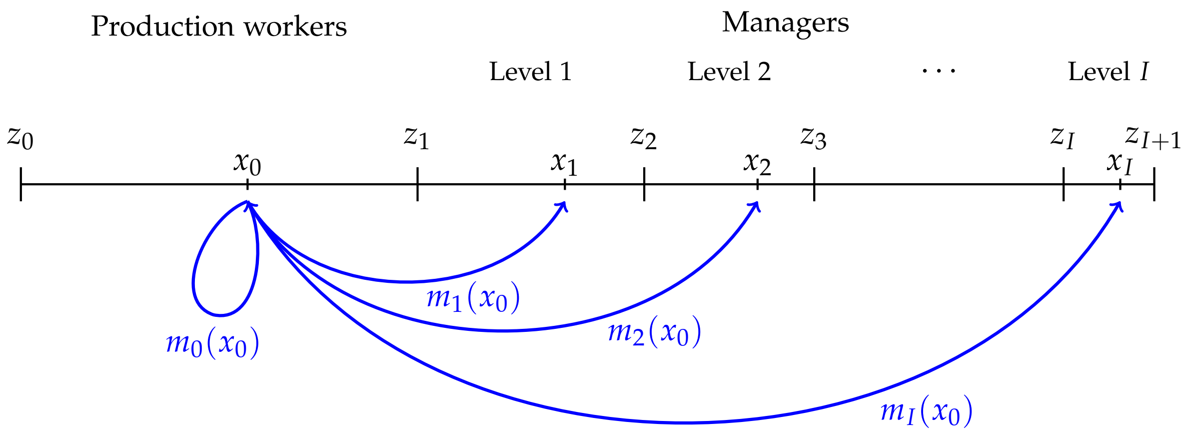


Figure 2: The equilibrium assignment $m_i(x_0)$ of workers to their superiors in level i.

Setting $\bar{\ell}(\tau) = 1$ and differentiating the market clearing condition (5) with respect to x yields a differential equation in m:

$$m'(x_0) = \theta(x_0) \tag{11a}$$

$$m'(x_i) = \frac{\theta(x_i)}{\theta(x_{i-1})}, \quad i = 1, \ldots, I, \tag{11b}$$

where we used the time constraints (4) to rewrite the expression. The rate at which managers skill increases with workers skill thus depends only on the relative density of both, and on the time cost function θ .

A second relationship between both thresholds is obtained from the managers' problem of choosing the type of his subordinates. Solving the managers' problem (6) yields a differential equation in production workers' wages

$$w'(x_0) = -\theta'(x_0)w(x_1) \tag{12a}$$

$$w'(x_i) = -\frac{\theta'(x_i)}{\theta(x_{i-1})}w(x_{i+1}). \tag{12b}$$

The intuition behind equation (12a) is very simple. Consider the marginal costs and marginal benefits of choosing a slightly better type. The marginal cost is the marginal increase in the production worker's wage, $w'(x_0)$. The marginal benefit is that a better production worker can solve more tasks himself, and saves time to his superior. The time saved is $-\theta'(x_0)$. What is the value of one unit of time for the superior? It is exactly his wage rate, $w(x_1)$. In the optimum, the marginal costs of a better production worker are equated to the marginal benefits, i.e. (12a) holds.

Rewriting the problem. It turns out that the equilibrium assignment can be easier to characterize by using matching functions that map the worker's knowledge x_0 directly to the knowledge of the manager in each layer. To that end, define a function $m_i(x_0)$ for $i = 0, \dots, I$ recursively by $m_0(x_0) = x_0$ and $m_{i+1}(x_0) = m(m_i(x_0))$. The function $m_i(x_0)$ represents the knowledge of a manager in layer i that is matched with a worker with knowledge x_0, as figure 2 illustrates. By the equilibrium assignment (4), one has

$$m_i(\underline{z}) = z_i, \quad m_i(z_0) = z_{i+1}. \tag{13}$$

Multiplying both sides of the differential equations (11) by $g(m(x_i))$, integrating and using the equilibrium conditions (13), we can then write the equilibrium assignment function as

$$m_i(x) = z_i + \rho_{i-1}(x), \quad i = 1, \dots, I, \tag{14}$$

where the function ρ_i is given by

$$\rho_i(x) = \int_{\underline{z}}^{x} \theta\left(m_i(t)\right) dt, \quad x \in [\underline{z}, z_1].$$

By differentiating (14), we immediately obtain that the functions m_i are differentiable and increasing in x. They are also concave if the time cost function θ satisfies Assumption 1.

Lemma 1. *Suppose that Assumption 1 holds. Then the matching function m_i is differentiable, increasing and concave for all $i = 1, \ldots, I$.*

Proof. Differentiating (14) with respect to x, we obtain that m_i is differentiable, with a derivative $m_i'(x) = \theta\left(m_{i-1}(x)\right)$, which is increasing in x. Differentiating again for $i = 1$, $m_1''(x) = \theta'(x)$, and, since θ is decreasing in x, m_1 is concave in x. Differentiating for $i = 2, \ldots, I$, $m_i''(x) = \theta'(m_{i-1}(x))m_{i-1}'(x)$, which is also negative, because m_{i-1}' is positive. ■

The main force the matching function is that more productive workers require less supervision, and so demand fewer managers. If the mass of production workers increases by one unit, the mass of managers that are needed to match with them must increase by less than one unit. This creates a concavity in the matching function.

Evaluating (14) at the thresholds z and using (4) yields a unique equilibrium condition for the threshold values:

$$z_{i+1} = z_i + \rho_{i-1}(z_1) \quad i = 1, \ldots, I. \tag{15}$$

where we take $z_{I+1} = 1$ in the last equation. Summing over, the equilibrium value of z_1 satisfies

$$1 - z_1 = \rho(z_1), \tag{16}$$

where $\rho(z_1) = \sum_{i=1}^{I} \rho_{i-1}(z_1)$ is only a function of z_1. Equation (16) determines the equilibrium value of z_1. Figure 3 illustrates how z_1 is determined. Since $\rho_{i-1}'(z) = \theta\left(m_i(z)\right) > 0$, $\rho(z)$ is strictly increasing and concave in z. Since the left-hand side of (16) starts at one, is strictly decreasing in z and ends at zero, there is a unique value of z_1 that satisfies the equilibrium assignment. At $z_1 = 0$, the left-hand side is strictly positive, while the right-hand side is zero, and at $z_1 = 1$, the left-hand side is zero, while the right-hand side is strictly positive. Thus, there is a unique solution to the equilibrium

