Entrepreneurial Taxation and Occupational Choice

Florian Scheuer[*]
Stanford University and NBER

This Version: May 2012

Abstract

This paper analyzes Pareto optimal non-linear taxation of profits and labor income in a private information economy with endogenous firm formation. Individuals differ in both their skill and their cost of setting up a firm, and choose between becoming workers and entrepreneurs. I show that a tax system in which entrepreneurial profits and labor income must be subject to the same non-linear tax schedule makes use of general equilibrium (or "trickle down") effects through wages to indirectly achieve redistribution between entrepreneurs and workers. As a result, constrained Pareto optimal policies can involve negative marginal tax rates at the top and, if available, input taxes that distort the firms' input choices. However, these properties disappear when a differential tax treatment of profits and labor income is possible. In this case, redistribution is achieved directly through the tax system rather than "trickle down" effects, and production efficiency is always optimal.

[*]Email address: scheuer@stanford.edu. I am grateful to Daron Acemoglu and Iván Werning for invaluable guidance and ongoing support. I also thank Richard Blundell, Eddie Dekel, Peter Diamond, Amy Finkelstein, Jonathan Gruber, Caroline Hoxby, Alessandro Pavan, James Poterba, Emmanuel Saez, Karl Scholz, Robert Shimer, Carmen Taveras, Robert Townsend and seminar participants at Chicago, MIT, Northwestern, Stanford, St. Gallen, UCL, Wisconsin Madison and the Review of Economic Studies European Tour at Toulouse, EUI Florence and Warwick for helpful comments and suggestions. All errors are my own.

1 Introduction

The question at what rate business profits should be taxed – notably relative to the tax rates on other forms of income such as labor earnings – is a recurring and controversial theme in the public policy debate. On the one hand, it is often argued that individuals who receive business profits, such as entrepreneurs, tend to be better off than those who do not. Therefore, arguments based on direct redistribution, or "tagging," seem to justify the taxation of profits at a higher rate than other forms of income, as for instance implemented by a corporate income tax and the resulting double taxation of profits both at the firm and individual level. On the other hand, proponents of "supply side" or "trickle down economics" typically emphasize the general equilibrium effects of the tax treatment of businesses. In particular, they point out that a reduction in the entrepreneurs' tax burden encourages entrepreneurial activity and labor demand. It thereby increases wages and hence "trickles down" to medium or lower income workers, achieving redistribution indirectly. From this perspective, a reduced taxation of firm profits, or even a subsidization of entrepreneurial activities, appears optimal.

Underlying these opposing arguments is the question to what degree an optimal tax system should rely on indirect general equilibrium, or "trickle down" effects to achieve redistribution and affect occupational choice. To study this issue formally, I construct a simple model in which the production side is managed by entrepreneurs and both wages and the decision to become a worker or an entrepreneur are endogenous. In particular, I consider a population of individuals characterized by two-dimensional heterogeneity: Agents differ in their cost of setting up a firm, and in their skill, both of which are private information. They can either choose to become a worker, in which case they supply labor at the endogenous wage rate, or select to be an entrepreneur. In this case, they hire workers and provide entrepreneurial effort, which are combined to produce the consumption good.

I characterize Pareto optimal allocations in this economy and demonstrate that the resulting multidimensional screening problem is tractable and allows for a transparent analysis of the issues raised above. The key result is that it crucially depends on the set of available tax instruments whether a Pareto optimal tax system uses general equilibrium effects to achieve redistribution indirectly through "trickle down." I start with characterizing constrained Pareto optimal allocations when the government imposes the same, non-linear tax schedule on both entrepreneurial profits and labor income. Analyzing this uniform taxation case is relevant from a policy perspective since the US and many other countries indeed impose the same (federal) income tax schedule on employed workers and self-employed entrepreneurs. In addition, this policy appears particularly appealing in view of the general presumption that introducing wedges between different forms of income is distorting and should therefore be avoided. For instance, in the Mirrlees Review, Crawford and Freedman (2010) argue that the tax system should aim at neutrality and align tax rates for the employed and self-employed.

However, even though such a tax policy does not explicitly distort the occupational choice margin, it puts severe limitations on the amount of redistribution that can be achieved between entrepreneurs and workers. Due to two-dimensional heterogeneity, the income distributions of workers and entrepreneurs have overlapping supports: There are high-skilled agents who remain workers since they have a high cost of setting up a firm, low-skilled agents who enter entrepreneurship because of their low cost of doing so, and vice versa. It is therefore impossible for a tax system to distinguish workers and entrepreneurs just based on their income. Formally, a policy that does not condition tax schedules on occupational choice puts a no-discrimination constraint on the Pareto problem, since it rules out discriminating between entrepreneurs and workers of different ability levels that are related by the endogenous wage rate.[1]

In the presence of this restriction, a Pareto optimal tax schedule indeed reflects some "trickle down" logic. I show that, if wages are not fixed by technology, the tax system explicitly manipulates incentives in order to induce general equilibrium effects through wages and thus achieve redistribution between entrepreneurs and workers indirectly, given that direct redistribution based on income is not possible. For instance, I provide conditions under which, if the government aims at redistributing from entrepreneurs to workers, top earning entrepreneurs are subsidized at the margin, as this encourages their effort and raises the workers' wage. This relaxes the no-discrimination constraints and therefore allows for additional redistribution in this case. As a result, optimal marginal tax rates not only depend on the skill distribution and wage elasticities of effort, as in standard models, but also on the degree of substitutability of labor and entrepreneurial effort in production. Moreover, I show that if the government has access to additional tax instruments, such as (non-linear) input taxes, it is generally optimal to distort marginal rates of substitution across firms in order to affect wages.

It turns out, however, that these non-standard properties of optimal tax systems, such as negative marginal tax rates at the top and production inefficiency, crucially rely on the restriction that there is only a single tax schedule for both entrepreneurs and workers. In

[1]For this reason, a comparison between uniform and differential taxation of entrepreneurs and workers cannot sensibly be done without accounting for multidimensional heterogeneity. Models with one-dimensional heterogeneity result in income distributions for the occupations that occupy non-overlapping intervals. In this special case, uniform income taxation is not restrictive.

fact, I show that they disappear as soon as the government can make firm profits and labor income subject to different non-linear tax schedules. A Pareto optimal tax policy can now achieve redistribution directly through differential taxation rather than indirectly through general equilibrium effects. For this reason, optimal marginal tax rate formulas no longer depend on substitution elasticities between different inputs in the firms' production function. Furthermore, even if the government could impose distorting input taxes in addition to the non-linear tax schedules on profits and labor income, this is not needed to implement constrained Pareto optima: With differential taxation, production efficiency is always optimal. I also show that, with differential taxation, the "trickle down" logic does not apply. In fact, when redistributing from entrepreneurs to workers, for instance, a Pareto optimal tax system does so in a way that depresses the workers' wage, who are of course more than compensated by tax transfers.

While it is sometimes argued that it may be difficult to distinguish entrepreneurial and other labor income, there are in fact countries that treat employed workers and self-employed small business owners differently for tax and social insurance purposes, as for instance in the UK, where e.g. social insurance contributions differ between employed and self-employed. A common argument against differential taxation is that it may be relatively easy for some individuals to shift their income between categories, leading to distortions. However, rather than a priori assuming that differential taxation is impossible as a result, this can in fact be interpreted in terms of the elasticity of occupational choice. Indeed, I show that this elasticity features prominently in the marginal tax rate formulas for differential taxation, and that the Pareto optimal tax schedules converge as the occupational choice margin becomes very elastic.

I finally compare the optimal tax schedules for profits and labor earnings in an economy that is calibrated to match income distributions and occupational choice between entrepreneurship and employment in the 2007 Survey Consumer Finances. Under various assumptions on the government's redistributive objectives, there robustly emerges an "excess profit tax," i.e. a higher taxation of entrepreneurial profits compared to labor income for individuals of the same skill level, as for instance implemented by a corporate income tax or a separate tax schedule for self-employed persons. I also simulate the effects of optimal tax policy on wages and entrepreneurship for various parameter combinations, with the finding that wages decrease and entrepreneurship is discouraged for most skill levels compared to the laissez-faire equilibrium.

