Organization of Knowledge and Taxation

Marek Kapička and Ctirad Slavík[*]

August 19, 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 substantially less progressive than the current one in the United States. However, if wages were exogeneous, the optimal tax progressivity would be much higher.

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 Ivo Bakota, Pawel Doligalski, Asier Mariscal, John Moore, B. Ravikumar, Florian Scheuer and Hakki Yazici, as well as audiences at various seminars, workshops and conferences for useful comments. We also thank Faisal Sohail for help with the CPS data, Marta Cota, Ante Sterc and Dusan Stojanovic for research assistance, and Andrea Downing for language editing. The authors acknowledge support from the Czech Science Foundation, grant 20-28979S.

1 Introduction

The 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 increased significantly, while the lower tail wage inequality (50/10 percentile wage ratio) has decreased, see for example Autor et al. (2006), Acemoglu and Autor (2011), Autor and Dorn (2013) or Garicano and Rossi-Hansberg (2015). 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 (Garicano (2000), Garicano and Rossi-Hansberg (2006)).

We augment the theory by endogenizing hours worked by the agents. Endogenous hours worked provide a key link between taxes and the wage structure of the economy: a change in taxes changes hours worked, which in turn changes the organization of production in the economy and the equilibrium wage structure, in addition to the standard effects on earnings. We assume that the income-tax function that the government uses has the constant-rate-of progressivity form as in Benabou (2000) and Benabou (2002) and solve for the optimal Ramsey taxes given this tax function. We find that the optimal tax progressivity is substantially lower than in the United States.

In our model, a change in taxes can simultaneously change the upper tail and the lower tail wage inequality in the opposite direction for the following reason. The efficient organization of production requires time to coordinate and communicate knowledge between managers and production workers. If managers decrease leisure due to a decrease in taxes, time available for communication between managers and workers increases. More time for communication makes more able managers relatively more useful in solving tasks, which increases wages of more able managers relative to less able managers (an increase in upper tail wage inequality). At the same time, more time for communication allows managers 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 show that this logic applies in our quantitative model. For a simplified version of the model, we provide comparative statics in closed form to illustrate the economic mechanisms in more detail. The equilibrium wage structure responds to a decrease in tax progressivity in two related ways: by changing the level of wages (the absolute effect), and the slope of the wage schedule (the relative effect). The relative effect increases the slope of the wage schedule for managers, and decreases it for production workers. This corresponds to an increase in the upper tail inequality, and a decrease in the lower tail inequality. The absolute effect is more complex: while the level of wages decreases for all production workers and higher skilled managers, lower skilled managers may see their wages increase.

We calibrate a quantitative version of the model to the U.S. wage data. The model matches the moments of the wage distribution well. It is also consistent with the empirical evidence on the elasticity of wage changes with respect to changes in marginal and average tax rates, both qualitatively and quantitatively, as we discuss in detail below. We then solve for the optimal tax progressivity using the calibrated version of the model and show that it is 0.108, which is substantially lower than 0.186, the value estimated for the United States by Heathcote et al. (2020). For comparison, we also calculate optimal wage progressivity assuming that wages are exogeneous. In this case, optimal tax progressivity is much higher at 0.341. Lower tax progressivity increases hours worked by everyone (analogously to a decrease in the communication costs), which increases wage of everyone, in particular of the poor. This leads to a decrease in bottom wage inequality. Ignoring endogenous wage changes would imply that the planner would choose excessively high progressivity, which in turn would lead to 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. Section 4 characterizes the equilibrium. Section 5 characterizes analytically the optimal tax progressivity. Section 6 discusses the calibration of a one-management-layer version of the model and presents the main quantitative results. Section 7 concludes.

2 Related Literature

Our setup generalizes Garicano (2000) and Garicano and Rossi-Hansberg (2006) in that we endogenize hours of work. Endogeneous labor supply choice implies a role for (progressive) income taxation since with inelastic labor supply income taxation would be non-distortionary. To match the key moments of the U.S. wage distribution, we also allow for the communication cost per task itself to depend on the skills of the subordinate.

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 these approaches can explain the changes observed 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.

There are several strands of literature that do consider the interaction between taxes and wages. 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 Brüggemann (2021) follow Quadrini (2000) and Cagetti and De Nardi (2006), and consider tax reforms in Bewley-Aiyagari economies with entrepreneurial activity. Taxes affect workers' wages firstly through changes in capital accumulation and, secondly through endogenous occupational choice. These papers 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 by top earners (CEOs) positively affects the productivity and profits of the firm. However, workers are not explicitly modeled either, and therefore there is no direct channel through 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- 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 a Mirrleesian optimal tax analysis. They show that the usual Mirrleesian tax formulas apply even with 'superstar' effects, but the convexity of the wage function implies higher elasticities at the top, leading to lower optimal top marginal rates. In their basic environment, matching between agents and firms is one-on-one, but they show that their main theoretical results carry over to an assignment model with fixed communication costs, which is similar to a special case of our model. We find that in such a model, it is impossible to match the bottom and top income inequality. Therefore, we focus on a richer model with a general communication cost function.

Lopez and Torres (2020) considers a framework very similar to ours, but with inelastic labor supply. Since labor income taxation is non-distortionary in their framework (as in Garicano (2000), Garicano and Rossi-Hansberg (2006)), they focus on the role of size-dependent policies in Mexico. In the same spirit, Tamkoc (2020) studies the implications of size-dependent distortions in an international context using a model similar to ours and Garicano et al. (2016) analyze the equilibrium and welfare effects of French size-contingent policies through the lens of a simple span-of-control model of Lucas (1978). Finally, Lawson (2019) studies efficient income taxation in hierarchical models with bargaining.

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 $\kappa > 0$.

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 $g(z)$. The agents receive a continuum of tasks 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 can thus produce $\ell F(z)$. If $\underline{z} > 0$ then there is a mass of problems $F(\underline{z}) > 0$ that every agent can solve.

Rather than 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. Thus, the agents 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.[1] 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 and upper bounds 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 $(z_I, z_{I+1}]$. The workers and the top layer managers are special: the workers because they are the only ones who produce output, and the top managers because they are residual claimants.

We denote the knowledge of the production worker by x_0 and the knowledge of the manager in layer i by x_i. After receiving advice, workers produce output. The production of the team is

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

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. Note that it is the skill of the layer I manager x_I that ultimately determines the mass of tasks that the workers solve.[2]

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

[1]We do not allow for self employment. A model with self employment shares many features with our model.

[2]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.

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 strictly positive, differentiable and decreasing in x_{i-1}.*

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 follows. Since the subordinate managers solve a fraction $F(x_{i-1})$ of problems, they pass the remaining fraction of problems $1 - F(x_{i-1})$ 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 $h[1 - F(x_{i-1})]$.

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 the 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.*

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 who employs a production 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 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 thus his profits, increase linearly with hours worked by 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 working. Equation (3) shows a second key complementarity in the model: there is *working time complementarity* between the hours worked by a worker, and those by 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

Assignment. We start the description of the equilibrium conditions by characterizing the assignment of workers to managers. There is positive assortative matching, where the worst production worker is matched with the worst managers, and the best production worker is matched with the best managers.[3] Let $m(x_i)$ for $x_i \in [z_{i-1}, z_i]$ be the knowledge of the manager at level $i+1$ who 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 at each level:

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

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 $n(x_i)$ 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

[3]We do not justify here that the equilibrium assignment takes this form. The reader is referred to Garicano and Rossi-Hansberg (2006), Proposition 1, for such a justification.

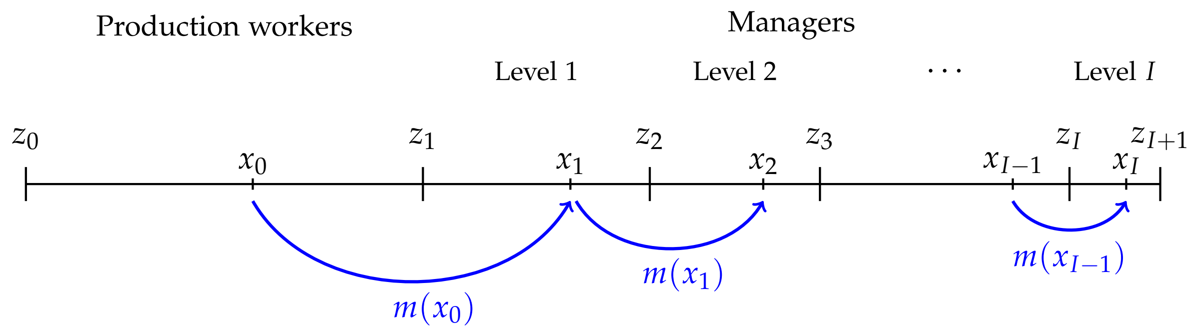


Figure 1: The equilibrium assignment $m(x)$ of subordinates to their immediate superiors. The subordinates are either workers or managers.

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 chooses 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 rate of the top manager is given by (3), and so the problem to maximize the top manager's wage rate is

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

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 $i = 1, \ldots, I-1$ 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. Aggregate output in the economy consists of workers' production (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 continuous at z, and the government budget constraint holds.*

Before proceeding further and characterizing the equilibrium, we will show that the model can be simplified substantially: 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 $\tilde{l}$ satisfy (6) given $\tilde{\theta}$ and $\tilde{F}$, rewrite the right-hand side of (6)

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

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. Therefore, 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 always be 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 , these 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.[4]

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 for 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 of 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. In what follows, we normalize $\bar{\ell}(\tau) = 1$.

Solving the assignment problem. It turns out that the equilibrium assignment can be easier to characterize by using matching functions that map the worker's knowledge

[4]From a practical perspective, the normalization is perhaps less important. In our quantitative analysis of Section 6 we find it convenient to normalize F to be uniform on $[0,1]$ and calibrate a flexible distribution of skills G.

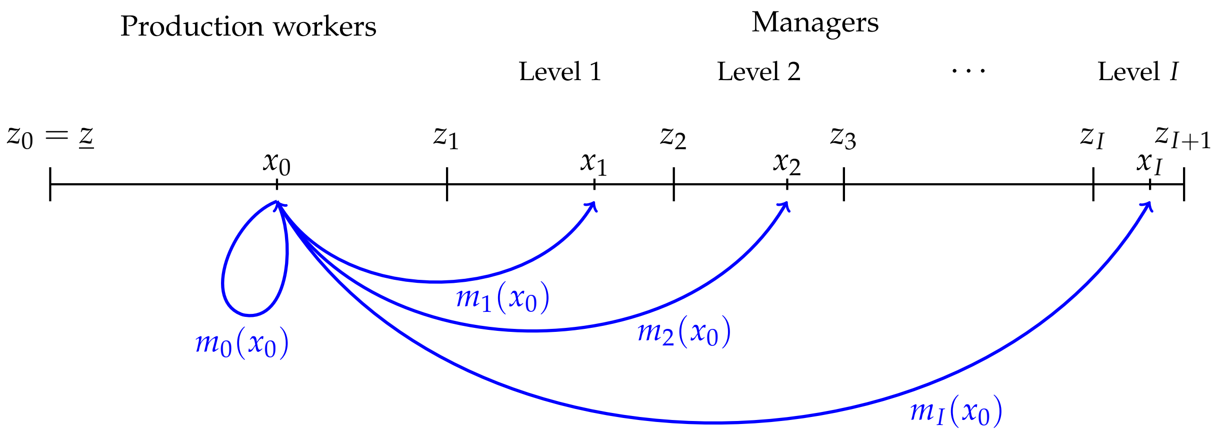


Figure 2: The equilibrium assignment $m_i(x_0)$ of workers to their superiors in level i.