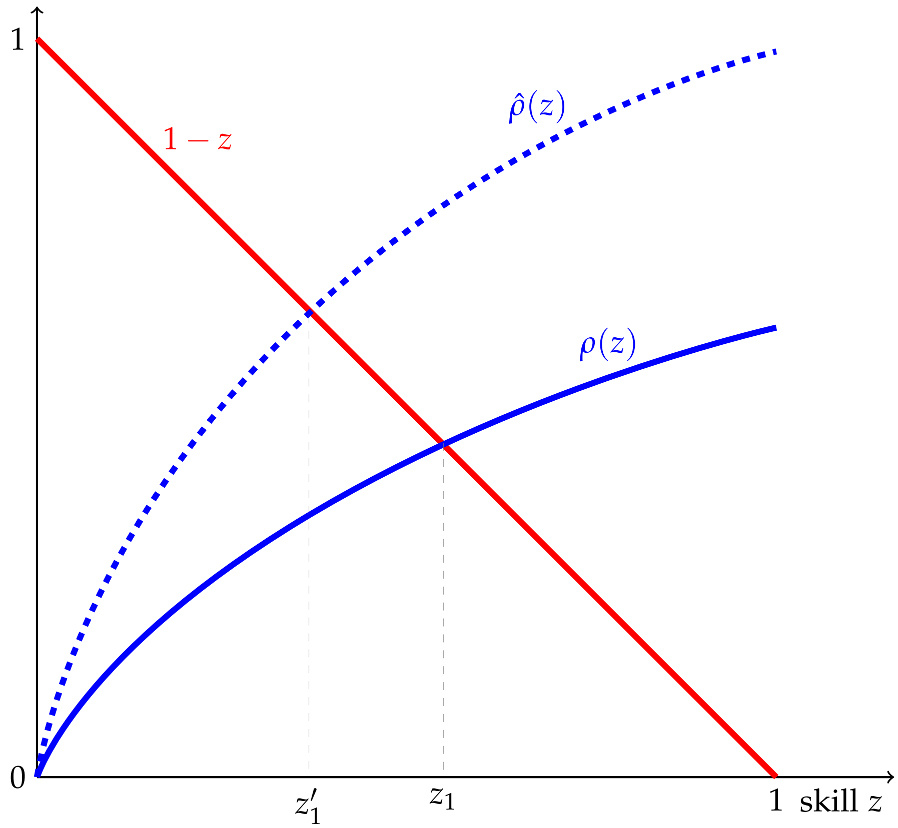


Figure 3: The equilibrium value of z_1. An increase in the time cost from θ to $\theta' > \theta$ increases ρ to $\hat{\rho} > \rho$ and decreases the equilibrium z_1.

condition (16). Moreover, when the time cost function increases from θ to $\hat{\theta} > \theta$,[4] the threshold knowledge z_1, and hence the fraction of workers, unambiguously decreases. Figure 3 illustrates the comparative statics. To summarize,

Proposition 2. *There is a unique threshold knowledge* $z_1 \in [0,1]$ *that solves (16). Moreover,* z_1 *is decreasing in* θ*.*

As with the matching function, we again find it easier to transform the wage function $w(x)$ to a sequence of functions mapping the worker type x_0 to the wage of his superiors. Define $w_i(x_0) = w(m_i(x_0))$ to be the wage of level-i manager that manages workers of type x_0. Note that $w_0(x_0) = w(x_0)$ is directly the wage of the production worker.

[4] $\hat{\theta} > \theta$ if $\hat{\theta}(x) > \theta(x)$ for all $x \in [0,1]$.

Moreover, the continuity of the function w implies

$$w_i(z_1) = w_{i+1}(\underline{z}) \quad i = 0, \dots I-1. \tag{17}$$

Differentiating the functions w_i and using the differential equations (12) yields the following differential equation for wages

$$w_i'(x) = -\theta'\left(m_i(x)\right) w_{i+1}(x), \quad i = 0, 1, \dots, I. \tag{18}$$

The differential equation is to be solved together with the residual definition of the top manager's wage,

$$w_I(x) = \frac{F\left(m_I(x)\right) - w_0(x)}{\theta\left(m_{I-1}(x)\right)} - \frac{\theta\left(x\right)}{\theta\left(m_{I-1}(x)\right)} w_1(x) \dots - \frac{\theta\left(m_{I-2}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_{I-1}(x). \tag{19}$$

Lemma 2. *The wage functions w_i are increasing in x for all $i = 0, 1, \dots, I+1$. If $\theta'' < 0$ and $f' \geq 0$ then the wage functions are convex for $i = 0, 1, \dots, I+1$.*

Proof. Differentiating the function w_I, we obtain

$$\begin{aligned} w_I'(x) = &-\frac{\theta'\left(m_{I-1}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_I(x) + f\left(m_I(x)\right) \frac{m_I'(x)}{\theta\left(m_{I-1}(x)\right)} - \frac{w_0'(x)}{\theta\left(m_{I-1}(x)\right)} \\ &- \frac{\theta'(x) w_1(x) + \theta(x) w_1'(x)}{\theta\left(m_{I-1}(x)\right)} - \frac{\theta'(m_1(x)) m_1'(x) w_2(x) + \theta(m_1(x)) w_2'(x)}{\theta\left(m_{I-1}(x)\right)} \\ &\dots - \frac{\theta'(m_{I-2}(x)) m_{I-2}'(x) w_{I-1}(x) + \theta(m_{I-2}(x)) w_{I-1}'(x)}{\theta\left(m_{I-1}(x)\right)}. \end{aligned}$$

Using the fact that $m_i'(x) = \theta\left(m_{i-1}(x)\right)$ and (18), we simplify to

$$\begin{aligned} w_I'(x) &= -\frac{\theta'\left(m_{I-1}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_I(x) + f\left(m_I(x)\right) \frac{m_I'(x)}{\theta\left(m_{I-1}(x)\right)} + \frac{\theta\left(m_{I-2}(x)\right) \theta'\left(m_{I-1}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_I(x) \\ &= f\left(m_I(x)\right) - \theta'\left(m_{I-1}(x)\right) \frac{1 - \theta\left(m_{I-2}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_I(x). \end{aligned}$$

Since $\theta < 1$ and $\theta' < 0$, the derivative is positive, and so w_I is increasing in x. Differentiating again,

$$w_I''(x) = f'\left(m_I(x)\right) m_I'(x) - \theta''\left(m_{I-1}(x)\right) \frac{1-\theta\left(m_{I-2}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_I(x) - \theta'\left(m_{I-1}(x)\right) \frac{1-\theta\left(m_{I-2}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_I'(x)$$
$$- \theta'\left(m_{I-1}(x)\right) w_I(x) \frac{-\theta'\left(m_{I-2}(x)\right) m_{I-2}'(x) \theta\left(m_{I-1}(x)\right) - \theta'\left(m_{I-1}(x)\right) m_{I-1}'(x)\left[1-\theta\left(m_{I-2}(x)\right)\right]}{\theta\left(m_{I-1}(x)\right)^2}.$$

If $f' \geq 0$ then all the terms on the right-hand side are positive, and so w_I is convex.