Related Literature. This paper contributes to a large literature that has studied the effects of tax policy on economies explicitly incorporating entrepreneurship. In particular, there has been considerable interest recently in using calibrated dynamic general equilibrium models with an entrepreneurial sector, such as those developed by Quadrini (2000), Meh and Quadrini (2004), and Cagetti and De Nardi (2006), to quantitatively explore how various stylized tax reforms affect the equilibrium wealth distribution, welfare, and investment. For instance, Meh (2005) and Zubricky (2007) have studied the effects of moving from a progressive to a flat income tax system in such economies, Cagetti and De Nardi (2009) have analyzed how an elimination of estate taxation would affect wealth accumulation and welfare, and Panousi (2008) and Kitao (2008) have computed the effects of capital taxation on entrepreneurial investment and capital accumulation. Yet none of these studies have aimed at characterizing and computing optimal tax systems in entrepreneurial economies, which is the focus of the present paper.[2]

In characterizing optimal allocations, my work therefore shares a common goal with Albanesi (2006, 2008) and Shourideh (2010) who have extended the framework of optimal dynamic taxation to account for entrepreneurial investment. More precisely, they consider moral hazard models where entrepreneurs exert some hidden action that affects a stochastic return to capital. Their focus is on characterizing the optimal savings distortions that entrepreneurs should face when the government provides insurance for entrepreneurial investment risk. Similarly, Chari, Golosov, and Tsyvinski (2002) examine optimal intertemporal wedges in a dynamic economy with start-up firms and incomplete markets. In contrast to this literature, I focus on characterizing the optimal taxation of profits and labor income in a static general equilibrium model that emphasizes endogenous entry into entrepreneurship and how taxes affect the effort-leisure wedge of entrepreneurs versus workers and thus wages.

The paper also builds on earlier research on optimal income taxation in models with endogenous wages and occupational choice, such as Feldstein (1973), Zeckhauser (1977), Allen (1982), Boadway, Marceau, and Pestieau (1991), and Parker (1999). This literature has restricted attention to linear taxation and typically ruled out a differential tax treatment of the occupational groups. An exception is the work by Moresi (1997), who considers non-linear taxation of profits. However, in his model, the occupational choice margin is considerably simplified and heterogeneity is confined to affect one occupation only, not the other. Stiglitz (1982) and Naito (1999) study optimal non-linear taxation in economies with two ability types and endogenous wages. While some of their results translate to

[2]There is also related research that has focused on how taxes affect more specific aspects of entrepreneurial activity. For example, Kanbur (1981), Kihlstrom and Laffont (1979), Kihlstrom and Laffont (1983), Christiansen (1990) and Cullen and Gordon (2007) have examined the effects of taxation on entrepreneurial risk-taking. Moreover, the consequences of a differential tax treatment of corporate versus non-corporate businesses (or of its removal) for investment have been the focus of Gordon (1985), Gravelle and Kotlikoff (1989) and Meh (2008). See Gentry and Hubbard (2000) for an overview of these issues. I abstract from a distinction of firms in corporate and non-corporate in this paper.

properties of Pareto optimal tax systems with uniform taxation of profits and income, their models do not include different occupational groups. Therefore, neither of these papers allow for the comparison of uniform and differential taxation of profits and income, and of the optimal (non-linear) tax schedules of workers and entrepreneurs in the case of differential taxation, which is performed here.

In addition, restricting heterogeneity to affect one occupation only, or tax schedules to be linear, sidesteps the complexities of multidimensional screening, which emerges naturally in the present model. In fact, few studies in the optimal taxation literature have attempted to deal with multidimensional screening problems until recently. Closest to the formal modelling approach used here is the recent contribution by Kleven, Kreiner, and Saez (2009) with an application to the optimal income taxation of couples. More generally, this paper builds on the large literature on optimal income taxation following the seminal contributions by Diamond and Mirrlees (1971) and Mirrlees (1971). However, rather than focusing on allocations that maximize some utilitarian social welfare criterion, I aim at characterizing the set of Pareto optimal tax policies, sharing the spirit of Werning (2007).

The structure of the paper is as follows. Section 2 introduces the baseline model and the equilibrium without taxation. In Section 3, I start with characterizing Pareto optimal tax policies when the same (non-linear) tax schedule is applied to both entrepreneurial profits and labor income. Properties of Pareto optimal tax schedules and the optimality of production distortions are discussed. As I show in Section 4, these properties disappear when profits and income can be made subject to different tax schedules. Section 4 also computes the two tax schedules for a calibrated economy. Finally, Section 5 concludes. Most of the proofs are relegated to the appendix.

2 The Baseline Model

2.1 Preference Heterogeneity and Occupational Choice

I consider a unit mass of heterogeneous individuals who are characterized by a two-dimensional type vector $(\theta,\phi) \in [\underline{\theta},\overline{\theta}] \times [0,\overline{\phi}_\theta]$, where θ will be interpreted as an individual's skill, and ϕ as an individual's cost of becoming an entrepreneur, as explained in more detail below.[3] $F(\theta)$ is the cumulative distribution function of θ and $G_\theta(\phi)$ the cumulative distribution function of ϕ conditional on θ, both assumed to allow for density functions $f(\theta)$ and $g_\theta(\phi)$. Note that this allows for an arbitrary correlation between θ and ϕ. Both θ and ϕ are an individual's private information.

[3] I assume $\underline{\theta} > 0$ and $\overline{\theta}, \overline{\phi}_\theta < \infty$ for most of the analysis.

Agents can choose between two occupations: They can become a worker, in which case they supply effective labor l at the (endogenous) wage w. Abstracting from income effects, I assume preferences over consumption c and labor to be quasi-linear with

$$U(c, l, \theta) \equiv c - \psi(l/\theta).$$

An individual's disutility of effort $\psi(.)$ is assumed to be twice continuously differentiable, increasing and convex. A particular specification, used later, is given by $\psi(l/\theta) = (l/\theta)^{1+1/\varepsilon}/(1+1/\varepsilon)$, which implies that the individual's elasticity of labor supply with respect to the wage is constant and equal to ε. θ captures an individual's skill type in the sense that a higher value of θ implies that the individual has a lower disutility of providing a given amount of effective labor l.

Alternatively, an agent may select to become an entrepreneur. In this case, she hires effective labor L and provides effective entrepreneurial effort E to produce output of the consumption good Y, where $Y(L, E)$ is a concave neoclassical firm-level production function with constant returns to scale. An entrepreneur's profits are then

$$\pi = Y(L, E) - wL,$$

and her utility is given by

$$U(\pi, E, \theta) - \phi \equiv \pi - \psi(E/\theta) - \phi.$$

ϕ is a heterogeneous utility cost of becoming an entrepreneur, which is distributed in the population as specified above, possibly depending on the skill type θ. Thus, θ determines an individual's skill in both occupations, but in addition, people differ in their idiosyncratic preferences for one of the two occupations, as captured by ϕ. The cost ϕ can therefore be interpreted as a shortcut for heterogeneity in the population that is not otherwise captured in the present model explicitly, such as a differences in setup costs, attitudes towards entrepreneurial risks, or access to entrepreneurial capital.[4] As a result of the two-dimensional heterogeneity, there will not be a perfect ranking between occupational choice and skill type (and thus income): For a given θ, there are individuals who

[4]While I assume $\phi \geq 0$, i.e. that entrepreneurship is associated with some cost for all individuals, the following analysis does not rely on this assumption. Rather, I could allow for the support of ϕ to include negative numbers, accounting for the fact that some individuals value non-pecuniary benefits from being an entrepreneur, such as flexibility of schedules and being one's own boss. The only advantage of assuming ϕ to be non-negative is that, in equilibrium, entrepreneurs receive a higher return on their effort than workers. See Section 4.2 for a detailed discussion of evidence on this.

enter entrepreneurship and others who become workers due to their different ϕ-type. This is an empirically attractive implication of the present specification, since it is true that, in reality, the income distributions of workers and entrepreneurs have overlapping supports.[5]