$x \in [\underline{z}, z_1]$ directly to the knowledge of the manager in each layer. To that end, define a function $m_i(x)$ for $i = 0, \ldots, I$ recursively by $m_0(x) = x$ and $m_{i+1}(x) = m(m_i(x))$. The function $m_i(x)$ represents the knowledge of a manager in layer i that is matched with a worker with knowledge x, as figure 2 illustrates. The equilibrium assignment (4) gives

$$m_i(\underline{z}) = z_i, \quad m_i(z_1) = z_{i+1}. \tag{11}$$

Differentiating the market clearing condition (5) with respect to x, multiplying both sides by $g(m(x_i))$, integrating back and using the equilibrium conditions (11), we write the equilibrium assignment function as

$$m_i(x) = z_i + \rho_{i-1}(x), \quad x \in [\underline{z}, z_1], \, i = 1, \ldots, I, \tag{12}$$

where the function ρ_i is given by

$$\rho_i(x) = \int_{\underline{z}}^{x} \theta\left(m_i(t)\right) \, dt, \quad x \in [\underline{z}, z_1], \, i = 0, \ldots, I-1.$$

We obtain immediately from (12) that the functions m_i are differentiable and strictly 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 functions m_i are strictly increasing, differentiable and concave for all $i = 1, \ldots, I$.*

Proof. Differentiating (12) with respect to x, we obtain that m_i is differentiable, with a derivative $m_i'(x) = \theta(m_{i-1}(x))$, which is strictly positive. 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 behind 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 functions.[5]

Evaluating (12) at the thresholds z and using (4) yields a unique equilibrium condition for the threshold values:

$$z_{i+1} = z_i + \rho_{i-1}(z_1) \quad i = 1, \ldots, I, \tag{13}$$

where we take $z_{I+1} = \bar{z}$ in the last equation. Summing over, the equilibrium value of z_1 is a solution to the following equation:

$$1 - z_1 = \rho(z_1), \tag{14}$$

where $\rho(z_1) = \sum_{i=1}^{I} \rho_{i-1}(z_1)$ is only a function of z_1. Figure 3 illustrates how z_1 is determined. Since $\rho_{i-1}'(z) = \theta(m_i(z)) > 0$, $\rho(z)$ is strictly increasing and concave in z. Since the left-hand side of (14) starts at one, is strictly decreasing in z and ends at

[5]Note that Lemma 1 does not claim that the original equilibrium mapping $m(x)$ is increasing, differentiable and concave. It can easily be shown that $m(x)$ is continuous and differentiable except for differentiability at z_1. Concavity is, however, not guaranteed for $x \geq z_1$.

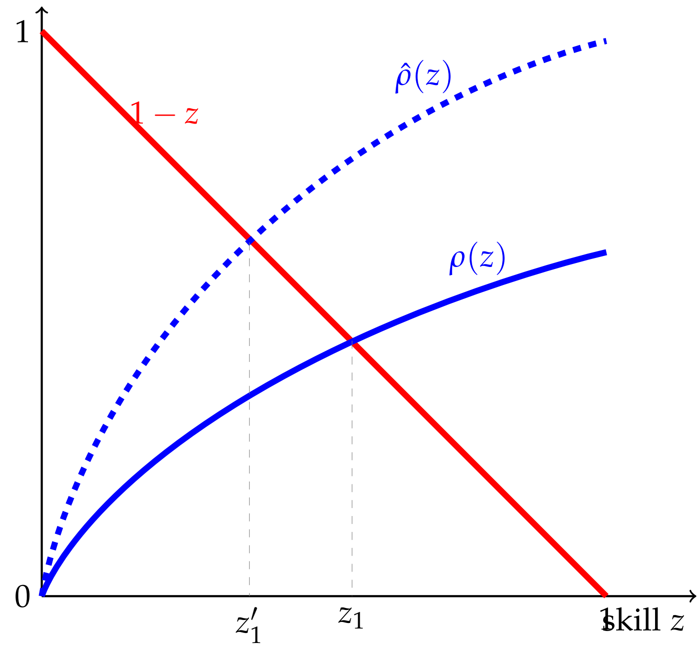


Figure 3: The equilibrium value of z_1. An increase in the time cost from θ to $\theta' > \theta$ increases ρ to $\hat{\rho} > \rho$ and decreases the equilibrium z_1.

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 condition (14). Moreover, when the time cost function increases from θ to $\hat{\theta} > \theta$,[6] the threshold knowledge z_1, and hence the fraction of workers, decreases unambiguously. Figure 3 illustrates the comparative statics. To summarize,

Proposition 2. *There is a unique threshold knowledge $z_1 \in [0, 1]$ that solves (14). Moreover, z_1 is decreasing in θ.*

Wage schedule. A second key equilibrium relationship obtained from the managers' problem of choosing the type of his subordinates. Solving the managers' problem (6)

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

yields a differential equation in production workers' wages

$$w'(x_0) = -\theta'(x_0)w(x_1) \tag{15a}$$

$$w'(x_i) = -\frac{\theta'(x_i)}{\theta(x_{i-1})}w(x_{i+1}), \quad i = 1, \ldots, I-1 \tag{15b}$$

$$w'(x_I) = \frac{f(x_I)}{\theta(x_{I-1})}. \tag{15c}$$

The intuition behind equation (15a) 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 his superior's time. 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. (15a) holds. The intuition behind (15b) is similar.

Lemma 2. *The wage function $w(x)$ is strictly increasing, differentiable for all x, except possibly at z_1 and z_I. If $\theta'' < 0$ and $f' \geq 0$ then $w(x)$ is convex for all x, except possibly at z_I.*

Proof. Differentiability inside each segment follows from (15) and Assumption 1. Evaluating the left-hand and right-hand derivatives at the thresholds gives

$$w'_-(z_1) = -\theta'(z_1)w(z_2) < \frac{-\theta'(z_1)}{\theta(z_0)}w(z_2) = w'_+(z_1)$$

$$w'_-(z_i) = w'_+(z_i) \quad i = 2, \ldots, I-1.$$

Hence the wage function is differentiable everywhere except at z_1 and z_I (for which the difference between left-hand and right-hand derivatives cannot be signed). Since it is continuous and the derivatives are all strictly positive, $w(x)$ is strictly increasing.

Differentiating (15a) and (15b) for $i = 1, \ldots, I-1$ again yields

$$w''(x_0) = -\theta''(x_0)w(x_1) - \theta'(x_0)w'(x_1)m_1'(x_0) > 0$$
$$w''(x_i) = \frac{-\theta''(x_1)w(x_{i+1})\theta(x_{i-1}) - \theta'(x_i)w'(x_{i+1})m_{i+1}'(x_0) + \frac{\theta'(x_{i-1})}{m_i'(x_0)}\theta'(x_i)w(x_{i+1})}{\theta(x_{i-1})^2} > 0$$
$$w''(x_i) = f'(x_I) - \frac{f(x_I)\theta'(x_{I-1})\theta(x_{I-2})}{\theta(x_{I-1})^2} > 0,$$

where the inequalities follow from $\theta' < 0$, $\theta'' < 0$ and $f' \geq 0$. This establishes convexity everywhere except for z_1 and z_I. But the wage function is convex at z_1 as well, since the left-hand derivative is smaller than the right-hand derivative.

■

Lemma 2 shows that the wage function is, in general, convex, except possibly at the threshold between the top managerial and the subordinate levels z_I. The reason for the potential nonconvexity comes from the fact that the top managers' wages are determined as residuals. If the wage schedule in the subordinate levels becomes steeper, the slope at the top level decreases. If the movement of slopes in the opposite direction is strong enough, nonconvexity at z_I follows. Section 4.1 establishes conditions under which nonconvexity happens in a special case.

Note also that, in general, convexity tends to 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 (15), 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 then w_{I-1} will tend to be quadratic in x.

4.1 An Example

In this section, we characterize the equilibrium in a special case of our model in which the distribution of wages and other relevant objects are easy to characterize analytically. We derive comparative statics results in the tax progressivity parameter τ, which are shown to hold in the richer version of the model (which we bring to the data in Section 6.1).

The model is assumed to have only one managerial layer ($I = 1$). The productivity shocks are drawn from a unit interval, $\underline{z} = 0$, $\overline{z} = 1$, and the underlying distributions are uniform with $F(z) = G(z) = z$. The normalized communication cost function is $\theta(x) = h(1-x)$. Hence, we interpret h as $h/\bar{\ell}(\tau)$ and assume that $h < 1$. An increase in h can then arise from an increase in the communication cost h or an increase in the tax progressivity parameter τ.[7] The assumptions made here imply that Assumption 1 is satisfied.

The problem has a closed form solution. The function ρ_0 is given by $\rho(z) = h[1 - (1-z)^2]/2$. Equation (14) is a quadratic equation in $1 - z_1$, and z_1 solves

$$z_1 = 1 - \frac{\sqrt{1+h^2} - 1}{h}. \tag{16}$$

The threshold value z_1 is clearly decreasing in h, confirming the results of Proposition 2. An increase in the effective communication cost h thus increases the fraction of managers in the economy. The reason is that it is more costly to supervise the production workers. The matching function is quadratic and concave in the workers' skill x:

$$m_1(x) = z_1 + hx - \frac{h}{2}x^2. \tag{17}$$

Since $m_1'(x) = h(1-x)$, an increase in h makes the matching function steeper. 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 gain a larger advantage by being matched with a comparatively better manager.

[7] This configuration delivers a distribution of wages with a Pareto right tail, see Geerolf (2017).

Wages of the production workers are increasing and quadratic in their skill, and wages of the managers are increasing and convex in their skill:

$$w(x) = z_1 + A(x-1) + \frac{h}{2}x^2, \quad 0 \le x < z_1, \tag{18}$$

$$= \frac{A}{h} + 1 - \sqrt{1 - 2\frac{x - z_1}{h}}, \quad z_1 \le x \le 1, \tag{19}$$

where $A = [1 + (1+h)^2]/\sqrt{1+h^2} - 2 - h$. It is easy to verify that A is positive. One can also show that A is increasing in h and that A/h is decreasing in h.