Wages in the layers below the top layer are increasing because of (18). Differentiating again, we get that for $i = 0, 1, \ldots, I$, we get

$$w_i''(x) = -\theta''\left(m_i(x)\right) w_{i+1}(x) - \theta'\left(m_i(x)\right) w_{i+1}'(x).$$

Under the assumptions of the lemma, $w_i'' > 0$ and so the wage function is convex. ■

Lemma 2 shows that the functions w_i are all convex. Furthermore, Lemma 1 implies that convexity will be further magnified if one expresses wages a function of the managers' own skill, rather than the skill of the production workers they are matched with. This is so because $w(x_i) = w_i(m_i^{-1}(x_i))$, and m_i is concave by Lemma 1.

In addition, convexity will be more pronounced in the lower layers. A given inequality in a layer i will be convexified in the subordinate layer $i-1$, because, by (18), the slope of the wage function w_{i-1}' decreases with the manager's skill in layer i. For example, if w_I is linear in x (which will happen if $I = 1$ and F is uniform), then w_{I-1} will tend to be quadratic in x.

5 An Example

We now solve for the equilibrium in a model with one managerial layer ($I = 1$). We assume that the underlying distributions F and G are both uniform, with $\bar{z} = 1$, $\underline{z} = 0$ and $\theta(x) = h[1 - F(x)]$. This configuration delivers a distribution of wages with a Pareto right tail, see Geerolf (2017). The cumulative distribution functions are

$$F(z) = G(z) = z.$$

The problem then has a closed form solution. The function ρ is given by $\rho(z) = h[1-(1-z)^2]/2$. Equation (16) is now a quadratic equation in $1-z_1$,[5] with the correct solution

$$z_1 = 1 - \frac{\sqrt{1+h^2}-1}{h}. \tag{20}$$

The threshold value z_1 is clearly decreasing in h, confirming the results of Proposition 2. Higher communication cost h thus increase sthe fraction of managers in the economy, because it is more costly to supervise the production workers. The matching function is quadratic and concave in x:

$$m_1(x) = z_1 + hx - \frac{h}{2}x^2. \tag{21}$$

An increase in the communication cost h makes the matching function steeper, since $m_1'(x) = h(1-x)$ increases in h. This is intuitive, since a smaller mass of workers is now matched with a larger mass of managers, and so workers' skills must be spread over a larger span of managers' skills. In other words, a worker that has only a small skill advantage over another worker will now be matched with now gain a larger advantage by being matched with a comparatively better manager. This will, as we shall see, have important implications for the wage structure.

[5] The quadratic equation is $h(1-z_1)^2 + 2(1-z_1) - h = 0$.

The wage function of the production workers w_0 is quadratic in z, while the wage function of the managers w_1 is linear in x:

$$w_0(x) = z_1 + A(x-1) + \frac{h}{2}x^2 \tag{22}$$

$$w_1(x) = \frac{A}{h} + x, \tag{23}$$

where A is only a function of h,

$$A = \frac{1+(1+h)^2}{\sqrt{1+h^2}} - 2 - h.$$

It is easy to verify that A is not only positive, but also increasing in h. However, A/h is decreasing in h. An increase in h affects the wage structure in two ways. First, there is an *absolute effect*: an increase in h increases the cost of creating teams, and shifts wages of everyone down. Equation (22) shows that the wage decreases for any given production worker, and so the whole wage function w_0 shifts down. To look at the wages of managers, it is useful to transform the managers wages w_1 back to a function of the manager's wage $w(x_1)$. The conversion yields

$$w(x_1) = \frac{A}{h} + 1 - \sqrt{1 - \frac{2}{h}(x_1 - z_1)}, \quad z_1 \leq x_1 \leq 1.$$

Differentiating, we find that wages for any given manager decrease as well. The absolute effect thus decreases wages for all.

Second, an increase in h produces a *relative effect*: individuals at different parts of a distribution are affected differently. The effect is asymmetric. While the wages of production workers become more unequal, the wages of managers become less unequal. To see this, note that, since $w_0'(x) = A + hx$, the wage function becomes steeper and so the inequality among individuals who continue being production workers is magnified. This so happens, because better workers are matched with relatively more productive managers, and a part of the relative efficiency gains is translated into wages. The wage of the best production worker also decreases. On the other hand, $w'(x_1)$ is decreasing in h, and so the wage differences among managers move in the opposite direction and become smaller. Managers' wage schedule becomes flatter for the same reason why workers' wage schedule becomes steeper. Two managers of given skills are now matched with more similar workers, and their productivity differences decrease.

The aggregate output of the economy has a closed form solution given by

$$Y = \int_0^{z_1} m_1(t)\,dt = \frac{1}{3}\left[z_1\left(h + 2(1+z_1)\right) - 1\right].$$

Differentiating, we get that the aggregate output is decreasing in the communication cost h. Higher h thus has a clear negative effect on aggregate output.

Lemma 3. *The aggregate output Y is decreasing in h.*

Proof. Differentiating Y with respect to h, we get

$$\frac{dY}{dh} = \frac{1}{3}\left[z_1 + (2 + h + 4z_1)\frac{dz_1}{dh}\right],$$

where

$$\frac{dz_1}{dh} = \frac{1-\sqrt{1+h^2}}{h^2\sqrt{1+h^2}} < 0.$$

Since $z_1 dz_1/dh < 0$, the Lemma will be proven if $z_1 + (2+h)dz_1/dh < 0$. To show this, write

$$z_1 + (2+h)\frac{dz_1}{dh} = \frac{2-h^2-2\sqrt{1+h^2}}{h^2\sqrt{1+h^2}} + 1 = \frac{2-h^3-(2-h^2)\sqrt{1+h^2}}{h^2\sqrt{1+h^2}}$$

The numerator is equal to zero for $h = 0$ and is decreasing in h, as one can verify by differentiating. Thus, the numerator is negative. Since the denominator is positive, the whole expression is negative, finishing the proof. ∎

While we were not able to obtain closed form expression for the variance of log wages, we verified numerically, that the variance of log wages increases with h. This is expected, since higher communication cost increases wage inequality among the production workers who do not switch occupation and, by the nature of a logarithmic transformation, this effect dominates a decrease in inequality among the managers who did not switch occupations.

6 Optimal progressivity

We now characterize the optimal value of the progressivity parameter τ, and its determinants. Let $\mathbb{E}[w|\tau] = Y[1, h/\ell(\tau)]$ be the average wage and rental rate, and $\mathbb{E}[w^{1-\tau}|\tau] = C[1, h/\ell(\tau)]/\lambda$ be the average of wages and rents to the power $1-\tau$. Putting back labor supply $\bar{\ell}(\tau)$ given by (10), we can write the resource constraint as

$$\lambda\bar{\ell}(\tau)^{1-\tau}\mathbb{E}[w^{1-\tau}|\tau] + G = \bar{\ell}(\tau)\mathbb{E}[w|\tau].$$

Solving for the equilibrium λ and substituting back to the expected utility yields the welfare for a given progressivity wedge τ

$$\mathcal{W} = \ln\left[\bar{\ell}(\tau)\mathbb{E}[w|\tau] - G\right] - \frac{1-\tau}{1+\eta} - \ln\mathbb{E}[w^{1-\tau}|\tau] + (1-\tau)\mathbb{E}[\ln w|\tau]. \tag{24}$$

The expression has a standard form, but the moments of the wage distribution are not exogenous, but depend on τ.