2.2 The Equilibrium without Taxes

In order to introduce the mechanics of this basic model, let me start with briefly discussing the equilibrium without taxes. Taking the wage w as given, conditional on becoming a worker, an individual of skill-type θ solves $\max_l wl - \psi(l/\theta)$ with solution $l^*(\theta, w)$ and indirect utility $v_W(\theta, w) \equiv wl^*(\theta, w) - \psi(l^*(\theta, w)/\theta)$. Similarly, conditional on becoming an entrepreneur, type θ solves $\max_{L,E} Y(L, E) - wL - \psi(E/\theta)$ with solution $L^*(\theta, w)$, $E^*(\theta, w)$ and indirect utility $v_E(\theta, w)$. Then the occupational choice decision for individuals of type θ is determined by the critical cost value

$$\tilde{\phi}(\theta, w) \equiv \begin{cases} 0 & \text{if} \quad v_E(\theta, w) - v_W(\theta, w) < 0 \\ \overline{\phi}_\theta & \text{if} \quad v_E(\theta, w) - v_W(\theta, w) > \overline{\phi}_\theta \\ v_E(\theta, w) - v_W(\theta, w) & \text{otherwise,} \end{cases} \tag{1}$$

so that all (θ, ϕ) with $\phi \leq \tilde{\phi}(\theta, w)$ become entrepreneurs, and the others workers. With this notation, an equilibrium without taxes can be defined as follows:

Definition 1. *An equilibrium without taxes is a wage w^* and an allocation $\{l^*(\theta, w^*), L^*(\theta, w^*), E^*(\theta, w^*)\}$ for all $\theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$ such that the labor market clears, i.e.*

$$\int_\Theta G_\theta(\tilde{\phi}(\theta, w^*))L^*(\theta, w^*)dF(\theta) = \int_\Theta (1 - G_\theta(\tilde{\phi}(\theta, w^*)))l^*(\theta, w^*)dF(\theta). \tag{2}$$

In fact, the entrepreneurs' utility maximization problem can be decomposed as follows. Since their labor demand L only affects profits and not the other components of their utility, for given E and w, entrepreneurs of all types θ solve the same problem $\max_L Y(L, E) - wL$ with the conditional labor demand function $L^c(E, w)$ as solution such that $Y_L(L^c(E, w), E) = w$. Under the assumption of constant returns to scale, Euler's

[5] This is in contrast to models where occupational choice is only based on skill heterogeneity, such as Boadway, Marceau, and Pestieau (1991) and Moresi (1997), and where it is assumed that one occupation rewards ability more than the other. Then there exists a critical skill level such that all higher skilled agents select into the high-reward occupation, and lower-ability agents into the other. This results in income distributions for the two occupations that occupy non-overlapping intervals (see e.g. Parker (1999)).

theorem implies

$$Y(L^c(E,w),E) = Y_L(L^c(E,w),E)L^c(E,w) + Y_E(L^c(E,w),E)E,$$

and thus an entrepreneur's profits are given by

$$\pi = Y(L^c(E,w),E) - wL^c(E,w) = Y_E(L^c(E,w),E)E.$$

Hence, entrepreneurs can be thought of just receiving a different wage $\tilde{w} \equiv Y_E$ on their effort. Moreover, there exists a decreasing one-to-one relationship between the workers' and the entrepreneurs' wage $\tilde{w}(w)$:[6] The entrepreneurs' wage $\tilde{w}$ is high if the entrepreneurial effort to labor ratio used in production is low, which means that the marginal product of labor and thus the workers' wage is low.

With these insights, the following properties of the equilibrium without taxes can be established:

Proposition 1. *Consider the no tax equilibrium as defined in Definition 1. Then*
(i) the entrepreneurs' wage exceeds the workers' wage, i.e. $\tilde{w}^* \equiv \tilde{w}(w^*) > w^*$, *and for all* $\theta \in \Theta$, $E^*(\theta,\tilde{w}^*) > l^*(\theta,w^*)$,
(ii) the critical cost value for occupational choice $\tilde{\phi}(\theta,w^*)$ *is increasing in* θ, *and*
(iii) the share of entrepreneurs $G_\theta(\tilde{\phi}(\theta,w^*))$ *is increasing in* θ *if* $G_{\theta'}(\phi) \succeq_{FOSD} G_\theta(\phi)$ *for* $\theta' \leq \theta$.

Proof. (i) Recall that $v_W(\theta,w^*) = \max_l w^*l - \psi(l/\theta)$ and $v_E(\theta,\tilde{w}^*) = \max_E \tilde{w}^*E - \psi(E/\theta)$. Suppose, by way of contradiction, $\tilde{w}^* \leq w^*$. Then $v_E(\theta,\tilde{w}^*) \leq v_W(\theta,w^*)$, and hence by (1), $\tilde{\phi}(\theta,w^*) = 0$ for all $\theta \in \Theta$. Therefore (2) cannot be satisfied. To see that $E^*(\theta,\tilde{w}^*) > l^*(\theta,w^*)$, note first that, since the function $wl - \psi(l/\theta)$ is supermodular in (w,l), $l^*(\theta,w)$ is increasing in w by Topkis' theorem (see Topkis (1998)). By the same argument, since $\tilde{w}^* > w^*$ from (i), $E^*(\theta,\tilde{w}^*) > l^*(\theta,w^*)$ for all $\theta \in \Theta$.
(ii) Using the results from (i),

$$\frac{\partial\tilde{\phi}(\theta,w^*)}{\partial\theta} = \psi'\left(\frac{E^*(\theta,\tilde{w}^*)}{\theta}\right)\frac{E^*(\theta,\tilde{w}^*)}{\theta^2} - \psi'\left(\frac{l^*(\theta,w^*)}{\theta}\right)\frac{l^*(\theta,w^*)}{\theta^2} > 0 \;\; \forall\theta \in \Theta$$

by the envelope theorem and convexity of ψ.
(iii) If $G_{\theta'}(\phi) \succeq_{FOSD} G_\theta(\phi)$ for $\theta' \leq \theta$, then

$$G_{\theta'}(\tilde{\phi}(\theta',w^*)) \leq G_{\theta'}(\tilde{\phi}(\theta,w^*)) \leq G_\theta(\tilde{\phi}(\theta,w^*)) \;\text{ for } \theta' \leq \theta,$$

where the first inequality follows from (ii) and the second from first-order stochastic dominance. □

[6] This is because, by linear homogeneity of Y, both Y_L and Y_E are homogeneous of degree zero and hence functions of $x \equiv E/L$ only. Then $\tilde{w}(w)$ is a decreasing function because $\tilde{w} = Y_E(x) = Y_E(Y_L^{-1}(w))$ and $Y_E(x)$ is decreasing and $Y_L(x)$ increasing in x by concavity of Y (and therefore the inverse $Y_L^{-1}(w)$ from $Y_L(x) = w$ is a decreasing function).

Proposition 1 summarizes intuitive properties of wages and occupational choice in equilibrium: First, the entrepreneurs' wage $\tilde{w}^*$ must be higher than that of the workers w^* in equilibrium. The reason is that, when deciding whether to become a worker or an entrepreneur, an individual of a given skill type considers two variables: The different wage that she can earn when becoming an entrepreneur rather than a worker, and the cost ϕ she has to incur when doing so. Clearly, if the entrepreneurs' wage were lower than that of workers, there would be no trade-off and nobody would choose to enter entrepreneurship, which cannot be an equilibrium. The entrepreneurs' higher wage then immediately implies that they exert more effort and earn higher profits than workers of the same ability level. While this is a direct consequence of the assumption that $\phi \geq 0$, it is in line with empirical evidence on returns to entrepreneurship. For instance, De Nardi, Doctor, and Krane (2007) find that entrepreneurs have higher incomes than workers, and Berglann, Moen, Roed, and Skogstrom (2009) confirm this pattern for wages, controlling for hours. Moreover, based on data from the 2007 Survey of Consumer Finances (SCF), I find the same relationship between returns to entrepreneurship and employment, as will be discussed in Section 4.2.[7] In addition, since this is a static model, Proposition 1 can be interpreted in terms of lifetime incomes, or wealth. There is strong evidence that entrepreneurs have more wealth than workers, for instance in Quadrini (2000) and Cagetti and De Nardi (2006).