What happens to the wage schedule if h changes? Individuals at different parts of the skill distribution are affected differently and we characterize the response both in terms of the level (the *absolute effect*) and slope of the wage function (the *relative effect*). In what follows, we explicitly index the wage function w by h and denote the partial derivative of the wage function with respect to skill x by $w_x(x,h) = \partial w(x,h)/\partial x$. Starting with the relative effect, we obtain:

Lemma 3. ***Relative Effect.*** *$\partial w_x(x,h)/\partial h > 0$ for all $x \in [0, z_1)$ and $\partial w_x(x,h)/\partial h < 0$ for all $x \in [z_1, 1]$.*

Proof. Differentiating the marginal wage schedule of the production workers with skill $x \in [0, z_1)$ with respect to h yields $\partial w_x(x,h)/\partial h = A'(h) + x$, which is strictly positive since $A' > 0$. The marginal wage of the managers with skill $x \in [z_1, 1]$ is $w_x(x,h) = h^{-1}[1 - 2\frac{x-z_1}{h}]^{-1/2}$. Differentiating with respect to h, we obtain

$$\frac{\partial w_x(x,h)}{\partial h} = B'(h) - \left(1 - 2\frac{x - z_1}{h}\right)^{-\frac{3}{2}} \frac{1 - \frac{h}{\sqrt{1+h^2}} + \frac{1-x}{h}}{h^2} < 0,$$

where $B(h) = A(h)/h$, and we show that $B'(h) < 0$. Since $1 - x > 0$, the numerator of the last term is positive and the expression must be negative. ■

Lemma 3 shows that the increase in the effective communication cost unambiguously increases the dispersion of wages of the production workers and decreases the dispersion of managers' wages. This happens because the number of managers increases and, as a result, better workers are matched with relatively more productive managers. Part of the relative efficiency gains is translated into their wages. The managers' wage schedule becomes flatter for the same reason the workers' wage schedule becomes steeper. Two managers of given skills are now matched with more similar workers, and their productivity differences decrease. The results also imply that the difference in wages between any two production workers increases, while the difference between any two managers decreases. [8]

The absolute effect is less straightforward because, under certain conditions, wages may move in either direction for lower skilled managers:

Lemma 4. ***Absolute Effect.*** *There is a value of $\bar{h}$ such that if $h > \bar{h}$ then $\partial w(x,h)/\partial h < 0$ for all $x \in [0,1]$. If $h < \bar{h}$ then $\partial w(x,h)/\partial h < 0$ for $x \in [0, z_1) \cup (\tilde{z}, 1]$ and $\partial w(x,h)/\partial h > 0$ for $x \in [z_1, \tilde{z})$ and $\partial w(\tilde{z},h)/\partial h = 0$ for some $\tilde{z} < 1$.*

Proof. The wage schedule for the production workers with skill $x \in [0, z_1)$ changes according to

$$\begin{aligned}\frac{\partial w(x,h)}{\partial h} &= z_1'(h) + A'(h)(x-1) + \frac{1}{2}x^2 \\ &= z_1'(h) + A'(h)(z_1 - 1) + \frac{1}{2}z_1^2 + A'(h)(x - z_1) + \frac{1}{2}x^2 - \frac{1}{2}z_1^2 \\ &= -\left(1 - \frac{1}{\sqrt{1+h^2}}\right)\frac{2-h}{h+h^3} + A'(h)(x-z_1) + \frac{1}{2}x^2 - \frac{1}{2}z_1^2 < 0.\end{aligned}$$

For the managers with skill $x \geq z_1$, we obtain

$$\frac{\partial w(x,h)}{\partial h} = B'(h) - \frac{1}{\sqrt{1 - 2\frac{x-z_1}{h}}}\frac{x - z_1 + hz_1'}{h^2}.$$

It has already been shown in Lemma 3 that $\partial w(x,h)/\partial h$ is decreasing in x. The wage schedule of all managers will thus decrease if the change at the threshold $\partial w(z_1,h)/\partial h = B'(h) - z_1'(h)/h$

[8] The fact that the slopes of the wage schedule move in the opposite direction in both segments indicates that the overall wage schedule will in general have a kink at z_1, as already shown in Lemma 2. For low values of h, the managers' schedule will be steeper than the workers' schedule at z_1. As h increases, the left and right derivatives at z_1 move in the opposite direction and reduce the kink. For sufficiently large h, in particular for $h > 0.918$, the slope of the production workers' wage schedule exceeds the slope of the managers' schedule. The overall wage schedule then exhibits nonconvexity at z_1.

is negative. After some algebra, we show that this happens if $h > \bar{h}$. If $h < \bar{h}$, $\partial w(z_1, h)/\partial h > 0$ and, since $\partial w(1,h)/\partial h = B'(h) + z_1'(h) < 0$, by continuity there exists a value of $\tilde{z} \in (z_1, 1)$ such that $\partial w(x,h)/\partial h > 0$ for $x \in [z_1, \tilde{z})$, and $\partial w(x,h)/\partial h < 0$ for $x \in (\tilde{z}, 1]$. ■

An increase in the effective communication costs thus unambiguously decreases wages for all production workers. The reason is that an increase in h increases the cost of creating teams. Higher costs imply that each production worker is being matched with a less productive manager, which shifts the worker's wage down. Wages also decrease for high skilled managers (those with skills greater than $\bar{x}$). Interestingly, lower skilled managers may see their wages increase. They, too, suffer from a higher cost of creating teams, but are matched with slightly better workers. The second effect dominates if h is low enough and the cost of creating teams is not as important.[9] Both cases are illustrated in Figure 4.

While it is not necessarily true that an increase in the effective communication cost h decreases wages for all workers, we show that the aggregate output of the economy must decrease:

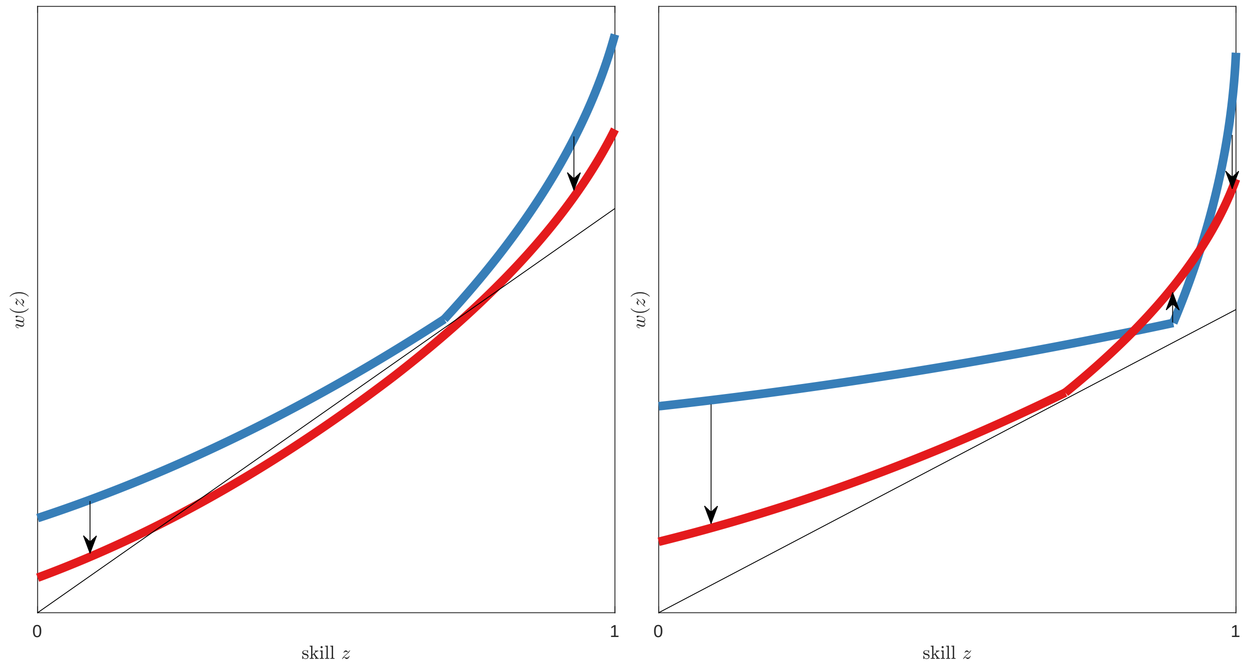
Lemma 5. *The aggregate output $Y = \int_0^{z_1} m_1(t)\, dt$ is decreasing in h.*

Proof. Y has a closed form solution $Y = [z_1 (h + 2(1 + z_1)) - 1]/3$. Differentiating with respect to h, we get $Y'(h) = [z_1 + (2 + h + 4z_1) z_1'(h)]/3$. Since $z_1 z_1' < 0$, the Lemma will be proven if $z_1 + (2 + h) z_1'(h) < 0$. To show this, write

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

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

[9] The threshold value of h is a solution to the polynomial equation $2h^3 - 2h^2 + 2h - 1 = 0$ and its value is approximately 0.648.



(a) $h \geq \bar{h}$ **(b)** $h \leq \bar{h}$.

Figure 4: A change in the wage schedule before (blue line) and after (red line) an increase in the effective communication cost h.

5 Optimal progressivity

We now characterize the optimal value of the progressivity parameter τ and its determinants. Let $\mathbb{E}[w|\tau] = Y\left[1, h/\ell(\tau)\right]$ be the average wage and rental rate, and $\mathbb{E}[w^{1-\tau}|\tau] = C\left[1, h/\ell(\tau)\right]/\lambda$ 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

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

Solving for the equilibrium λ and substituting back to the expected utility yields the aggregate welfare (where each agent's utility receives equal weight) for a given progressivity wedge τ

$$\mathcal{W} = \ln\left[\bar{\ell}(\tau)\mathbb{E}\left[w|\tau\right] - G\right] - \frac{1-\tau}{1+\eta} - \ln \mathbb{E}[w^{1-\tau}|\tau] + (1-\tau)\mathbb{E}\left[\ln w|\tau\right]. \quad (20)$$

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

To further inspect the novel role of taxes in determining the wage distribution, we approximate the penultimate term in (20) as follows:

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

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 (20) and canceling terms yields an approximate expression for welfare:

$$\mathcal{W} \approx \ln\left[\bar{\ell}(\tau)\mathbb{E}\left[w|\tau\right] - G\right] - \frac{1-\tau}{1+\eta} - \frac{(1-\tau)^2}{2}\mathbb{V}[\ln w|\tau], \tag{21}$$

where $\mathbb{V}[\ln w|\tau] = \mathbb{E}\left[(\ln w - \mathbb{E}\ln w)^2|\tau\right]$ is the variance of log wages. The expression (21) 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. Similarly, increasing the progressivity parameter τ is expected to increase the variance of *log* wages by increasing bottom wage inequality. While we cannot prove these results analytically, we show that they apply in the calibrated model in the next section. Taken together, these two forces suggest that the optimal tax progressivity will not be high, as we confirm in what follows.