To further inspect the novel role of taxes in determining the wage distribution, we approximate the penultimate term in (24) as follows:

$$
\begin{aligned}
\ln \mathbb{E}[w^{1-\tau}|\tau] &= \ln \mathbb{E}[e^{(1-\tau)\ln w}|\tau] \\
&\approx (1-\tau)\mathbb{E}\left[\ln w|\tau\right] \\
&\quad + \ln \mathbb{E}\left[1 + (1-\tau)\left(\ln w - \mathbb{E}[\ln w|\tau]\right) + \frac{(1-\tau)^2}{2}\left(\ln w - \mathbb{E}[\ln w|\tau]\right)^2|\tau\right] \\
&\approx (1-\tau)\mathbb{E}\left[\ln w|\tau\right] + \ln\left[1 + \frac{(1-\tau)^2}{2}\mathbb{E}\left[\left(\ln w - \mathbb{E}\ln w\right)^2|\tau\right]\right] \\
&\approx (1-\tau)\mathbb{E}[\ln w|\tau] + \frac{(1-\tau)^2}{2}\mathbb{E}\left[\left(\ln w - \mathbb{E}\ln w\right)^2|\tau\right],
\end{aligned}
$$

where the second line uses a Taylor approximation around $\mathbb{E}\ln w$ and rearranges terms, and the last line uses a well known property of logarithm. The approximation is exact, if the distribution of wages is lognormal. We cannot, of course, assume that this is the case. Substituting into (24) and cancelling terms yields an approximate expression for welfare:

$$
\mathcal{W} \approx \ln\left[\bar{\ell}(\tau)\mathbb{E}\left[w|\tau\right] - G\right] - \frac{1-\tau}{1+\eta} - \frac{(1-\tau)^2}{2}\mathbb{V}[\ln w|\tau], \tag{25}
$$

where $\mathbb{V}[\ln w|\tau] = \mathbb{E}\left[\left(\ln w - \mathbb{E}\ln w\right)^2|\tau\right]$ is the variance of log wages. The expression (25) makes it clear how endogenous wage distribution affects welfare. First, and perhaps most importantly, it changes the mean of wages, $\mathbb{E}\left[w|\tau\right]$. Second, it can change the variance of log wages $\mathbb{V}[\ln w|\tau]$.

One might expect that a higher progressivity parameter τ, by decreasing hours worked and so increasing the effective communication cost $\theta/\bar{\ell}$, will decrease the average wage in the economy. Quantitatively, we find that increasing the progressivity parameter τ increases the variance of *log* wages by increasing the bottom wage inequality. These two forces suggest that the optimal tax progressivity will not be high, as we document in a calibrated version of the model below.

7 Model Calibration

We assume that $I = 1$, and so there is only one layer of management in the organization. Knowledge is distributed on a unit interval, with $\underline{z} = 0$ and $\overline{z} = 1$. The distribution of problems F is uniform (recall that this assumpiton is witout loss of generality due to 1 and the underlying distribution of skills G is polynomial:

$$F(x) = x, \quad G(x) = 1 - (1 - x)^{1+\rho}.$$

The case when $\rho = 0$ corresponds to the uniform distribution of skills. If $\rho > 0$ then skill density decreases with skills, while if $\rho < 0$ then it increases in skills. We consider the following time cost function θ:

$$\theta(x) = h(1 - x)^{\gamma} [1 - F(x)] = h(1 - x)^{1+\gamma}, \quad \gamma \geq 0.$$

One can interpret the communication cost as follows. Workers incur two types of costs on their managers. First, they need help with a larger fraction of problems, which is represented by the second term $1 - F(x)$. Second, the time needed to communicate each problem is $h(1 - x)^{\gamma}$, which is larger for lower skilled workers.The special case with $\gamma = 0$ correspond to the specification in Garicano (2000) or Garicano and Rossi-Hansberg (2006). As noted earlier, given γ and ρ, only the ratio of $h/\bar{\ell}(\tau)$ matters from the wage distribution.

We calibrate the model parameters ρ and γ and well as relative costs $h/\bar{\ell}$ to match three empirical moments: a fraction of individuals in managerial positions in the population, the 90/50 log wage ratio, and the 50/10 log wage ratio. The data are taken from 2018 CPS March supplement. The fraction of managers is 20.7 percent, the 90/50 log wage ratio is 0.916, and the log 50/10 ratio is 0.788. This yields $h/\bar{\ell} = 0.467$, $\rho = 1.387$ and $\gamma = 1.819$. A relatively high value of ρ means that the density of high skill is much smaller than the density of low skills. The decreasing density is needed to match especially the wage

Table 1: Baseline Parameters

κ	η	τ	ρ	γ	h
1.000	2.000	0.181	1.387	1.819	0.437

distribution at the top. While the worst production worker can solve none of tasks that arrive, the best production worker can solve about 48 percent of all tasks.

A relatively high value of γ means that a substantial amount of heterogeneity in the communication cost is needed to match the calibration target: communicating about a given task with the worst production worker takes about three times more time than communicating about it with the best production worker. The heterogeneity in the communication costs is needed to differentiate the worst production worker from the best production worker, and to match the wage inequality at the bottom of the distribution.

The calibration of γ and ρ and $h/\bar{\ell}(\tau)$ is independent of the rest of the parameters, but comparative statics and optimal taxes are, in general, not independent of the rest of the parameters. Hence, they need to be speficied. One exception is κ, which is a scaling factor, irrelevant due to log utility (notice that κ does not appear separately in equation (25), unlike η). We thus normalize κ to 1. We set $\eta = 2$ in the benchmark, impying a Frisch labor supply elasticity of 0.5. The current U.S. tax system can be approximated by $\tau = 0.181$ as estimated by Heathcote et al. (2017) (see also Guner et al. (2014) for estimates of this and other tax functions). As a result, $\bar{\ell} = 0.936$, and do $h = 0.467 \times 0.936 = 0.437$. The resulting benchmark parameters are in Table 1.

Figure 4 plots the resulting distribution of wages in the benchmark economy, and compares it to the empirical distribution of wages. Table 2 shows additional moments of the wage distribution in the benchmark economy (and uder the optimal tax progressivity $\tau = 0.197$, which is discussed below). Both distributions are remarkably close, despite the fact that we are matching only the 50-10 and 90-50 log wage ratio.

Table 2: Moments of the Wage Distribution

	Data	Benchmark	Optimum
Variance log wages	0.436	0.407	0.412
log 50/10 ratio	0.789	0.787*	0.797
log 90/50 ratio	0.916	0.916*	0.920
log 99/50 ratio	1.920	1.677	1.678

Moments with an asterisk are calibrated to match the corresponding empirical moments.