The second result in the proposition is that, the higher the skill type θ, the more the wage difference matters compared to the cost, which is why the critical cost value $\tilde{\phi}(\theta, w^*)$ increases with θ. Finally, the same holds for the share of entrepreneurs in equilibrium as a function of skill whenever skill and disutility from entrepreneurship are independent or such that higher skills tend to have a lower disutility from being an entrepreneur in the first-order stochastic dominance sense. More generally, while such a correlation between θ and ϕ may strike as plausible, the model is flexible enough to generate more complicated relationships between income and the share of entrepreneurs through the dependence of the cost distribution on θ, as captured by $G_\theta(\phi)$.[8] Proposition 1 thus demonstrates that, while the basic model is admittedly stylized and quite different from other models of entrepreneurship, it is able to produce reasonable predictions about empirical relationships, and to point out how they depend on the underlying heterogeneity in the population.

[7] Hamilton (2000) and Blanchflower (2004) find lower returns to entrepreneurship than to employment. However, their concept of entrepreneurship is different, setting it equal to self-employment. As I will discuss in Section 4.2, I consider individuals as entrepreneurs if they are not only self-employed, but also own and actively manage a business and hire at least two employees.

[8] In Section 4.2, $G_\theta(\phi)$ will be calibrated to match the relationship between income and entrepreneurship found in the data.

3 Uniform Tax Treatment of Profits and Income

3.1 A Constrained Pareto Problem

While the no tax equilibrium represents a particular point on the Pareto-frontier, other Pareto optimal allocations can be implemented by suitable tax policies. Let me start with characterizing the resulting Pareto-frontier under the assumption that the government imposes a single non-linear tax schedule $T(.)$ that applies to both the workers' labor income $y \equiv wl$ and the entrepreneurs' profits π in the same way. Such a tax system may seem particularly appealing on the grounds of neutrality, since it does not explicitly distort the occupational choice margin (see e.g. Crawford and Freedman (2010) in the Mirrlees Review). It is also the system that is in place for employed workers and self-employed small business owners in many countries, including the US. Then the question is to what degree a Pareto-optimal tax policy makes use of general equilibrium ("trickle down") effects through the workers' wage to achieve redistribution indirectly.

With a tax on profits $T(\pi)$, entrepreneurs solve $\max_{L,E} Y(L,E) - wL - T(Y(L,E) - wL) - \psi(E/\theta)$ and thus their labor demand is always undistorted such that $Y_L = w$ for all skill types θ. This implies that, by the same arguments as in the preceding section, entrepreneurs can be viewed as just receiving a different wage $\tilde{w} = Y_E$ than workers on their effort E. Hence, entrepreneurs of type θ choose their effort so as to solve $\max_E \tilde{w}E - T(\tilde{w}E) - \psi(E/\theta)$, and workers of type θ solve $\max_l wl - T(wl) - \psi(l/\theta)$. Since they face the same tax schedule $T(.)$, it immediately follows that the profits generated by an entrepreneur of type θ and the income earned by a worker of type θ' with the same "total" wage on their effort, i.e. such that $\tilde{w}\theta = w\theta'$, must be equal:

$$\tilde{w}E(\theta) = wl\left(\frac{\tilde{w}}{w}\theta\right) \tag{3}$$

for all $\theta \in [a,b]$ with $a = \max\{\underline{\theta}, (w/\tilde{w})\underline{\theta}\}$ and $b = \min\{\overline{\theta}, (w/\tilde{w})\overline{\theta}\}$. This is a no-discrimination constraint on the Pareto-problem that results from the restriction that both profits and income must be subject to the same tax schedule $T(.)$: With this instrument, it is impossible for the government to discriminate between individuals who earn the same overall wage, even if in different occupations, namely entrepreneurs of skill θ and workers of the rescaled skill $(\tilde{w}/w)\theta$, whereby the rescaling factor $\tilde{w}/w$ is endogenous and corresponds to the ratio between the marginal products of entrepreneurial effort and labor. The same no-discrimination constraints have to hold for consumption (or, equivalently, utility), as I will note formally below.

In addition to the no-discrimination constraints, any allocation that can be implemented with the single non-linear tax schedule $T(.)$ must satisfy the following incentive compatibility constraints by the revelation principle. Suppose the social planner assigns labor supply $l(\theta)$ and consumption $c_W(\theta)$ to each individual of skill type θ who chooses to become a worker, and a labor demand and entrepreneurial effort bundle $L(\theta), E(\theta)$ and consumption $c_E(\theta)$ to each θ-type who selects into entrepreneurship.[9] Then the incentive constraints can be written as

$$c_W(\theta) - \psi\left(\frac{l(\theta)}{\theta}\right) \geq c_W(\hat{\theta}) - \psi\left(\frac{l(\hat{\theta})}{\theta}\right) \quad \forall \theta, \hat{\theta} \in \Theta, \tag{4}$$

$$c_E(\theta) - \psi\left(\frac{E(\theta)}{\theta}\right) \geq c_E(\hat{\theta}) - \psi\left(\frac{E(\hat{\theta})}{\theta}\right) \quad \forall \theta, \hat{\theta} \in \Theta \tag{5}$$

and

$$Y_L(L(\theta), E(\theta)) = w \;\; \forall \theta \in \Theta. \tag{6}$$

Constraint (6) is a result of the fact that the profit tax $T(.)$ does not distort the entrepreneurs' labor demand, and so all firms set it so as equalize the marginal product of labor to the workers' wage. Hence, the marginal products of entrepreneurial effort are also equalized across firms with

$$Y_E(L(\theta), E(\theta)) = \tilde{w} \;\; \forall \theta \in \Theta. \tag{7}$$

Defining the indirect utility functions as

$$v_W(\theta) \equiv \max_{\hat{\theta} \in \Theta} c_W(\hat{\theta}) - \psi\left(\frac{l(\hat{\theta})}{\theta}\right) \quad \text{and} \quad v_E(\theta) \equiv \max_{\hat{\theta} \in \Theta} c_E(\hat{\theta}) - \psi\left(\frac{E(\hat{\theta})}{\theta}\right) \quad \forall \theta \in \Theta,$$

and observing that preferences satisfy single-crossing, it is a standard result that the incentive constraints (4) and (5) are satisfied if and only if the envelope conditions

$$v_W'(\theta) = \psi'\left(\frac{l(\theta)}{\theta}\right)\frac{l(\theta)}{\theta^2} \quad \text{and} \quad v_E'(\theta) = \psi'\left(\frac{E(\theta)}{\theta}\right)\frac{E(\theta)}{\theta^2} \;\; \forall \theta \in \Theta \tag{8}$$

hold and

$$l(\theta) \quad \text{and} \quad E(\theta) \quad \text{are non-decreasing.}^{10} \tag{9}$$

[9] Since the cost ϕ enters utility additively, it is straightforward to see that, conditional on occupational choice, individuals cannot be further separated based on ϕ. Hence, indexing the allocation $\{l(\theta), c_W(\theta), L(\theta), E(\theta), c_E(\theta)\}$ by θ only is without loss of generality.

[10] See, for instance, Fudenberg and Tirole (1991), Theorems 7.2 and 7.3, and Kleven, Kreiner, and Saez (2009), online appendix.