6 Quantitative Analysis

This section calibrates the model to the U.S. data and solves for the optimal taxes.

6.1 Model Calibration

The model is calibrated to the U.S. economy for the 2012 - 2016 period. 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 (this assumption is without loss of generality due to Proposition 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:

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

One can interpret the communication cost as follows. Workers incur two types of costs on their managers. First, lower skilled workers 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 $\gamma = 0$ corresponds to the specification in Garicano (2000) or Garicano and Rossi-Hansberg (2006).

Clearly, Assumption 1 is satisfied; θ is differentiable and decreasing in x. Proposition 1 implies that an isomorphic way of writing the model would assume that G is uniform. This is equivalent to writing the cost function θ and the distribution of tasks F as a function of the percentiles, $\tilde{\theta}(p) = h(1-p)^{\frac{1+\gamma}{1+\rho}}$ and $\tilde{F}(p) = 1-(1-p)^{\frac{1}{1+\rho}}$. These considerations imply that the sufficent conditions for Lemma 6 ($\tilde{f}' < 0, \tilde{\theta}'' < 0$) are satisfied if $0 < \gamma < \rho$.

As noted earlier, given γ and ρ, only the ratio of $h/\bar{\ell}$ matters from the wage distribution. We calibrate ρ and γ and the 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 the CPS March supplement, where we compute the averages over the 2012-2016 period. The details are relegated to Appendix A. The fraction of managers is 18.7 percent, the 90/50 log wage ratio is 0.877, and the log 50/10 ratio is 0.743. This yields $\rho = 1.540$, $\gamma = 2.153$ and $h/\bar{\ell} = 0.429$.[10]

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 specified. One exception is κ, which is a scaling factor, irrelevant for comparative statics in τ and the optimal tax exercise due to log utility (notice that κ does not appear separately in equation (21), 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 $\tau = 0.186$ as estimated for the 2012-2016 period by Heathcote et al. (2020) (see also Guner et al. (2014)

[10]A relatively high value of γ means that a substantial amount of heterogeneity in the communication cost is needed to calibrate the model. Communicating about a given task with the worst production worker takes about four times longer than communicating about it with the best production worker. The heterogeneity in the communication costs is needed to match the wage inequality at the bottom of the distribution, as we discuss in the next subsection. A relatively high value of ρ means that the density of high skills is much smaller than that of low skills. While the worst production worker can solve none of the tasks that arrive, the best production worker can solve about 48 percent of all tasks. The decreasing density is needed to match the wage distribution at the top, as we show in Section 6.4.

Table 1: Baseline Parameters

Parameter	Value	Target	Value	Source
Calibrated parameters				
ρ	1.540	log 90/50 wage ratio	0.877	CPS
γ	2.153	log 50/10 wage ratio	0.743	CPS
$h/\bar{\ell}$	0.429	Share of managers	0.187	CPS
Parameters set outside the model				
κ	1	Disutility of labor		Normalization
η	2	Frisch elasticity		
τ	0.186	Tax progressivity		Heathcote et al. (2020)
G/Y	0.16	Gvt consumption		NIPA

This table shows the benchmark parameters. CPS stands for the March supplement of the Current Population Survey (the values are averages over 2012-2016) and NIPA stands for the National Income and Product Accounts. The details of the data constructions are contained in Appendix A.

for estimates of this and other tax functions). As reported in the National Income and Product Accounts (NIPA), the government consumption-to-output ratio has been fairly stable with an average of about 16% since the 1980's. This is the value we use for the share of government expenditure relative to GDP. This is also the income-weighted average tax in the economy. The resulting benchmark parameters are summarized in Table 1.

We solve the model by discretizing the set of workers' skills $[0, z_1]$ into 1000 equally spaced gridpoints. This effectively makes the grid endogeneous to the threshold value z_1. All the relevant functions, including the matching function, are then computed using this grid. As a result, the grid of managers' skills are not necessarily equally spaced.

6.2 Model Fit

Figure 5 plots the resulting distribution of wages in the benchmark economy (the blue segment of the solid line corresponds to the wages of the workers and the red segment to those of the managers) and compares it to the empirical distribution of wages (black dotted line). The horizontal axis depicts the population percentiles. Table 2 shows

Table 2: Moments of the Wage Distribution

	CPS Data	Model
Calibrated moments		
log 50/10 ratio	0.743	0.743
log 90/50 ratio	0.877	0.877
Uncalibrated moments		
log 25/10 ratio	0.329	0.302
log 50/25 ratio	0.415	0.441
log 75/50 ratio	0.444	0.411
log 90/75 ratio	0.433	0.466
log 95/90 ratio	0.260	0.306
log 99/95 ratio	0.643	0.521
Variance of log wages	0.430	0.378
Gini of wages	0.386	0.362

This table shows the moments of the wage distribution computed using the CPS data (averaged over 2012-2016) and in the calibrated model.

additional moments of the wage distribution in the benchmark economy relative to the CPS data. Both distributions are remarkably close, despite the fact that we are matching only the 50-10 and 90-50 log wage ratio. The model-implied wage function is convex even though the sufficient conditions for Lemma 6 are not satisfied, as $\gamma > \rho$ and $\theta'' > 0$.

As both Figure 5 and Table 2 show, the model is slightly less succesful in matching the wage distribution at the very bottom and the very top. At the bottom of the wage distribution, the model predicts larger wages than those observed in the data, while at the very top, the model predicts a somewhat thinner upper tail. For example, the log 99/95 wage inequality, not targeted by the model, is 0.643 in the data and 0.521 in the model. Overall, however, the model matches the empirical distribution of wages well.

Wage Elasticity to Tax Changes. In our model with endogeneous wages, changes in taxes imply changes in wages. To the extent that we are interested in optimal taxes, we need to make sure that the relationship between taxes and wages in the model matches that in the data. The empirical literature estimates the elasticity of wages to the (net of) marginal tax rate $1 - MTR$ as well as the elasticity of wages to the (net

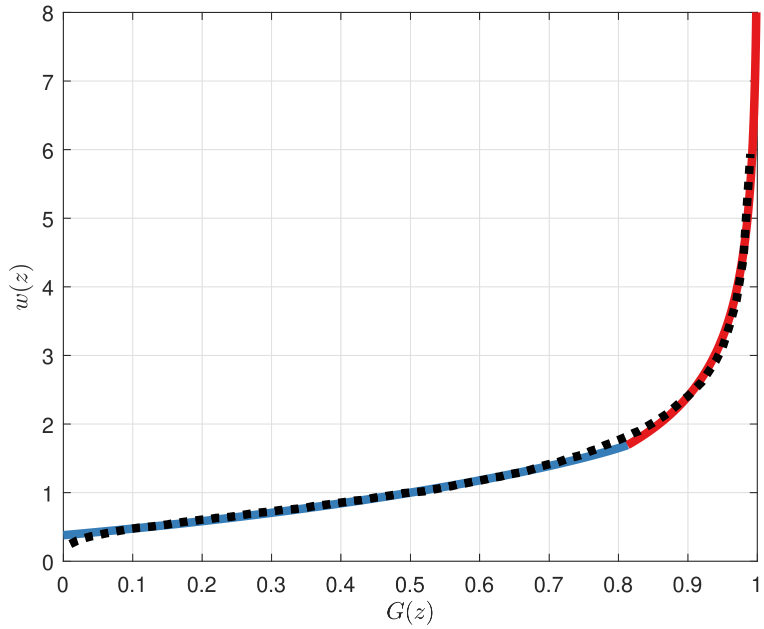


Figure 5: Distribution of wages, benchmark economy. Median normalized to one. Blue and red segments of the solid line represent production workers and managers. Dotted black line represents CPS data, averaged over 2012-2016. The horizontal axis shows the population percentiles.

of) average tax rate $1 - ATR$. The broad agreement in the empirical literature is that: $d \log w / d \log(1 - MTR) > 0$ while $d \log w / d \log(1 - ATR) < 0$, see Aronsson et al. (1997), Hansen et al. (2000), Blomquist and Selin (2010) or Holmlund and Kolm (1995). These papers use Scandinavian data. Lockwood and Manning (1993), Schneider (2005) and Frish et al. (2020) report similar findings using British, German and Israeli data, respectively. There is less work on the effects of tax changes on wages in the United States. A notable exception is Moffitt and Wilhelm (1998), who show that wages of rich males have increased in response to the 1986 U.S. (marginal) tax rate cuts, consistent with the above literature.

To compare our model to the empirical papers, we solve another version of it in which we vary the tax progressivity marginally relative to the benchmark value $\tau = 0.186$. We then run two regressions on the model-generated data before and after this tax change to estimate the two tax elasticities. We find that: (i) $d \log w / d \log(1 - MTR) = 0.21$, and,

(ii) $d \log w / d \log(1 - ATR) = -0.46$.[11] These elasticity estimates are not very sensitive to the magnitude of the tax change. These results are consistent with the literature not only qualitatively, but also quantitatively. The estimates of $d \log w / d \log(1 - MTR)$ in the literature range from 0.15 to 0.8 and the estimates of $d \log w / d \log(1 - ATR)$ tend to be between -0.2 and -0.7. [12]

Section B.1 in the Appendix discusses additional features of the calibrated model.

6.3 Optimal Taxes

This section solves for optimal taxes from the perspective of a utilitarian government that puts equal weight on each agent's utility, as in Section 5. The planner chooses tax progressivity τ and tax-level parameter λ to maximize welfare subject to being able to finance a fixed level of government expenditure G; held fixed at the level from the calibrated model. All remaining parameters, including the communication costs h are held constant. As a result of the tax change(s), however, labor supply $\bar{\ell}(\tau)$ changes and the normalized communication costs $h/\bar{\ell}(\tau)$ change as well. We can thus compare the results of this section to the theoretical results derived for a simpler version of the model in Section 4.1.

For illustration purposes, we plot aggregate welfare as a function of τ (bearing in mind that there is an underlying change in λ) as the blue line in Figure 6. The optimal level of progressivity is 0.108, which is substantially lower than in the current U.S. tax code, namely 0.186. The welfare gains of optimal taxes are non-negligible at 0.36% in consumption units. Decreasing tax-progressivity is redistributive: implementing the

[11] Holding the ATR constant when estimating the elasticity with respect to MTR (and the other way around), as sometimes done in the literature, yields the following estimates on the model-generated data: (i) $d \log w / d \log(1 - MTR) = 0.27$, and, (ii) $d \log w / d \log(1 - ATR) = -0.53$.

[12] Schneider (2005) estimates that the effect of changes in the ATR on wages is larger in absolute terms than the effect of changes in the MTR (as in our model). Aronsson et al. (1997) report the same finding for some specifications of their empirical model.