As both Figure 4 and Table 2 show, the model is slightly less succesful in matching the wage distribution at the very bottom and at the very top. At the bottom of the wage distribution, the model predicts larger wage gains than what is observed in the data, while at the very top of the wage distribution (about the top 3 percent), the model predicts somewhat thinner upper tail. For example, the log 99/50 wage inequality, not targeted by the model, is 1.920 in the data, but only 1.677 in the model. Overall, however, the model is very successful in producing a realistic distribution of wages.

The distribution of wages differs significantly from the underlying distribution of abilities. Agents at all skill levels experience significant welfare gains relative to autarky, where every agent would have a wage equal to $F(z)$. One can show that the gains are U-shaped in the agents' abilities. Lowest ability agents, who would otherwise produce nothing, gain from being matched with a manager that solves some of their problems. Very high ability agents gain from being able to supervise a relatively large number of production workers. Agents in the middle of the distribution gain as well, but their gains are smaller compared to the gains at both endpoints of the distribution.

7.1 Sources of Wage Inequality

The calibrated parameter values show that the model requires i) heterogeneity in the time cost of communication, and ii) a decreasing density of skills. Without either of those two ingredients, the model cannot simultaneously match both the 90/50 and 50/10

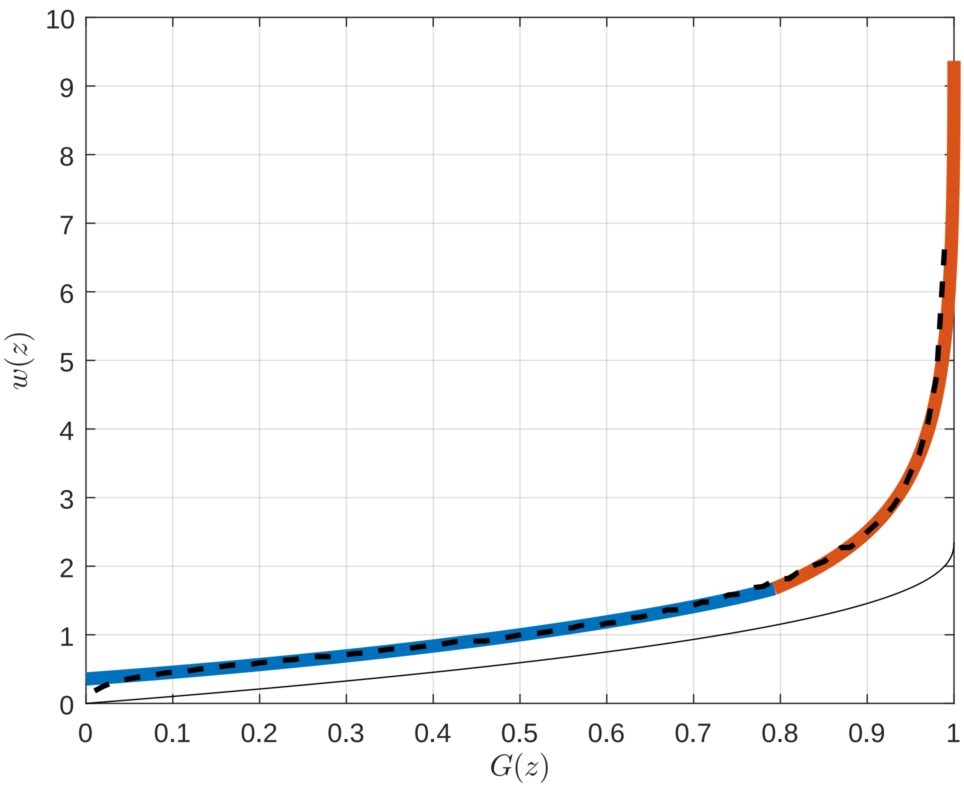


Figure 4: Distribution of wages, benchmark economy. Median normalized to one. Blue and red segments represent production workers and managers. Dashed black line represent CPS data. Solid black line represent distribution of skills.

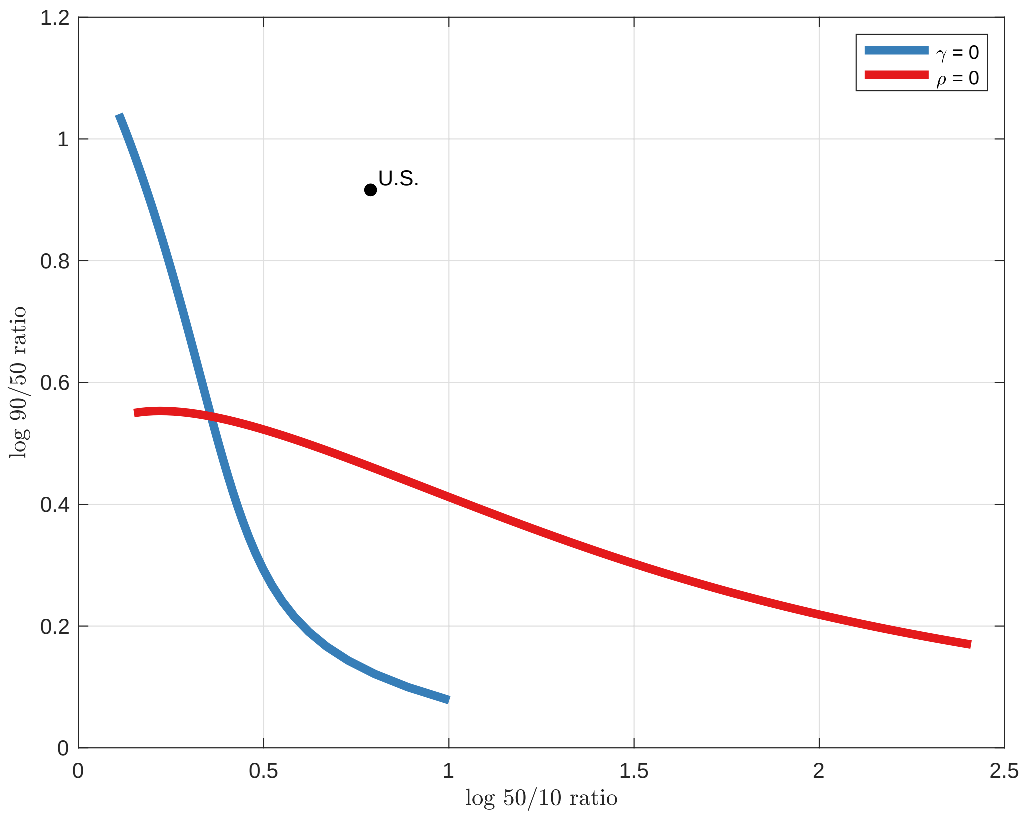


Figure 5: Feasible combinations of the log 50/10 ratio and the log 90/50 ratio conditional on the fraction of production workers being 0.793. Blue line: a model with polynomial distribution but no time cost heterogeneity ($\gamma = 0$). Red line: a model with a uniform distribution ($\rho = 0$) but time cost heterogeneity. Their intersection corresponds to $\rho = \alpha = 0$.