Incentive compatibility also requires that the critical cost values for occupational choice are given by

$$\tilde{\phi}(\theta) = v_E(\theta) - v_W(\theta) \;\; \forall \theta \in \Theta.^{11} \tag{10}$$

Finally, the fact that a single tax schedule cannot discriminate between entrepreneurs of skill θ and workers of skill $(\tilde{w}/w)\theta$ implies that their consumption (and, by (3), their utility) must be the same, i.e.

$$v_E(\theta) = v_W\left(\frac{\tilde{w}}{w}\theta\right) \;\; \forall \theta \in [a, b]. \tag{11}$$

Summarizing these insights, the Pareto problem can be written as follows. Let the social planner attach Pareto-weights to individuals depending on their two-dimensional type vector, as captured by cumulative distribution functions $\tilde{F}(\theta)$ and $\tilde{G}_\theta(\phi)$. Then the program is

$$\max_{\substack{\{E(\theta), L(\theta), l(\theta), v_E(\theta), \\ v_W(\theta), \tilde{\phi}(\theta), w, \tilde{w}\}}} \int_\Theta \left[\tilde{G}_\theta(\tilde{\phi}(\theta)) v_E(\theta) - \int_{\underline{\phi}}^{\tilde{\phi}(\theta)} \phi d\tilde{G}_\theta(\phi) + (1 - \tilde{G}_\theta(\tilde{\phi}(\theta))) v_W(\theta) \right] d\tilde{F}(\theta)$$

subject to

$$\int_\Theta G_\theta(\tilde{\phi}(\theta)) L(\theta) dF(\theta) \leq \int_\Theta (1 - G_\theta(\tilde{\phi}(\theta))) l(\theta) dF(\theta), \tag{12}$$

$$\begin{aligned} \int_\Theta G_\theta(\tilde{\phi}(\theta)) \left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi(E(\theta)/\theta) \right] dF(\theta) & \\ - \int_\Theta (1 - G_\theta(\tilde{\phi}(\theta))) \left[v_W(\theta) + \psi(l(\theta)/\theta) \right] dF(\theta) & \;\; \geq \;\; 0, \end{aligned} \tag{13}$$

and constraints (3), (6), (7), (8), (9), (10) and (11). Inequality (12) requires the total amount of labor demand assigned to entrepreneurs not to exceed the total amount of labor supply assigned to workers. Similarly, (13) is the resource constraint that makes sure that the total amount of resources produced by the entrepreneurs in the economy covers the consumption allocated to entrepreneurs and workers.[12]

[11] Again, additive separability of ϕ implies that any incentive compatible allocation must take a threshold form such that, for all θ, there is some critical value $\tilde{\phi}(\theta)$ such that all $\phi \leq \tilde{\phi}(\theta)$ become entrepreneurs and the others workers.

[12] As is standard in the screening literature, I solve the Pareto problem ignoring the monotonicity constraint (9), assuming that it is not binding. Otherwise, the Pareto optimum would involve bunching of some types. In the numerical analysis in Section 4.2, I check whether the monotonicity constraint is satisfied at the optimum, and find that bunching does not arise.

3.2 Properties of Constrained Pareto Optimal Tax Systems

Inspection of the constrained Pareto problem reveals that the wages w and $\tilde{w}$ enter the program through the no-discrimination constraints (3), a property that is referred to as a pecuniary externality. Intuitively, wages have first-order effects on welfare as their ratio determines to what extent the income distributions of the two occupations overlap, and hence which workers and entrepreneurs must be treated the same as a result of the non-discriminating tax treatment of profits and labor income. This has consequences for the amount of redistribution that can be achieved with a single tax schedule. For this reason, whenever wages are not fixed by technology, the optimal tax policy exhibits some non-standard properties. The following two propositions summarize characteristics of constrained Pareto optimal tax systems.

Proposition 2. *(i) At any Pareto-optimum, $\tilde{w} > w$, and $\tilde{w}E(\theta) > wl(\theta)$ for all $\theta \in \Theta$.*
(ii) $T'(wl(\underline{\theta})) = T'(\tilde{w}E(\overline{\theta})) = 0$ if $Y(L, E)$ is linear.
(iii) Otherwise, $T'(wl(\underline{\theta}))$ and $T'(\tilde{w}E(\overline{\theta}))$ have opposite signs whenever (3) and (11) bind for some $\theta \in \Theta$.
(iv) Suppose the workers' effort $l(\theta)/\theta$ is increasing in θ. If, at the optimum, the no-discrimination constraints (3) and (11) bind in the $\geq$-direction, then $T'(wl(\underline{\theta})) > 0$ and $T'(\tilde{w}E(\overline{\theta})) < 0$ (otherwise, the opposite holds).

Proof. See Appendix A.1. □

The first part of Proposition 2 holds for the same reason as in the equilibrium without taxes: Since profits and labor income are subject to the same tax treatment, the entrepreneurs' marginal product must be higher than the workers', because otherwise nobody would choose to set up a firm. This implies that the top earner at any Pareto optimum is an entrepreneur, and the bottom earner a worker.[13]

Part (ii) establishes that the standard results are obtained for the bottom and top marginal tax rates if technology is linear so that wages are fixed: Both the bottom and the top earners should face a zero marginal tax rate, as in Mirrlees (1971). However, this is no longer necessarily true when technology is not linear, as shown in part (iii) of Proposition 2. In this case, since the tax system is restricted not to treat labor income and profits differently, and the ratio of wages determines which types of workers and entrepreneurs have to be treated the same as a result, the optimal policy manipulates effort incentives

[13] It also implies that the no-discrimination constraints (3) and (11) do not bind at the top of the skill distribution: There does not exist a worker who achieves the same labor income as the highest skill entrepreneurs' profits, since $\tilde{w}\overline{\theta} > w\theta$ for all $\theta \in \Theta$. Hence $a = \underline{\theta}$ and $b = (w/\tilde{w})\overline{\theta}$.

and thus wages to relax these no-discrimination constraints. This then allows for additional redistribution depending on the set of Pareto-weights.

The tax system can increase the workers' relative to the entrepreneurs' wage (i.e. decrease $\tilde{w}/w$) by encouraging entrepreneurial effort and discouraging labor supply. Therefore, and since part (i) has shown that the set of top earners is exclusively given by entrepreneurs and the lowest income is only earned by workers, the optimal tax schedule involves a negative marginal tax rate at the top and a positive marginal tax rate at the bottom in this case. If, by contrast, the Pareto-weights are such that the no-discrimination constraints are relaxed by increasing $\tilde{w}/w$, the opposite pattern holds. Part (iv) in the proposition provides conditions under which these cases occur. Under the natural assumption that the optimal effort schedule for workers $l(\theta)/\theta$ is increasing, it shows that the top marginal tax rate is negative (and the bottom rate positive) whenever the optimum ignoring the no-discrimination constraints would involve

$$v_E(\theta) < v_W\left(\frac{\tilde{w}}{w}\theta\right) \text{ and } \tilde{w}E(\theta) < wl\left(\frac{\tilde{w}}{w}\theta\right) \quad \forall\theta \in [a,b],$$

so that (3) and (11) bind in the $\geq$-direction at the constrained optimum. I will show in Section 4 (Proposition 5) that this is the case if Pareto-weights are such that redistribution from low-ϕ agents to high-ϕ agents is desirable, and thus from entrepreneurs to workers who earn the same overall wage on their effort.

In addition to redistributing across income/profit-levels directly through the tax schedule $T(.)$, the tax system thus makes use of the indirect general equilibrium effects through wages to achieve redistribution indirectly. This shows that optimal marginal tax rates depend on the degree of substitutability between the inputs of the two occupations in the firms' production function. While most of the public finance literature has typically focused on wage elasticities of effort and the skill distribution to derive optimal tax rates (e.g. Saez, 2001), Proposition 2 demonstrates that production elasticities are similarly important when tax policy is restricted to a single schedule.

This intuition is similar, although more intricate, to earlier models of taxation with endogenous wages, notably Stiglitz (1982). He considers a two-class economy where high and low ability workers' labor supply enter a non-linear aggregate production function differently. Then the top marginal tax rate is negative if the government aims at redistributing from high to low skill agents, because subsidizing the high ability individuals' labor supply reduces their wage and thus relaxes the binding incentive constraint preventing high skill agents from imitating low skill agents.[14] In the present occupational

[14] Allen (1982) analyzes optimal *linear* taxation with endogenous wages. In this case, the incentive effects

choice model with two-dimensional heterogeneity, however, the income distributions of entrepreneurs and workers overlap, so that the no-discrimination constraints can bind in either direction. In particular, higher ability workers may have to be prevented from mimicking lower skilled entrepreneurs, but since $\tilde{w} > w$, it is also possible that lower skilled entrepreneurs want to imitate higher ability workers given that the tax system does not condition on occupational choice, even if the Pareto-weights imply redistribution from high to low wage individuals. In fact, as shown by part (iv) of Proposition 2, what matters are the redistributive motives across occupations, and therefore across ϕ-types who earn the same total wage, as implied by the Pareto-weights.

The next proposition contains two results on the effects and desirability of additional tax instruments.