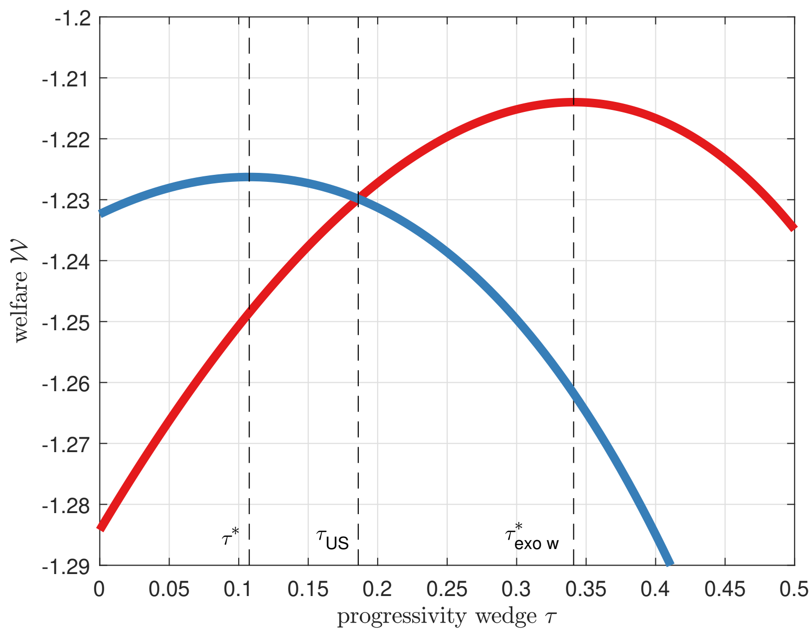


Figure 6: Welfare as a function of the progressivity wedge τ, with endogenous wage distribution (blue line) and exogenous wage distribution (red line).

optimum would harm (approximately) the lower half of the distribution and help the upper half, as we discuss in more detail below.

For comparison, the red line in Figure 6 shows aggregate welfare as a function of τ assuming that wages are exogeneously given at the levels from the benchmark calibrated model. In this case, the optimal progressivity is much higher at $\tau = 0.341$. Indeed, if the planner implements the progressivity that would be optimal with exogeneous wages in a world in which wages are endogeneous, the welfare loss is substantial at 3.15%.

With endogeneous wages, the optimal (lower) progressivity implies higher pre-tax wages for everyone, as shown in Figure 7. This increases welfare by increasing the first element of the welfare function (21). The result is also consistent with Figure 4 in Section 4.1. The wage effect is particularly strong at the low end of the wage distribution. This is the first reason for why with endogeneous wages, optimal tax progressivity is much lower than with exogeneous wages. The interesting finding is that taxes should be substantially less progressive than they currently are in the United States. Decreasing the tax progressivity to the optimal level of $\tau = 0.108$ boosts labor supply (increasing the

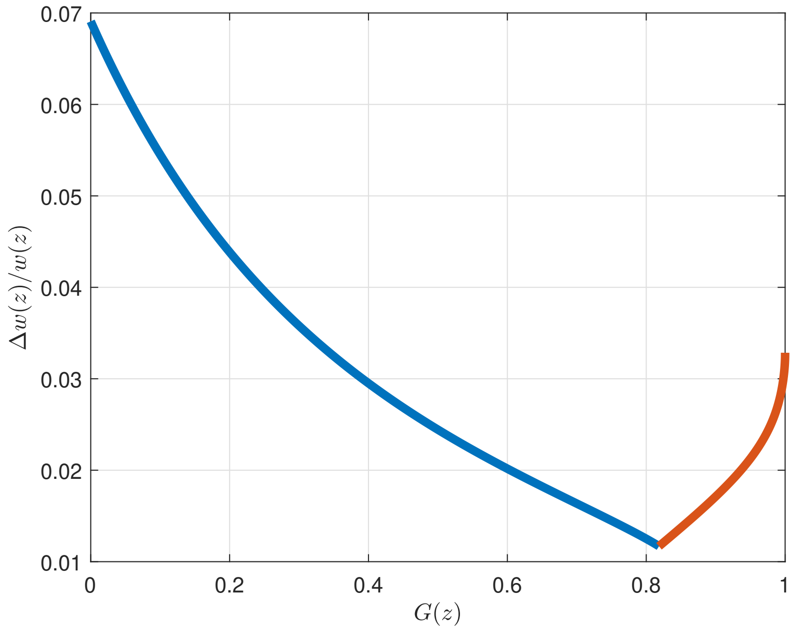


Figure 7: Change in wages between the benchmark progressivity $\tau = 0.186$ and the optimal progressivity $\tau = 0.108$. The horizontal axis shows the population percentiles. Agents who are workers with $\tau = 0.108$ correspond to the blue part of the curve, while managers correspond to the red part.

first element of the welfare function (21) further) increasing the number of workers each manager can supervise, thereby decreasing the share of managers from the benchmark 18.7% to 18.2%.[13] Average wages increase and while wage inequality increases at the top, it decreases at the bottom, consistent with the analytical results of Section 4.1. Table 3 summarizes these results by reporting crucial statistics in our endogeneous-wage model for the various levels of progressivity.

Figures 8 and 9 make the same points graphically. While inequality at the very top decreases with tax progressivity, it increases elsewhere in the distribution, see Figure 8. More general inequality indices, such as variance of logs or the Gini coefficient increase with τ as well, see Figure 9. As equation (21) indicates, the increase in the variance of *log* wages is detrimental for welfare. This is the second reason for why our model with endogeneous wages features low optimal tax progressivity.

[13]Similar effects are described in the context of size-dependent policies in Lopez and Torres (2020) and Garicano et al. (2016).

Table 3: Moments of the Wage Distribution

	CPS Data	Benchmark $\tau = 0.186$	Optimum $\tau = 0.108$	Exo-wage Optimum $\tau = 0.341$
log 50/10 ratio	0.743	0.743	0.707	0.835
log 90/50 ratio	0.877	0.877	0.860	0.914
log 95/90 ratio	0.260	0.306	0.311	0.296
log 99/95 ratio	0.643	0.521	0.528	0.503
Variance of log wages	0.430	0.378	0.358	0.431
Gini of wages	0.386	0.362	0.355	0.401
Mean wages	–	–	1.23%	-3.03%
Welfare gains	–	–	0.36%	-3.15%

This table shows the moments of the wage distribution computed using the CPS data (averaged over 2012-2016). The column denoted 'Benchmark, $\tau = 0.186$' shows the statistics for the calibrated model. The column denoted 'Optimum, $\tau = 0.108$' shows the results for the optimal tax level. The column denoted 'Exo-wage Optimum, $\tau = 0.341$' shows the statistics in our endogeneous-wage model, if the tax progressivity τ is set at 0.341, which would be optimal if wages were exogeneous. 'Mean wages' and 'Welfare gains' are shown relative to the benchmark model. 'Welfare gains' are shown as a fraction of consumption.

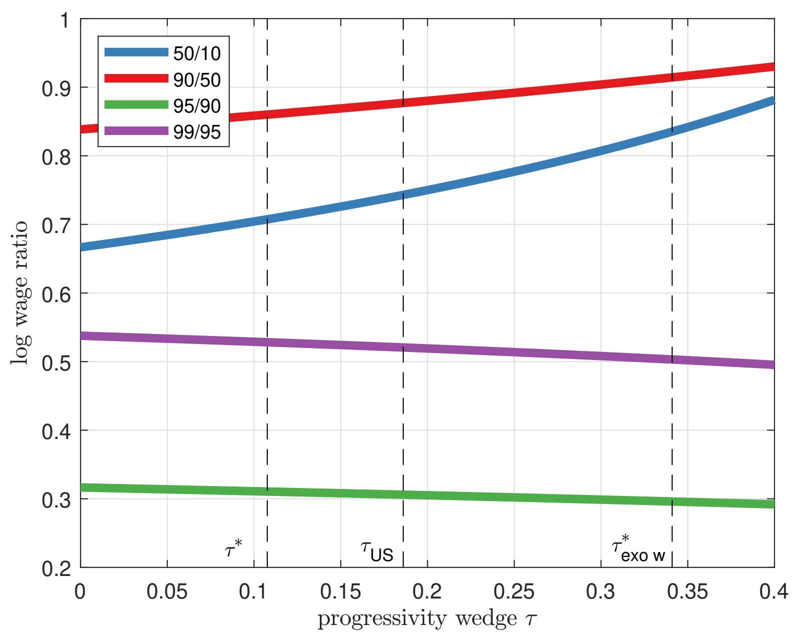


Figure 8: Log-wage ratios as a function of the progressivity wedge τ.

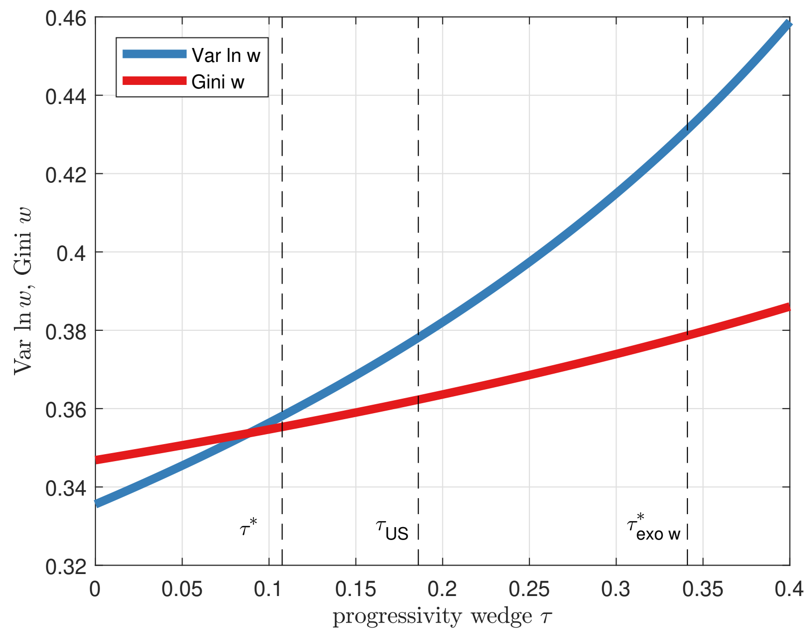


Figure 9: Variance of *log* wages and Gini of wages as a function of the progressivity wedge τ.

As for the tax rates that agents face, the first obvious implication of the decline in progressivity from the calibrated to the optimal level is that marginal taxes decrease for most agents, see Figure 15. Average taxes, on the other hand, decrease for high-income agents and increase for low-income agents, see Figure 16. As a result, low-income agents receive lower transfers (or pay higher taxes), and even though their pre-tax wages increase, their after-tax wages (=consumption) decrease, see Figure 17. Overall, the decrease in tax progressivity thus has the expected effect: it increases efficiency by boosting output, but provides a lower degree of redistribution decreasing welfare of the poor and increasing that of the rich, see Figure 18. These figures are relegated to Appendix B.2.