inequality, and the fraction of production workers as in the data. Figure 5 illustrates the negative result. Both lines in the figure are negatively sloped, meaning that a change at the top of the wage distribution is always accompanied by the opposite change at the bottom of the wage distribution. The blue line represents all the combinations of the log 50/10 ratio and the log 90/50 ratio that can be generated by a model with no time cost heterogeneity (i.e. in a model with $\gamma = 0$). In producing the blue line, we vary the density parameter ρ but use the value of h that simultaneously keeps the fraction of production workers to be 0.793. This reduces the two-dimensional parameter space to a one-dimensional one. Clearly, a model with no heterogeneity in the time cost can produce the required log 50/10 ratio only at the expense of counterfactually reducing the log 90/50 ratio to very low levels. This scenario requires a density that is significantly increasing in z ($\rho \approx -0.7$), and so produces relatively few workers of low ability. Alternatively, the model with no time cost heterogeneity can produce realistic values for the 90/50 low wage ratio, at the expense of too little inequality at the bottom.

A model with a uniform density of skills cannot match both inequality targets either. Figure 5 shows that the time cost parameter γ is key in determining the wage inequality at the bottom of the distribution. This is not surprising, given that higher γ makes low skill workers more costly to their employers relative to high skilled workers. However, the model now generates too little wage inequality at the top, regardless of the value of γ.

8 Tax Reforms

Consider now a change in the tax progressivity τ. Start with a reform that counterfactually assumes that the distribution of wages is exogenous, and does not respond to changes in the tax system. The red line in Figure 6 shows the resulting welfare as a function of the progressivity wedge τ. The (incorrectly measured) welfare is maximized at a very high rate of progressivity $\tau = 0.429$. That is, if the wage distribution is exogenous, it is

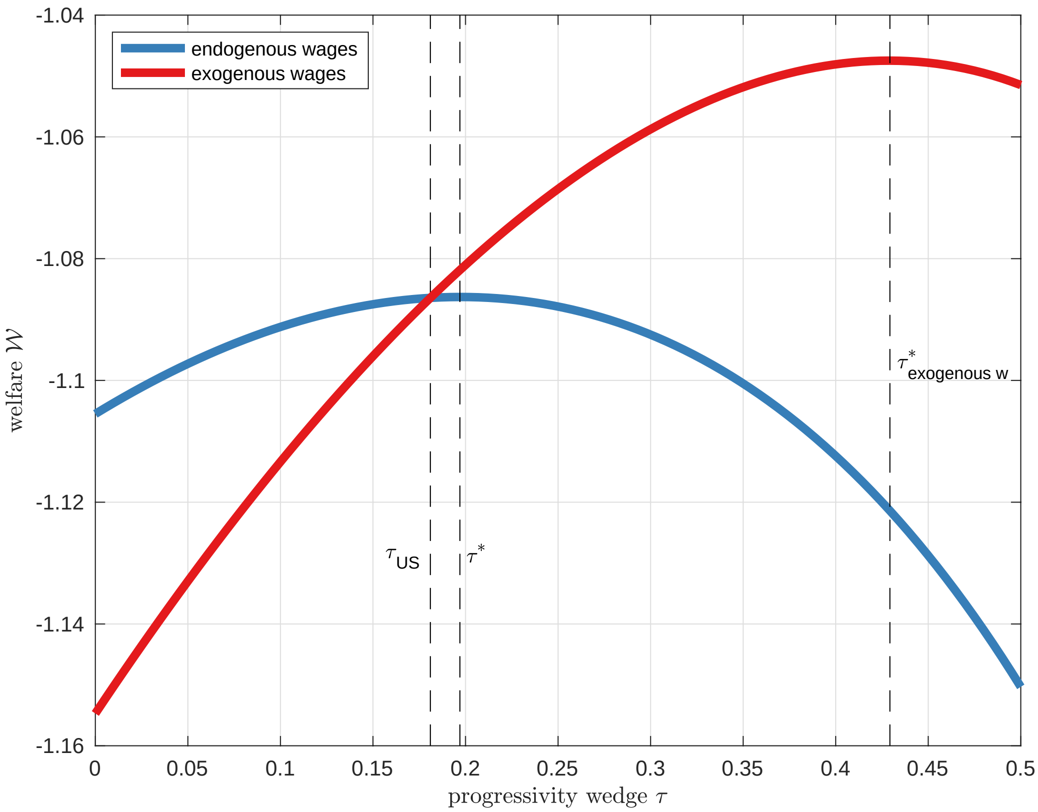


Figure 6: Welfare as a function of the progressivity wedge τ, with endogenous wage distribution (blue line) and exogenous wage distribution (red line).

optimal to significantly increase progressivity of the tax system. The blue line in Figure 6, on the other hand, shows the welfare (24) when the distribution of wages is endogenous (both lines cross at the benchmark level of $\tau = 0.181$). The optimal level of progressivity is now 0.197, only slightly higher than the benchmark value of τ, and significantly below the optimal progressivity when the wage structure is taken as exogenous. The normative predictions are thus completely different.

The reason why the optimal progressivity is lower than with exogenous wages lies in the fact that wages adjust to changes in progressivity. In particular, lower progressivity increases hours worked, which gives managers more time to supervise production workers (increase in hours worked is equivalent to a decrease in the communication time h). This increases inequality at the very top, for example as measured by the 99-90 ratio, but decreases inequality at the bottom and in the middle, as can be seen from Figure 7. The