Proposition 3. *(i) If, in addition to the non-linear tax $T(.)$ on profits and income, the government can impose a proportional tax on the firms' labor input, then a Pareto-optimal tax system satisfies $T'(wl(\underline{\theta})) = T'(\tilde{w}E(\bar{\theta})) = 0$.*
(ii) Moreover, if the government can distort $Y_L(L(\theta), E(\theta))$ across firms, e.g. through a non-linear tax on labor input, then it is optimal to do so whenever $Y(L, E)$ is not linear and the no-discrimination constraints (3) and (11) bind for some $\theta \in \Theta$.

Proof. See Appendix A.2. □

The first part of Proposition 3 demonstrates that the properties derived in the last part of Proposition 2 disappear when the government disposes of an additional instrument. With a proportional tax on the firms' labor input, entrepreneurs face a wage cost of τw on their labor rather than the wage w that workers receive. This decouples the scaling factor $\tilde{w}/w$ in the no-discrimination constraints (3) and (11) from the marginal products of entrepreneurial effort and labor in constraints (6) and (7), so that there remains no need to affect them through the nonlinear tax schedule $T(.)$. As a result, the top and bottom marginal tax rates are again zero at any Pareto optimum, even if technology is not linear.

Whereas a pure profit tax, even when complemented by a proportional tax on labor inputs, always implies that marginal products of labor (and thus of entrepreneurial effort) are equalized across all firms, part (ii) shows that such production efficiency is not necessarily optimal in this framework. Intuitively, by distorting marginal products of labor and effort across firms, e.g. through a non-linear tax on the firms' labor input, the government can make the entrepreneurs' wage $\tilde{w}$ vary with skill type. As a result, the rescaling factor $\tilde{w}/w$ in the no-discrimination constraints can also vary with θ, depending

of taxes on wages through the labor supply of different income groups are less clear, since all agents face the same marginal tax rate.

on how much (and in which direction) the no-discrimination constraint binds at that skill level. Then the government faces a trade-off between production efficiency and relaxing the no-discrimination constraints, which generally involves some degree of production inefficiency at the optimum.[15]

4 Differential Tax Treatment of Profits and Income

In this section, I relax the assumption that the government can only impose a single non-linear tax schedule that applies to both labor income and entrepreneurial profits. In contrast, suppose the government is able to condition taxes on occupational choice and thus set different tax schedules $T_y(.)$ for labor income $y \equiv wl$ and $T_\pi(.)$ for profits π. Moreover, suppose the government can use any additional tax instrument that is contingent on observables, such as the firms' outputs or labor inputs. Then the main results compared to the previous section will be that (i) the non-linear tax schedules T_y and T_π are enough to implement the resulting constrained Pareto optima, so that production distortions are no longer desirable, and (ii) redistribution is no longer achieved indirectly through general equilibrium effects, but directly through the tax system. As a result, optimal marginal tax rate formulas for workers and entrepreneurs no longer depend on elasticities of substitution in production.

Beyond contrasting these results with the uniform taxation case, the analysis of differential taxation is also relevant from a policy point of view, since some countries indeed discriminate between the employed and self-employed, as discussed in the introduction. Even if, as is sometimes argued, some individuals are able to shift their incomes between those categories, this does not a priori rule out differential taxation, but can be captured by the elasticity of occupational choice. Indeed, I show in the following that the tax rates crucially depend on this elasticity and that the differential tax treatment vanishes as the occupational choice margin becomes highly elastic.

[15] This is in contrast to the well-known Diamond-Mirrlees Theorem (Diamond and Mirrlees (1971)) in settings without pecuniary externalities. See also Naito (1999) for a related result in the two-class economy introduced by Stiglitz (1982), where production inefficiency is shown to be optimal in an economy with a private and public sector.

4.1 A Theoretical Characterization

4.1.1 Pareto Optimal Tax Formulas

When the planner is not restricted to a single tax schedule on profits and income, the no-discrimination constraints (3) disappear, as do the constraints (6) and (7) that required the equalization of marginal products across all firms. I am therefore left with the following relaxed Pareto problem:

$$\max_{\substack{E(\theta),L(\theta),l(\theta),\\ v_E(\theta),v_W(\theta),\tilde{\phi}(\theta)}} \int_\Theta \left[\tilde{G}_\theta(\tilde{\phi}(\theta))v_E(\theta)+(1-\tilde{G}_\theta(\tilde{\phi}(\theta)))v_w(\theta)\right]d\tilde{F}(\theta)-\int_\Theta\int_{\underline{\phi}}^{\tilde{\phi}(\theta)}\phi d\tilde{G}_\theta(\phi)d\tilde{F}(\theta)$$

$$\text{s.t.}\quad \tilde{\phi}(\theta)=v_E(\theta)-v_W(\theta)\ \ \forall\theta\in\Theta$$

$$v_E'(\theta)=E(\theta)\psi'\left(E(\theta)/\theta\right)/\theta^2,\quad v_W'(\theta)=l(\theta)\psi'\left(l(\theta)/\theta\right)/\theta^2\ \ \forall\theta\in\Theta$$

$$\int_\Theta G_\theta(\tilde{\phi}(\theta))L(\theta)dF(\theta)\le\int_\Theta(1-G_\theta(\tilde{\phi}(\theta)))l(\theta)dF(\theta)$$

$$\int_\Theta G_\theta(\tilde{\phi}(\theta))\left[Y(L(\theta),E(\theta))-v_E(\theta)-\psi(E(\theta)/\theta)\right]dF(\theta) - \int_\Theta(1-G_\theta(\tilde{\phi}(\theta)))\left[v_w(\theta)+\psi(l(\theta)/\theta)\right]dF(\theta)\ \ \ge\ \ 0$$

Clearly, the remaining incentive, labor market clearing and resource constraints are the same as before. It can be seen from this formulation that the wages $\tilde{w}$ and w have now dropped out of the planning problem. In other words, the pecuniary externality that resulted from ruling out differential tax treatment in the previous section has disappeared. This leads to the following proposition characterizing the Pareto-optimal tax policy.

Proposition 4. *(i) At any Pareto optimum, $Y_L(L(\theta),E(\theta))$ is equalized across all $\theta\in\Theta$.*
(ii) If there is no bunching, $T_\pi'(\pi(\theta))$ and $T_y'(y(\theta))$ satisfy

$$\frac{T_\pi'(\pi(\theta))}{1-T_\pi'(\pi(\theta))}=\frac{1+1/\varepsilon_\pi(\theta)}{\theta f(\theta)G_\theta(\tilde{\phi}(\theta))}\int_{\underline{\theta}}^{\theta}\left[\tilde{G}_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\tilde{f}(\hat{\theta})-G_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})+g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\Delta T(\hat{\theta})f(\hat{\theta})\right]d\hat{\theta}$$

$$\frac{T_y'(y(\theta))}{1-T_y'(y(\theta))}=\frac{1+1/\varepsilon_y(\theta)}{\theta f(\theta)(1-G_\theta(\tilde{\phi}(\theta)))}\int_{\underline{\theta}}^{\theta}\left[(1-\tilde{G}_{\hat{\theta}}(\tilde{\phi}(\hat{\theta})))\tilde{f}(\hat{\theta})-(1-G_{\hat{\theta}}(\tilde{\phi}(\hat{\theta})))f(\hat{\theta})-g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\Delta T(\hat{\theta})f(\hat{\theta})\right]d\hat{\theta}$$

with $\Delta T(\theta)\equiv T_\pi(\pi(\theta))-T_y(y(\theta))$.
(iii) $T_\pi'(\pi(\underline{\theta}))=T_\pi'(\pi(\bar{\theta}))=T_y'(y(\underline{\theta}))=T_y'(y(\bar{\theta}))=0$.