6.4 Sources of Wage Inequality

The calibrated parameter values show that the model requires i) heterogeneity in the time cost of communication, $\rho > 0$, and ii) a decreasing density of skills, $\gamma > 0$. Without either of those two ingredients, the model cannot simultaneously match both the 90/50 and the 50/10 wage ratio, and the fraction of managers in the data. This is an important result, as it shows that it is necessary to generalize the models of Garicano (2000), Garicano and Rossi-Hansberg (2006) (as well as that of Scheuer and Werning (2017)) for a quantitative assessment of optimal taxes in this class of models.

Figure 10 illustrates this point graphically. 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 keeps the fraction of managers to 18.7%, as in the data. 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, as shown by the red line, which fixes $\rho = 0$ and varies γ. The figure 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.

6.5 Additional Quantitative Exercises

Sensitivity to Labor Supply Elasticity. Labor supply elasticity affects optimal taxes in the expected way. With higher labor supply elasticity, optimal progressivity is expected to be lower, since labor supply responds more strongly to changes in progressivity, see Equation 10. In our model, there is an additional effect. Changes in labor supply map into changes in effective communication costs affecting matching and teamsize, and, ultimately, wages. As as result, optimal tax progressivity responds even more strongly

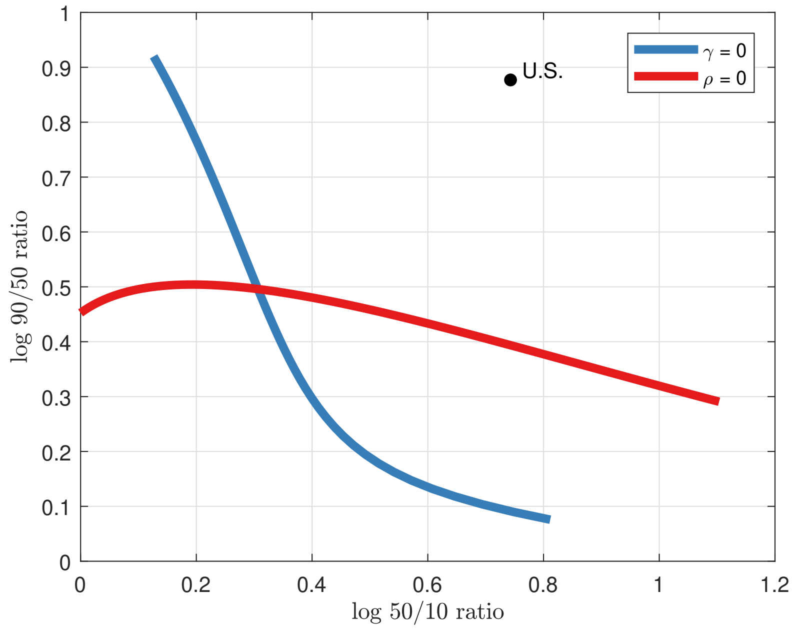


Figure 10: Feasible combinations of the log 50/10 ratio and the log 90/50 ratio conditional on the fraction of managers being 18.7%. Blue line: a model with polynomial distribution but no time cost heterogeneity, $\gamma = 0, \rho \in [-0.72, 2]$. Red line: a model with a uniform skill distribution with time cost heterogeneity, $\rho = 0, \gamma \in [-1, 2]$. Their intersection corresponds to $\rho = \alpha = 0$.

to changes in labor supply elasticity. More specifically, we calibrate our model to a higher labor supply elasticity of 0.75 as suggested by Chetty et al. (2011), implying $\eta = \frac{4}{3}$. As we have argued, wages themselves do not depend on η, but we need to calculate government consumption so that the government consumption-to-output ratio equals 16%. Government consumption is then held constant in levels when we compute optimal taxes. We find that for this level of labor supply elasticity optimal $\tau = 0.024$, meaning that the optimal tax system is almost linear. More generally, our finding that the optimal tax system is more progressive that the current one in the United States holds true for a range of parameters, namely for any labor supply elasticity larger than approximately $\frac{1}{3}$.

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

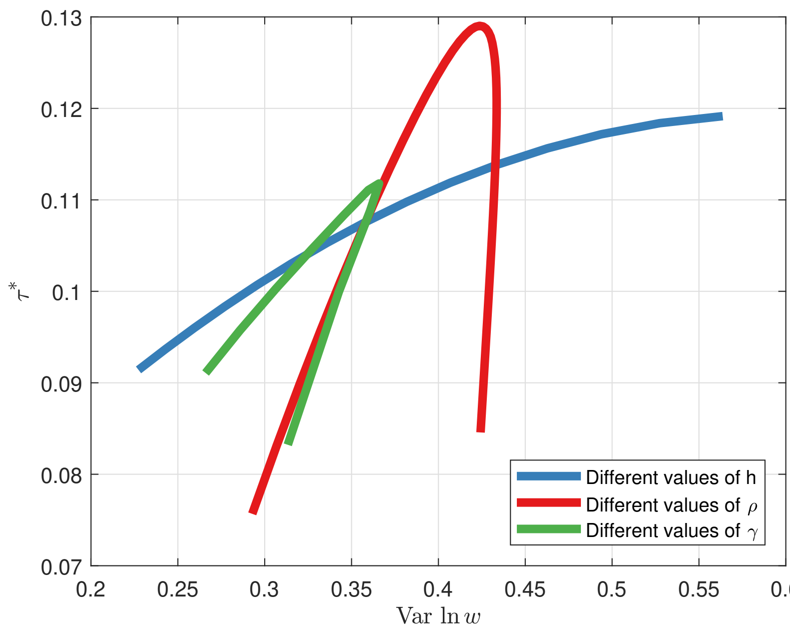
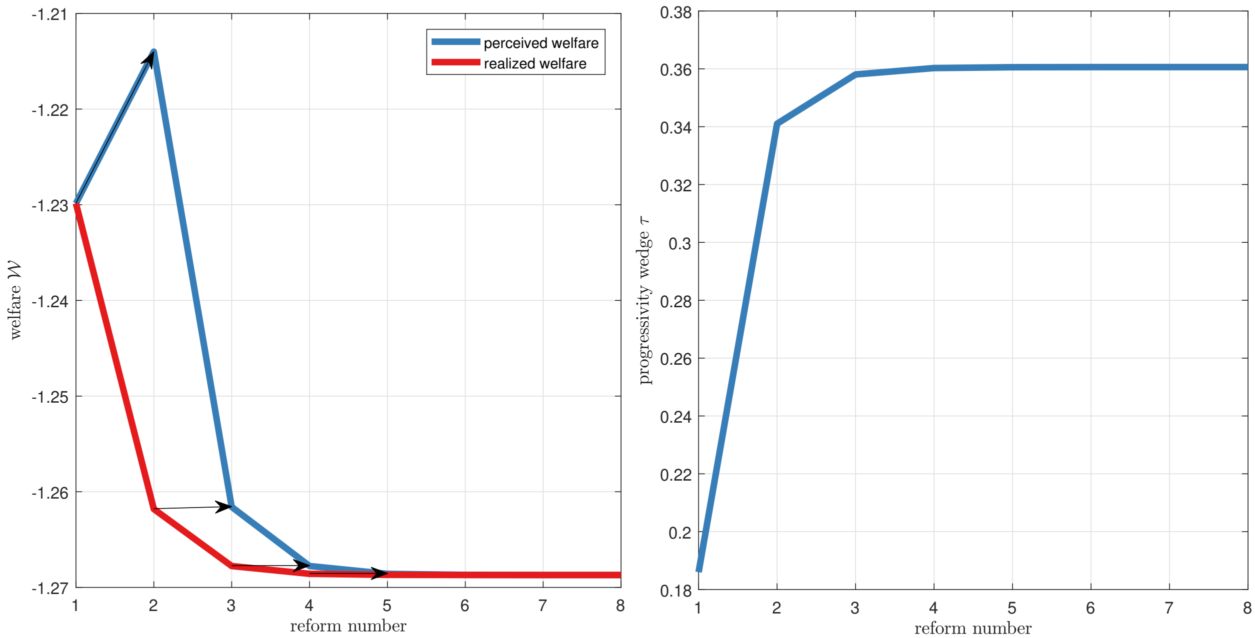


Figure 11: Optimal tax progressivity τ^* as a function of the variance of log wages, depending on the source of the difference.

might imply the same change in wage inequality, as measured, for example, by the variance of log wages. However, the implications for the optimal tax progressivity can be very different. Figure 11 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 the source of the differences is, but that there could also be more than one value of the optimal progressivity for a given variance of log wages; depending on where the wage inequality comes from. This reasoning would also hold for other inequality statistics.

A Naive Government. We now consider choices of a naive government, which believes that the wage structure is exogenous, and maximizes welfare under that perception. Obviously, that perception is wrong, and the naive government will realize that 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



(a) Perceived and realized welfare.

(b) Progressivity wedge chosen.

Figure 12: Solution to the naive government's problem, first 8 rounds of reform. Arrows show perceived welfare gains.

of reforms converges to a self-confirming equilibrium, where the naive planner takes the current wage distribution as given, and the distribution replicates itself post-reform. Figure 12 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. 0.341. However, the wage distribution changes as a result of the tax reform: average wages decrease, wage dispersion increases, and welfare decreases drastically, see Figure 12a. Paradoxically, an increase in the wage dispersion compels the naive government to increase the progressivity wedge even further as Figure 12b 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 0.361. The resulting welfare is substantially lower, namely by 3.82%, than in the original U.S. benchmark.

Optimal Taxes Over Time. Inequality has increased substantially in the last several decades. For example, according the the CPS data, the log 90-50 wage ratio increased

Table 4: Calibrated Parameters Over Time

Parameter	Target	Parameter Value 2012-2016	Target Value 2012-2016	Parameter Value 1979 - 1983	Target Value 1979 - 1983
ρ	log 90/50 ratio	1.540	0.877	1.622	0.697
γ	log 50/10 ratio	2.153	0.743	3.168	0.686
$h/\bar{\ell}$	Share managers	0.429	0.187	0.367	0.141

This table shows calibrated parameters along with the targets for two time periods, 2012-2016 (benchmark) and 1979 - 1983. All target values are averages computed using the CPS data. The details of the data constructions are contained in the appendix.

from 0.697 to 0.877 from 1979-1983 to 2012-2016 and the log 50-10 wage ratio increased from 0.686 to 0.743 over the same time period. At the same time, the fraction of managers increased from 14.1% to 18.7%. In this section we calibrate the model to 1979 - 1983 and compute optimal tax progressivity. We then compare the optimal tax progressivity in 1979-1983 to that which we computed for the benchmark period 2012-2016.

Table 4 compares the calibration of the model in 1979-1983 to the benchmark. Heathcote et al. (2020) report that the progressivity wedge τ was equal to the same value of 0.186 in 1979-1983 and we use this estimate in the calibration. The government consumption-to-GDP ratio is held constant at 16%. No parameter, other than the calibrated ones, changes between the calibrations. As a result, labor supply $\bar{\ell}(\tau)$ does not change and the change in $h/\bar{\ell}(\tau)$ is, in fact, a change in h.