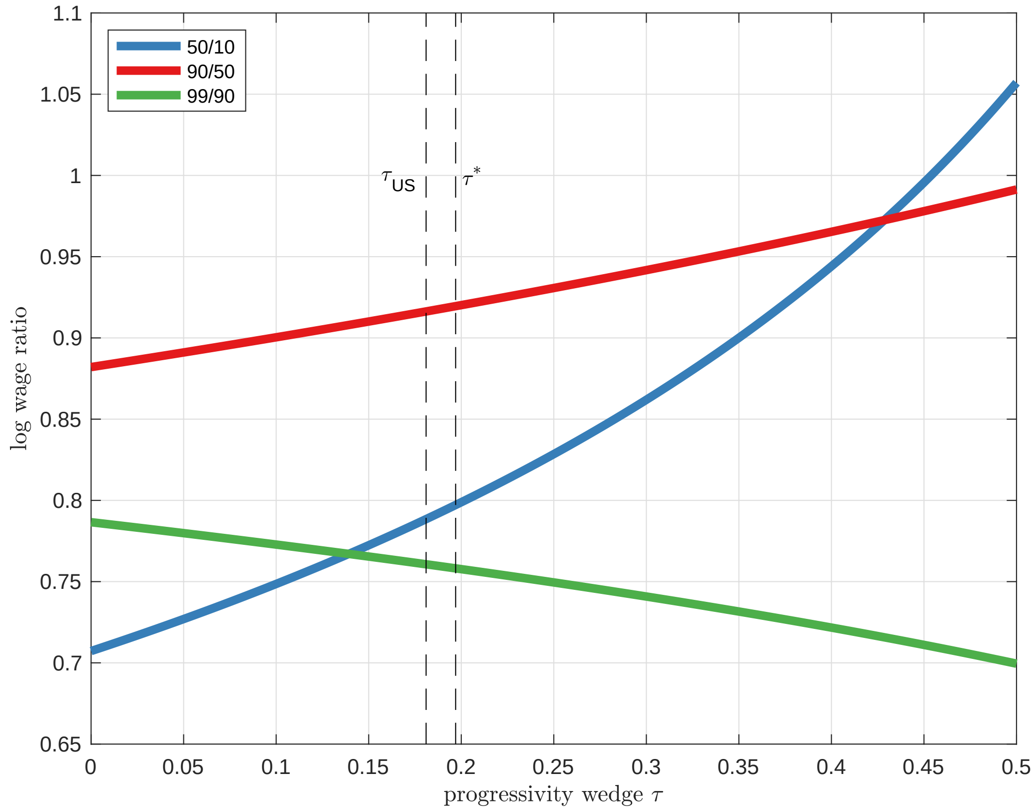


Figure 7: Log-wage ratios as a function of the progressivity wedge τ.

decrease of inequality at the bottom is in particular reflected in higher wages for the least able agents. An increase in tax progressivity is thus endogenously counteracted by a less unequal distribution of wages at the bottom. Table 2 shows that the log of 50/10 ratio increases, from 0.787 to 0.797. It is the bottom of the distribution, not the top of it, that is critical for the welfare in the economy. The optimal progressivity then decreases relatively to a model with exogenous wages. Note also from Figure 6 that implementing a naive optimum with $\tau = 0.423$ would lead to substantial welfare losses relative to the benchmark. The forces now work in the opposite directions, and the distribution of wages becomes more unequal at the bottom.

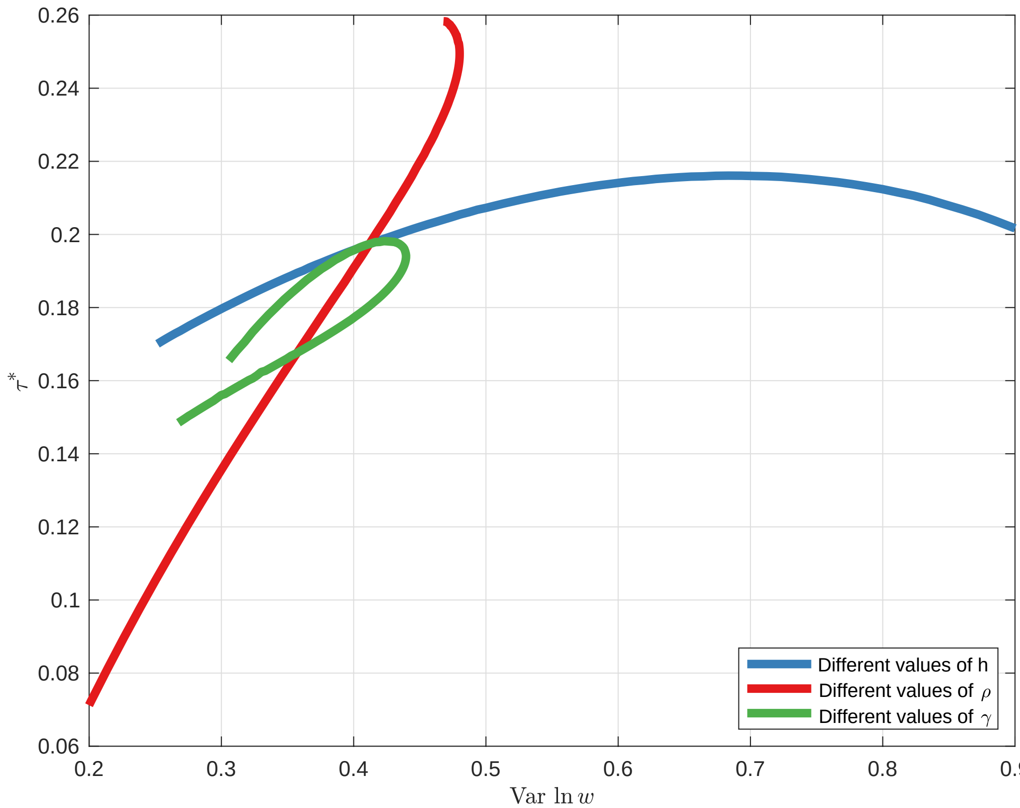
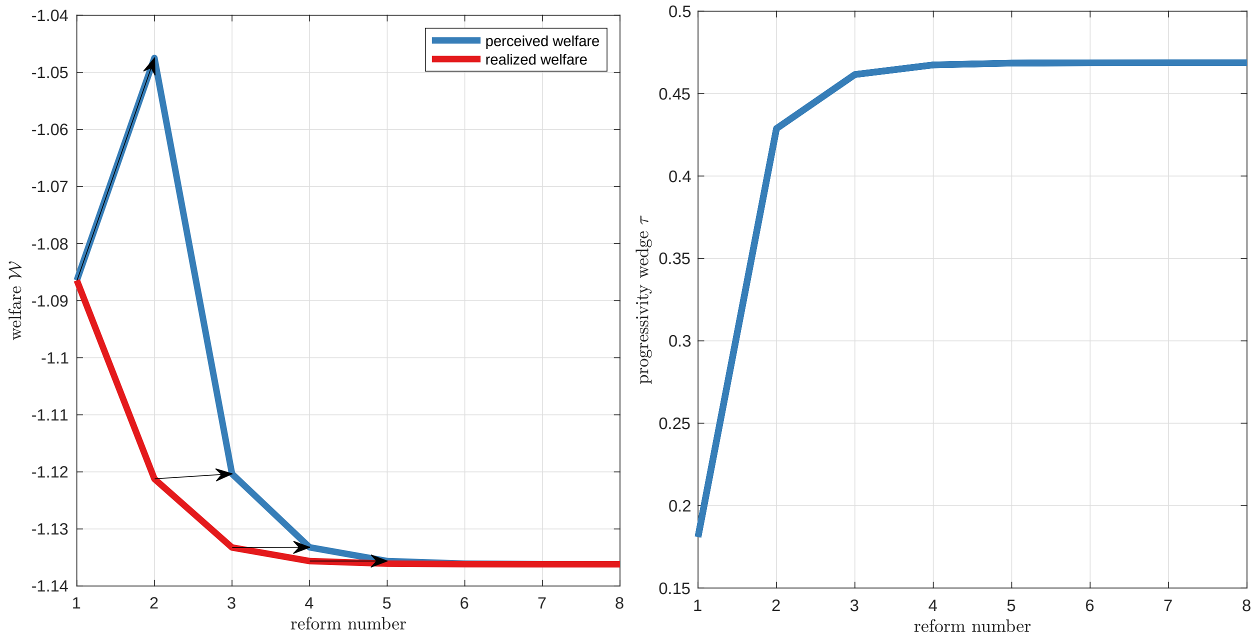


Figure 8: Optimal Tax Progressivity as a function of the variance of log wages, depending on the source of changes.