Proof. See Appendix A.3. □

Proposition 4 shows first that, when allowing for different tax schedules T_π and T_y, production efficiency is always optimal, since the marginal products of labor and entrepreneurial effort are equalized across all firms. Thus, the non-linear profit and income taxes are actually sufficient to implement any Pareto optimum: No additional tax instruments distorting the firms' input choices are required.[16]

Part (ii) of the proposition derives formulas for the optimal marginal profit and income tax rates. As usual, the optimal marginal tax rate faced by skill type θ is negatively related to the elasticity of profits (income) with respect to the after-tax wage

$$\varepsilon_\pi(\theta) \equiv \frac{\partial \pi(\theta)}{\partial \tilde{w}(1 - T'_\pi(\pi(\theta)))} \frac{\tilde{w}(1 - T'_\pi(\pi(\theta)))}{\pi(\theta)}$$

(and analogously for income) and the mass of entrepreneurs $f(\theta)G_\theta(\tilde{\phi}(\theta))$ at θ (this mass is $f(\theta)(1 - G(\tilde{\phi}(\theta)))$ for workers). This accounts for the local effort (labor supply) distortion generated by the marginal tax. The first two terms in the integral, in turn, capture the redistributive effects of the tax schedule, comparing the mass of Pareto-weights $\tilde{G}_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\tilde{f}(\hat{\theta})$ for all skill types $\hat{\theta}$ below θ to that of the population densities $G_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})$ (and again equivalently for workers). The last term in the integrals, finally, captures the effect of differential profit and labor income taxation on occupational choice. Specifically, the mass of agents of skill θ driven out of entrepreneurship by an infinitesimal increase in profit taxation T_π is given by $g_\theta(\tilde{\phi}(\theta))f(\theta)$, i.e. those individuals who were just indifferent between entrepreneurship and employment before the change. The resulting effect on the government budget is captured by the excess entrepreneurial tax $\Delta T(\theta)$, which is the additional tax payment by an entrepreneur of type θ compared to a worker of the same skill. Of course, this budget effect appears with opposite signs in the optimality formulas for the entrepreneurial profit and labor income tax schedule.[17]

As can be seen from the formulas in Proposition 4, key properties of the restricted tax schedule characterized in the preceding section disappear as soon as differential taxation is allowed. Notably, the tax formulas no longer depend on whether technology is linear or not. Hence, no knowledge about empirical substitution elasticities in production is required to derive optimal marginal tax rates. Differential taxation thus justifies the focus

[16] In a response to the results by Naito (1999), Saez (2004) has argued that the optimality of production inefficiency disappears when the individuals' decision is not along an intensive (effort) margin, but along an extensive (occupational choice) margin. The present model includes both margins, and points out that it is the availability of tax instruments that is crucial for whether there exists a pecuniary externality, which in turn is the underlying reason for the desirability of production distortions.

[17] See Kleven, Kreiner, and Saez (2009) for similar results and interpretations in a model with a secondary earner participation margin. Rather than tracing out the Pareto-frontier, however, they work with a concave social welfare function, which gives rise to different optimal tax formulas.

of much of the public finance literature on estimating labor supply elasticities and identifying skill distributions, quite in contrast to the case of uniform taxation considered in the preceding section.

In fact, the wages $\tilde{w}$ and w earned by entrepreneurs and workers do not even appear in the formulas. Moreover, the bottom and top marginal tax rates are always zero, both for workers and entrepreneurs. In the present setting with a bounded support of the skill distribution, these results show that differential taxation generally allows for a Pareto improvement compared to uniform taxation: Since any Pareto optimum with differential taxation must be such that the bottom and top marginal tax rates for both workers and entrepreneurs are zero, any allocation that does not satisfy these properties must be Pareto inefficient. But Proposition 2 has shown that, whenever uniform taxation leads to binding no-discrimination constraints, the bottom and top marginal tax rates are not zero. Hence, starting from such an allocation, there must exist a Pareto improvement using differential taxation.

The following result is an immediate corollary of Proposition 4.

Corollary 1. *With a constant elasticity ε,[18] the average marginal tax across occupations satisfies*

$$G_\theta(\tilde{\phi}(\theta))\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))} + (1-G_\theta(\tilde{\phi}))\,\frac{T'_y(y(\theta))}{1-T'_y(y(\theta))} = \frac{1+1/\varepsilon}{\theta f(\theta)}\left(\tilde{F}(\theta)-F(\theta)\right). \qquad (14)$$

Note that the formula for the average marginal tax rate across entrepreneurs and workers of a given skill type is given in closed form on the right-hand side of equation (14): It only depends on the elasticity parameter ε, the distribution of skill types as captured by $f(\theta)$ and $F(\theta)$, and the redistributive motives of the government in the skill dimension, determined by the cumulative Pareto-weights $\tilde{F}(\theta)$. In particular, the distribution of cost types ϕ, or redistributive motives in the cost dimension as captured by the Pareto-weights $\tilde{G}_\theta(\phi)$, play no role. This implies a separation result for the implementation of Pareto optima: Average marginal taxes across occupational groups are set so as to achieve the desired redistribution in the skill dimension. Then any redistribution across cost types and hence between entrepreneurs and workers of the same skill is achieved by varying the marginal profit and income taxes, leaving the average tax unaffected. In fact, the formula for a Pareto optimal average marginal tax rate in (14) is the same as the one

[18] Even without a constant elasticity, a modified version of (14) holds, which is that

$$\frac{G_\theta(\tilde{\phi}(\theta))}{1+1/\varepsilon_\pi(\theta)}\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))} + \frac{(1-G_\theta(\tilde{\phi}))}{1+1/\varepsilon_y(\theta)}\frac{T'_y(y(\theta))}{1-T'_y(y(\theta))} = \frac{\tilde{F}(\theta)-F(\theta)}{\theta f(\theta)}.$$

Thus, except for the nicer expression, Corollary 1 does not depend on a constant elasticity.

that would be obtained in a standard quasi-linear Mirrlees-model without occupational choice and with only one-dimensional heterogeneity in θ.[19]

4.1.2 Testing the Pareto Efficiency of Tax Schedules

Rather than determining the optimal shape of tax schedules for a given specification of Pareto-weights, the results in Proposition 4 can also be used as a test for whether some given tax schedules T_π and T_y are Pareto optimal. This approach has been pursued by Werning (2007) in the standard Mirrlees model, and provides an interesting reintepretation of the formulas in Proposition 4 in the present framework. In fact, since the Pareto-weights $\tilde{G}_\theta(\tilde{\phi}(\theta))\tilde{f}(\theta)$ and $(1-\tilde{G}_\theta(\tilde{\phi}(\theta)))\tilde{f}(\theta)$ must be non-negative, the following corollary can be obtained immediately from Proposition 4:

Corollary 2. *Given the utility function $u(c,e) = c - e^{1+1/\varepsilon}/(1+1/\varepsilon)$, a skill distribution $F(\theta)$ and cost distribution $G_\theta(\phi)$, the tax schedules T_π, T_y inducing an allocation $(\pi(\theta), y(\theta))$ and occupational choice $\tilde{\phi}(\theta)$ are Pareto optimal if and only if*

$$\frac{\theta f_E(\theta)}{1+1/\varepsilon}\,\frac{T_\pi'(\pi(\theta))}{1-T_\pi'(\pi(\theta))} + F_E(\theta) - \int_{\underline{\theta}}^{\theta} \frac{g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})}{\overline{G}}\Delta T(\hat{\theta})d\hat{\theta} \quad \text{and} \tag{15}$$

$$\frac{\theta f_W(\theta)}{1+1/\varepsilon}\,\frac{T_y'(y(\theta))}{1-T_y'(y(\theta))} + F_W(\theta) + \int_{\underline{\theta}}^{\theta} \frac{g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})}{1-\overline{G}}\Delta T(\hat{\theta})d\hat{\theta} \tag{16}$$

are non-decreasing in θ, where $\overline{G} \equiv \int_\Theta G_\theta(\tilde{\phi}(\theta))dF(\theta)$ is the overall share of entrepreneurs in the population, $f_E(\theta) \equiv G_\theta(\tilde{\phi}(\theta))f(\theta)/\overline{G}$ and $f_W(\theta) \equiv (1-G_\theta(\tilde{\phi}(\theta)))f(\theta)/(1-\overline{G})$ are the skill densities for entrepreneurs and workers, and $F_E(\theta)$ and $F_W(\theta)$ the corresponding cumulative distribution functions.