Inequality and the share of managers was lower in 1979-1983. The model's way of matching this difference is a smaller value of h, which decreases the share of managers, the 50-10 log wage ratio and the 90-50 log wage ratio. The larger γ decreases the communication costs overall, but makes differences among workers larger. This decreases the share of managers in 1979-1983 even further, but mitigates the change in the 50-10 log wage ratio. However, it further decreases the 90-50 log wage ratio. As a result, the 90-50 log wage ratio changes more than the 50-10 log wage ratio. The change in ρ has a small effect.[14]

[14]Given Assumption 1 and the assumption that the task distribution is not changing over time, our results should be interpreted as saying that the model can match the observed changes in the U.S. wage

We find that the optimal labor tax progressivity in 1979-1983 is $\tau = 0.095$, lower than in the U.S. tax code and lower than the optimal τ for 2012-2016, which is 0.108. In response to an increase in inequality the planner makes the tax code more progressive.[15]

7 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 as well as those 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 substantially less progressive than the one currently in place in the United States. Ignoring the fact that wages are endogeneous would make the planner choose a substantially higher level of progressivity leading to large welfare losses.

inequality over time by changes in time communication costs *relative* to the distribution of tasks. Caicedo et al. (2019) report a similar finding in a more complex model.

[15] Suppose now, for comparison, that the planner would have to finance the same *level* of G in 2012-2016 as in 1979-1983. In addition, let us take into account the empirical fact that GDP per capita grew by a factor of 1.75 over that period. This would effectively mean that the planner has to finance a smaller G-to-Y in 2012-2016 compared to 1979-1983. In this case, the optimal progressivity in 2012-2016, driven solely by changes in the parameters h, γ and ρ that affect wage inequality, would be even larger, namely 0.142. Assuming that the same G-to-Y needs to be financed does not affect wage inequality, but brings the 2012-2016 optimal τ down to 0.108.

References

Acemoglu, D. and D. Autor (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4 of *Handbook of Labor Economics*, Chapter 12, pp. 1043–1171. Elsevier.

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

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.

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

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

Aronsson, T., M. Wikstrom, and R. Brannlund (1997). Wage determination under nonlinear taxes: Estimation and an application to panel data. *Oxford Economic Papers 49*(3), 404–418.

Autor, D. H. and D. Dorn (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review 103*(5), 1553–1597.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006, May). The Polarization of the U.S. Labor Market. *American Economic Review 96*(2), 189–194.

Benabou, R. (2000). Unequal societies: Income distribution and the social contract. *American Economic Review 90*(1), 96–129.

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

Blomquist, S. and H. Selin (2010). Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates. *Journal of Public Economics 94*(11), 878–889.

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.

Brüggemann, B. (2021). Higher taxes at the top: The role of entrepreneurs. *American Economic Journal: Macroeconomics 13*(3), 1–36.

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

Caicedo, S., J. Lucas, Robert E., and E. Rossi-Hansberg (2019). Learning, career paths, and the distribution of wages. *American Economic Journal: Macroeconomics 11*(1), 49–88.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review 101*(3), 471–75.

Dorn, D. (2009). *Essays on Inequality, Spatial Interaction, and the Demand for Skills*. Ph. D. thesis, University of St. Gallen.

Frish, R., N. Zussman, and S. Igdalov (2020). The Wage Response to a Reduction in Income Tax Rates: The Israeli Tax Reform. *The B.E. Journal of Economic Analysis & Policy 20*(2), 1–22.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy 108*, 874–904.

Garicano, L., C. Lelarge, and J. Van Reenen (2016, November). Firm size distortions and the productivity distribution: Evidence from france. *American Economic Review 106*(11), 3439–79.

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

Garicano, L. and E. Rossi-Hansberg (2015). Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics 7*, 1–30.

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

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

Hansen, C. T., L. H. Pedersen, and T. Sløk (2000). Ambiguous effects of tax progressivity theory and danish evidence. *Labour Economics 7*(3), 335–347.

Heathcote, J., F. Perri, and G. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 19672006. *Review of Economic Dynamics 13*, 15–51.

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.

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

Heathcote, J., K. Storesletten, and G. L. Violante (2020). Presidential Address 2019: How Should Tax Progressivity Respond to Rising Income Inequality? *Journal of the European Economic Association 18*(6), 2715–2754.

Holmlund, B. and A.-S. Kolm (1995). Progressive taxation, wage setting, and unemployment: Theory and swedish evidence. *Swedish Economic Policy Review 2*.

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

Lawson, N. (2019). Taxing the job creators: Efficient taxation with bargaining in hierarchical firms. *Labour Economics 56*(C), 1–25.

Lockwood, B. and A. Manning (1993). Wage setting and the tax system theory and evidence for the united kingdom. *Journal of Public Economics 52*(1), 1–29.

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.

Lucas, Jr., R. E. (1978). On the size distribution of business firms. *Bell Journal of Economics 9*, 508–523.

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

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

Moffitt, R. A. and M. Wilhelm (1998). Taxation and the labor supply: Decisions of the affluent. Working Paper 6621, National Bureau of Economic Research.

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

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

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

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.

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

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

Schneider, K. (2005). Union wage setting and progressive income taxation with heterogeneous labor: theory and evidence from the german income tax reforms 19861990. *Labour Economics 12*(2), 205–222.

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

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

Tamkoc, M. N. (2020). Production complexity, talent misallocation and development. Working paper.

Appendix

A Data Construction

CPS Data. We use the March Current Population Survey (CPS) supplements from 1962 to 2018 (contains information covering 1961 to 2017). In all computations, the CPS sample weights are used. We follow Acemoglu and Autor (2011) in our sample selection and include only full time, full year private employees between the ages of 16 and 64. The restriction of full time, full year employees includes only those employees who worked at least 40 weeks in a year and more than 35 hours in a week. Self-employed are excluded from the sample as we do not have self-employed in the model.

Annual earnings are simply wage income, variable *incwage* (not including business income, variable *incbus*). Hourly earnings (wages) are then the ratio of annual earnings and the product of weeks worked in a year and typical hours worked in a week. Since data on hours is only available after the 1976 CPS, our final sample includes data from this sample to the 2018 CPS data. We only make use of a subset of the years as explained in the main text. Following Heathcote et al. (2010), we drop those employees that report a wage rate less than half the federal minimum wage.

The share of managers is computed as the ratio of the number of managers divided by the total number of managers and non-managerial employees (self-employed are excluded again), where the definition of a managerial occupation follows Dorn (2009) and Autor and Dorn (2013).

Government Consumption-to-GDP Ratio. The government consumption-to-output ratio is recovered from the National Income and Product Accounts (NIPA) data. It is defined as the ratio of nominal government consumption expenditure (line 15 in NIPA Table 3.1) to nominal GDP (line 1 in NIPA Table 1.1.5).

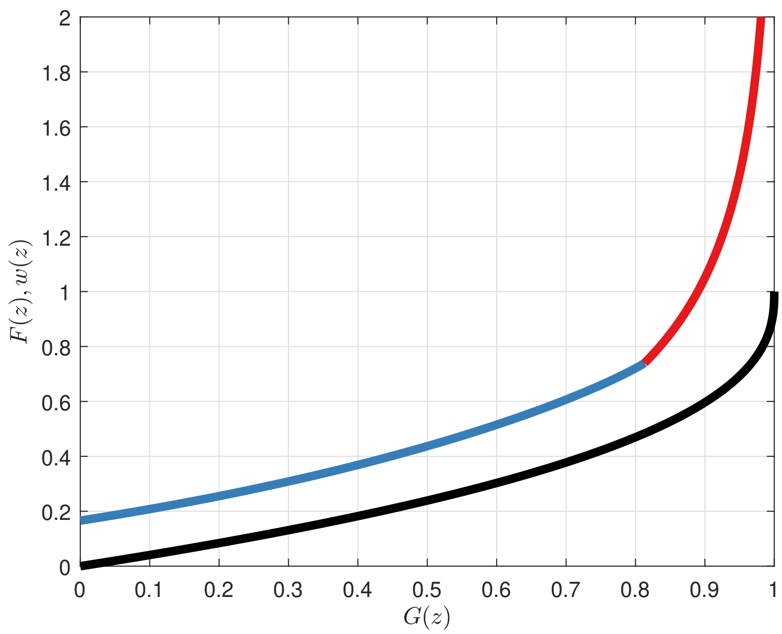


Figure 13: Distribution of wages $w(z)$ (red and blue line) and productivities $F(z)$ (black solid line) in the benchmark economy. Wages are *not* normalized in this figure. The horizontal axis shows the population percentiles.

B Additional Results

B.1 Benchmark Model Properties

The wage distribution differs significantly from the underlying distribution of abilities. Agents at all skill levels experience significant welfare gains relative to autarky, in which every agent would have a wage equal to $F(z)$. Workers gain from being matched with a manager that provides advice for some of their problems, while managers gain from being able to supervise production workers and not having to solve tasks on their own. Figure 13 shows the distribution of skills vs. the distribution of wages/productivities.

Figure 13 shows that the efficiency gains are largest at the top of the wage distribution. This is because very high ability agents are able to supervise a relatively large number of production workers, as shown in Figure 14. The figure shows that while the worst manager supervises about two workers, the top manager supervises ten times as many workers.

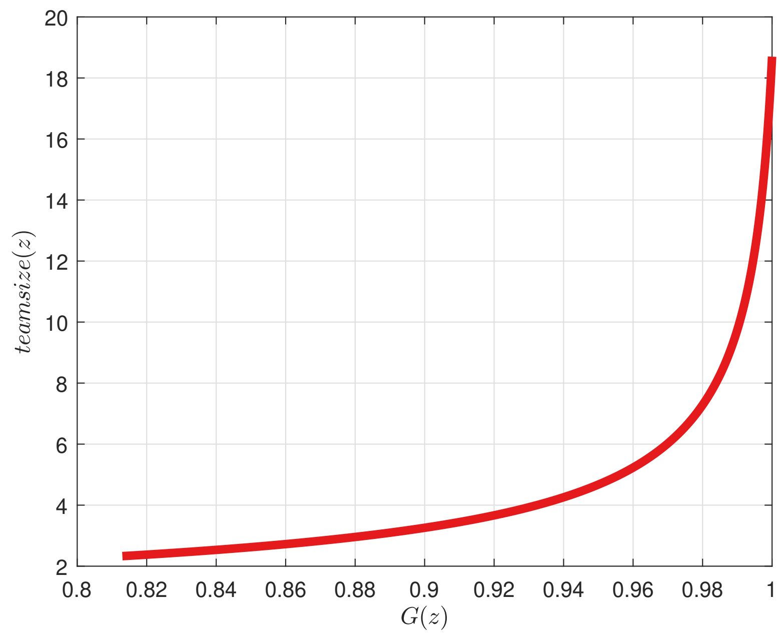


Figure 14: Teamsize in benchmark economy. The horizontal axis shows the population percentiles (of the managers). The vertical axis shows the number of workers supervised by a given manager.