8.1 Wage Inequality and Optimal Tax Progressivity

Changes in wage inequality can occur for several reasons: the underlying distribution of skills G can change, the level time cost parameter h can change, or the heterogeneity in time cost can change. All changes might imply the same change in the dispersion of wages, as measured, for example, by the standard deviation in log wages. However, the implications for the optimal tax progressivity can be very different. Figure 8 shows the optimal progressivity parameter τ for each of the three cases, as a function of the resulting variance of log wages. It shows not only that the optimal tax response depends very strongly on what is the source of the changes, but also that there could be more than one value of the optimal progressivity at a given variance of log wages.



(a) Perceived and realized welfare.

(b) Progressivity wedge chosen.

Figure 9: Solution to the naive government's problem, first 8 rounds of reform. Arrows show perceived welfare gains.

8.2 A Naive Government

We now consider choices of a naive government, who believes that the wage structure is exogenous, and maximizes welfare under that perception. Obviously, that perception is wrong, and the naive government will realize it once the reform is implemented. We thus consider a sequence of reforms, where the planner, realizing that the wage distribution is not what he expected, implements additional reform, again under the assumption that the wage distribution is exogenous. Such a sequence of reforms converges to a self-confirming equilibrium, where the naive planner takes the current wage distribution as given, but the distribution replicates itself post-reform. Figure 9 shows such a sequence of reforms. In the first round of the reform, the progressivity wedge chosen is equal to the one that was chosen under the exogenous wage distribution above, i.e. 0.429. However, the wage distribution changes as a result of the tax reform: average wages decrease, wage dispersion increases, and welfare decreases drastically, see Figure 9a.

Paradoxically, an increase in the wage dispersion compels the naive government to increase the progressivity wedge even further as Figure 9b shows, and the vicious cycle is repeated. After several rounds of such ill-conceived tax reforms, the progressivity wedge converges to a self-confirming value of 0.469. The resulting welfare is substantially lower than in the original U.S. benchmark.

9 Conclusions

In this paper, we study the effects of taxation in a model with knowledge based hierarchies. In the model, agents self-select into being workers or managers based on their ability to solve tasks. Individual labor supply is endogeneous leading to important interactions between taxes and wage inequality. If taxes become more progressive, managers work less which decreases their wages but also the wages of their employees (workers). We calibrate a one-management-layer version of the model to the U.S. wage data. We find that the optimal tax schedule is only modestly more progressive than the one currently in place in the United States. We leave the task of studying optimal taxation with more flexible tax functions (including the Mirrleesian approach of placing no ad-hoc restrictions on the labor income tax schedule) as a fruitful avenue for future research.

References

Albanesi, S. (2011). Optimal taxation of entrepreneurial capital with private information. Working paper, Columbia University. 4

Ales, L., A. Bellofatto, and J. Wang (2017). Taxing atlas: Executive compensation, firm size, and their impact on optimal top income tax rates. *Review of Economic Dynamics 26*, 62–90. 5

Ales, L., M. Kurnaz, and C. Sleet (2015). Technical change, wage inequality, and taxes. *American Economic Review 105*, 3061–3101. 4

Ales, L. and C. Sleet (2016). Taxing top CEO. *American Economic Review 106*, 3331–3306. 4

Benabou, R. (2002, March). Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency? *Econometrica 70*(2), 481–517. 3, 10

Boháček, R. and J. Zubrický (2012). A flat tax reform in an economy with occupational choice and financial frictions. *The Economic Journal 122*, 1313–1345. 4

Bruggemann, B. (forthcoming). Higher taxes at the top: The role of entrepreneurs. *American Economic Journal: Macroeconomics*. 4

Cagetti, M. and M.-C. De Nardi (2006). Entrepreneurship, frictions and wealth. *Journal of Political Economy 114*, 835–870. 4

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy 108*, 874–904. 2, 6, 8, 9, 11, 28

Garicano, L. and E. Rossi-Hansberg (2006). Organization and inequality in a knowledge economy. *Quarterly Journal of Economics 121*, 383–435. 2, 6, 8, 9, 28

Garicano, L. and E. Rossi-Hansberg (2011). Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics 7*, 1–30. 2

Geerolf, F. (2017). A theory of pareto distributions. Technical report, UCLA. 6, 23

Guner, N., R. Kaygusuz, and G. Ventura (2014, October). Income Taxation of U.S. Households: Facts and Parametric Estimates. *Review of Economic Dynamics 17*(4), 559–581. 29

Heathcote, J., K. Storesletten, and G. Violante (2014). Consumption and labor supply with partial insurance: An analytical framework. *American Economic Review 104*(7), 2075–2126. 10

Heathcote, J., K. Storesletten, and G. L. Violante (2017, 06). Optimal Tax Progressivity: An Analytical Framework*. *The Quarterly Journal of Economics 132*(4), 1693–1754. 2, 3, 10, 29

Kapička, M. (2020). Quantifying the welfare gains from history dependent income taxation. Working paper, CERGE-EI. 3, 10

Lopez, J. J. and J. Torres (2020, October). Size-dependent policies, talent misallocation, and the return to skill. *Review of Economic Dynamics 38*, 59–93. 5

Meh, C. (2005). Entrepreneurship, wealth inequality, and taxation. *Review of Economic Dynamics 8*, 688–719. 4

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies 38*(2), 175–208. 2

Quadrini, V. (2000). Entrepreneurship, saving and social mobility. *Review of Economic Dynamics 3*, 1–40. 4

Rosen, S. (1982). Authority, control, and the distribution of earnings. *Bell Journal of Economics 13*(2), 311–323. 5

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the roy model. *Quarterly Journal of Economics 128*, 623–668. 4

Saez, E., S. Stancheva, and T. Piketty (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *Journal of Public Economics 6*, 230–271. 4

Scheuer, F. (2014). Enterpreneurial taxation with endogenous entry. *American Economic Journal: Economic Policy 6*, 126–163. 4

Scheuer, F. and I. Werning (2017). The Taxation of Superstars. *The Quarterly Journal of Economics 132*(1), 211–270. 5

Shourideh, A. (2012). Optimal taxation of wealthy individuals. Working paper, Carnegie Mellon University. 4

Slavík, C. and H. Yazici (2014). Machines, buildings, and optimal dynamic taxes. *Journal of Monetary Economics 66*, 47 – 61. 4