B.2 Optimal Tax Progressivity Relative to Calibrated Benchmark

This section presents additional results comparing the properties of the optimum with the calibrated benchmark.

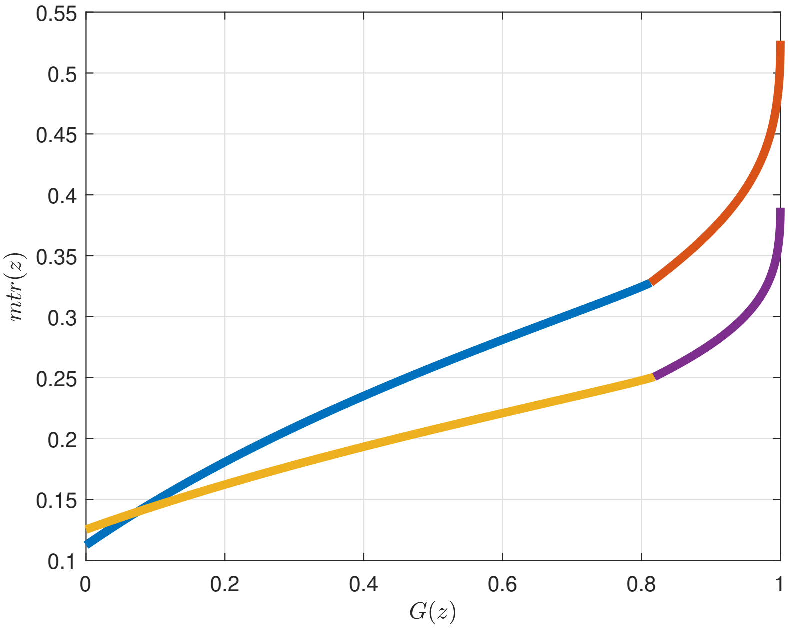


Figure 15: Marginal tax rates. The blue (workers) and red (managers) line is the benchmark calibrated economy with $\tau = 0.186$. The yellow (workers) and purple (managers) line is the counterfactual economy with $\tau = 0.108$. The horizontal axis shows the population percentiles.

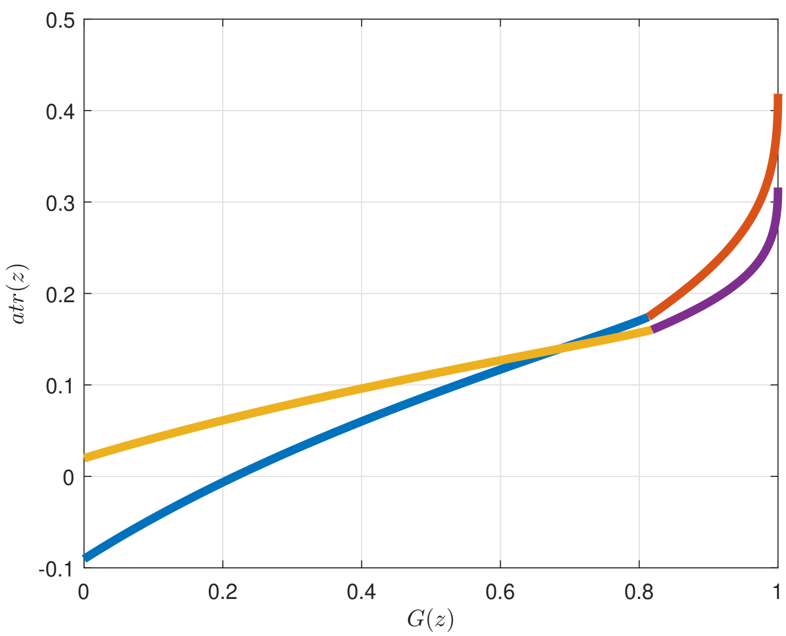


Figure 16: Average tax rates. Blue (workers) and red (managers) line is the benchmark calibrated economy with $\tau = 0.186$. Yellow (workers) and purple (managers) line is the counterfactual economy with $\tau = 0.108$. The horizontal axis shows the population percentiles.

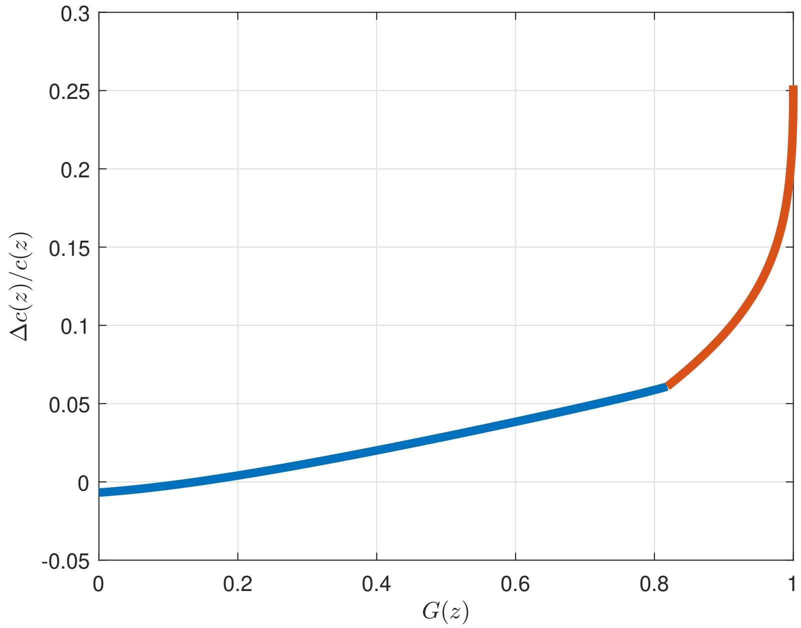


Figure 17: Change in after-tax income/consumption between the benchmark progressivity $\tau = 0.186$ and the optimal progressivity $\tau = 0.108$. The horizontal axis shows the population percentiles. Agents who are workers with $\tau = 0.108$ correspond to the blue part of the curve, while managers to the red part.

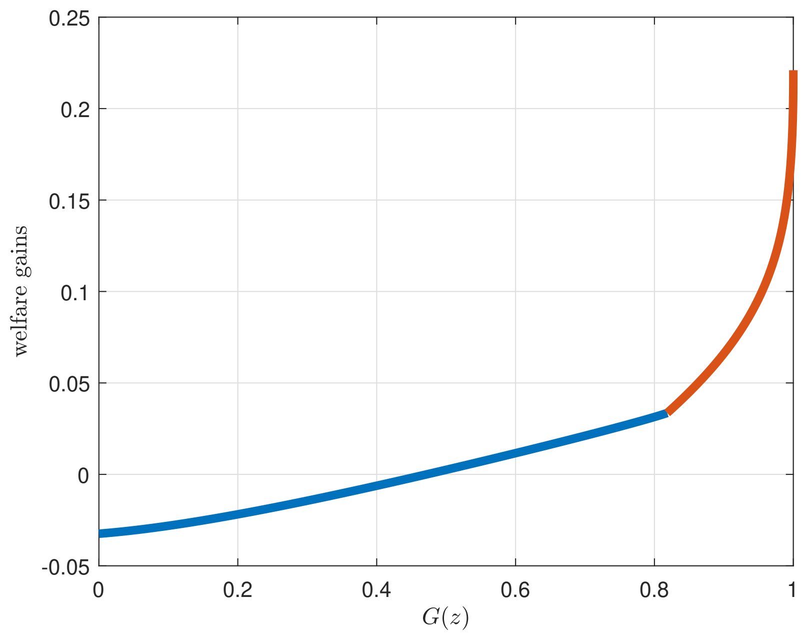


Figure 18: Welfare gains in fraction of consumption between the benchmark progressivity $\tau = 0.186$ and the optimal progressivity $\tau = 0.108$. The horizontal axis shows the population percentiles. Agents who are workers with $\tau = 0.108$ correspond to the blue part of the curve, while managers to the red part.

C Proofs Not For Publication

Lemma 6.

1. *The functions w_i are increasing in x for all $i = 0, \ldots, I+1$.*
2. *If $\theta'' < 0$ and $f' \geq 0$ then the functions are convex for all $i = 0, \ldots, I+1$.*
3. *The functions w_i satisfy $w_i'(z_1) = w_{i+1}'(\underline{z})$ and $w_i''(z_1) = w_{i+1}''(\underline{z})$ for $i = 0, \ldots, I-2$.*

Proof. *Part 1.* The wage functions satisfy, using (15), the following differential equations

$$w_i'(x) = -\theta'\left(m_i(x)\right) w_{i+1}(x), \quad i = 0, 1, \ldots, I. \tag{22}$$

and so are increasing. The top manager's wage is

$$w_I(x) = \frac{F\left(m_I(x)\right) - w_0(x)}{\theta\left(m_{I-1}(x)\right)} - \frac{\theta\left(x\right)}{\theta\left(m_{I-1}(x)\right)} w_1(x) \ldots - \frac{\theta\left(m_{I-2}(x)\right)}{\theta\left(m_{I-1}(x)\right)} w_{I-1}(x). \tag{23}$$

Differentiating the function w_I, we obtain

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

Using $m_i'(x) = \theta\left(m_{i-1}(x)\right)$, (22) and cancelling terms we simplify to

$$w_I'(x) = f\left(m_I(x)\right).$$

The derivative is positive, and so w_I is increasing in x.

Part 2. Differentiating again, $w_I''(x) = f'\left(m_I(x)\right) m_I'(x)$, which is positive if $f' \geq 0$, and so w_I is convex. Differentiating marginal ages below the top layer, we get that for $i = 0, 1, \ldots, I-1$,

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

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

Part 3. Equating the derivatives for $i = 0, 1, \ldots I-2$,

$$w_i'(z_1) = -\theta'\left(m_i(z_1)\right) w_{i+1}(z_1) = -\theta'\left(z_{i+1}\right) w_{i+1}(z_1) = -\theta'\left(m_{i+1}(\underline{z})\right) w_{i+2}(\underline{z}) = w_{i+1}'(\underline{z}),$$

where the penultimate inequality follows from the properties of the matching function and continuity of the age function. Similarly for second derivatives:

$$\begin{aligned}
w_i''(z_1) &= -\theta''\left(m_i(z_1)\right) w_{i+1}(z_1) - \theta'\left(m_i(z_1)\right) w_{i+1}'(z_1) \\
&= -\theta''\left(z_{i+1}\right) w_{i+1}(z_1) - \theta'\left(z_{i+1}\right) w_{i+1}'(z_1) \\
&= -\theta''\left(m_{i+1}(\underline{z})\right) w_{i+2}(\underline{z}) - \theta'\left(m_{i+1}(\underline{z})\right) w_{i+2}'(\underline{z}) \\
&= w_{i+1}''(\underline{z})
\end{aligned}$$

■