For a given elasticity ε, conditions (15) and (16) can be tested after identifying the skill and cost distributions from the observed income distributions and shares of entrepreneurs and workers for a given tax system. This identification step has been pioneered by Saez (2001) in a one-dimensional taxation model, and will be extended in Section 4.2 to the setting with two-dimensional heterogeneity and occupational choice considered here.

Two remarks on Corollary 2 are in order. First, adding conditions (15) and (16) yields another test for Pareto optimality, which is weaker but requires less information to be

[19] See Diamond (1998) for such an analysis. However, since in his model redistribution is determined by a concave social welfare function rather than by Pareto-weights that trace out the entire Pareto-frontier, a closed form solution for the optimal marginal tax rates as in (14) cannot be obtained.

implemented. In particular, a necessary condition for T_π, T_y to be Pareto optimal is that

$$\frac{\theta f(\theta)}{1+1/\varepsilon}\left[G_\theta(\tilde{\phi}(\theta))\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))}+(1-G_\theta(\tilde{\phi}(\theta)))\frac{T'_y(y(\theta))}{1-T'_y(y(\theta))}\right]+F(\theta)$$

is non-decreasing in θ. This condition, relying on the average marginal tax rate of entrepreneurs and workers at a given skill level, only requires the identification of the skill distribution $F(\theta)$, not of the cost density $g_\theta(\phi)$ (note that $G_\theta(\tilde{\phi}(\theta))$ can be easily inferred from the share of entrepreneurs at a given profit and hence skill level). However, this test is obviously weaker since some tax systems that pass it may fail the test in Corollary 2 and thus be Pareto inefficient.

Another special case of conditions (15) and (16) occurs when there is no occupational choice, so that whether an individual is an entrepreneur or a worker is a fixed characteristic. This can be thought of as a special case of the general formulation considered so far, where the cost ϕ has a degenerate distribution with only two mass points, at 0 and $\overline{\phi}$, and $\overline{\phi}$ is sufficiently high. Then T_π must be such that

$$\frac{\theta f_E(\theta)}{1+1/\varepsilon}\,\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))}+F_E(\theta)$$

is non-decreasing, and analogously for T_y replacing the (fixed) skill distribution for entrepreneurs by that for workers, $F_W(\theta)$. This coincides with the integral version of the condition derived in Werning (2007) for a standard Mirrlees model. Hence, the key difference arising from the present framework are the terms $-\int_{\underline{\theta}}^{\theta} g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})\Delta T(\hat{\theta})d\hat{\theta}/\overline{G}$ and $\int_{\underline{\theta}}^{\theta} g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})\Delta T(\hat{\theta})d\hat{\theta}/(1-\overline{G})$, reflecting the effects of taxation on occupational choice and thus on the resource constraint. Note that, since these terms enter conditions (15) and (16) with opposite signs, whenever one term is increasing in θ, the other is decreasing, so that ceteris paribus it becomes harder for differential taxation with $\Delta T(\theta) \neq 0$ to pass the test for Pareto efficiency the more elastic the occupational choice margin (and thus the higher the cost density at the critical level $\tilde{\phi}(\theta)$). Corollary 2 thus demonstrates that the differential tax treatment disappears at Pareto optima as the elasticity of occupational choice increases (for example as it becomes easier for individuals to shift their income and if ϕ is interpreted as the associated cost).

4.1.3 Comparing Optimal Profit and Income Tax Schedules

How do the optimal tax schedules for entrepreneurial profits and labor income compare under given redistributive objectives and thus Pareto-weights? To shed light on this question, I make the following two assumptions.

Assumption 1. *θ and ϕ are independent and $g(\phi)$ is non-increasing.*

These assumptions are strong, and will be relaxed in the numerical explorations that follow in Section 4.2. Nonetheless, they allow me to obtain a theoretical characterization of the pattern of differential taxation of profits and income. I start with the case where the government aims at redistributing from entrepreneurs to workers.

Proposition 5. *Suppose that $\tilde{F}(\theta) = F(\theta)$, $\tilde{g}(\phi) < g(\phi)$ for all $\phi \leq \tilde{\phi}(\overline{\theta})$ and Assumption 1 holds. Then*
(i) $T_y'(y(\theta)) < 0$, $T_\pi'(\pi(\theta)) > 0$ for all $\theta \in \Theta$,
(ii) $\Delta T(\theta) > 0$ and $\Delta T'(\theta) > 0$ for all $\theta \in \Theta$,
(iii) compared to the no tax equilibrium, w decreases and $\tilde{w}$ and $L(\theta)/E(\theta)$ increase for all $\theta \in \Theta$,
(iv) $\tilde{w}E(\theta) < wl((\tilde{w}/w)\theta)$ and $v_E(\theta) < v_W((\tilde{w}/w)\theta)$ for all $\theta \in [a, b]$.

Proof. See Appendix A.4. □

The assumptions in Proposition 5 focus on the benchmark case where the government does not aim at redistributing across skill types (since $\tilde{F}(\theta) = F(\theta)$ for all θ), but puts a lower social welfare weight on low ϕ-types (who end up as entrepreneurs) than their density in the population. This generates a redistributive motive from low to high cost types, and thus from entrepreneurs to workers. Corollary 1 immediately implies that, in this case, the average marginal tax rate must be zero for all skill types. The first part of Proposition 5 shows that, in fact, workers face a negative marginal tax rate and entrepreneurs a positive one at the optimum. Moreover, as a result of the redistributive motive from entrepreneurs to workers, there is a strictly positive excess profits tax $\Delta T(\theta)$, which increases with the skill level.

It also turns out that the optimal policy involves a decrease in the workers' wage, and makes the input mix of all firms more labor intensive compared to the no tax equilibrium. This is quite in contrast to the intuition based on a "trickle down" argument, which would have suggested a policy that increases the workers' wage in order to benefit them indirectly. Here, however, this is not necessary since workers can be overcompensated for the decrease in their wage through the differential tax treatment directly, as captured by the positive excess tax on entrepreneurs. The reason for the depressed wage w is that the excess profit tax discourages entry into entrepreneurship, and therefore the workers' wage must fall so that each firm hires more labor and the labor market remains cleared.

Finally, part (iv) shows that the optimal differential tax policy involves a lower income and utility for entrepreneurs compared to workers who earn the same total wage on their effort. This implies that, under Assumption 1 and the conditions on Pareto-weights in Proposition 5, the no-discrimination constraints (3) and (11) in the previous Section 3 all bind in the same direction and such that the optimal uniform tax schedule involves a positive bottom and a negative top marginal tax rate (see Proposition 2 part (iv)).

If the Pareto-weights are such that $\tilde{F}(\theta) \neq F(\theta)$ for some θ, so that redistribution across skill types is also desirable, then a comparison of the tax schedules for entrepreneurs and workers becomes more involved. A theoretical result is available for the following benchmark case. Suppose that $\tilde{G}(\phi) = G(\phi)$ for all $\theta \in \Theta$, but $\tilde{F}(\theta) \neq F(\theta)$. Also, suppose there is no occupational choice margin, but each individual's occupation is in fact fixed and independent of the skill type, so that $G_\theta = \overline{G}$ for all $\theta \in \Theta$. Then Proposition 4 implies

$$\frac{T'_\pi(\pi(\theta))}{1 - T'_\pi(\pi(\theta))} = \frac{T'_y(y(\theta))}{1 - T'_y(y(\theta))} = \frac{1 + 1/\varepsilon}{\theta f(\theta)} \left(\tilde{F}(\theta) - F(\theta)\right) \tag{17}$$

for any $w, \tilde{w}$. Hence, when the occupational choice margin is removed, the optimal marginal tax rates are the same for entrepreneurs and workers (and equal to the average marginal tax rate from Corollary 1), independently of the different wages in the two occupations. This makes clear that any difference in the optimal tax schedules for profits and income must be the result of an active occupational choice margin or a non-zero correlation between ability and occupational choice, which will be further explored in the following numerical simulations.[20]

4.2 A Quantitative Exploration

To further explore the importance of a differential tax treatment of profits and income, I provide a quantitative illustration of the analysis so far by computing optimal tax systems under various redistributive objectives. Notably, the formulas in Proposition 4 can be used to compute the tax schedules T_π and T_y once distributions for θ and ϕ, Pareto-weights, a production function and preferences are specified.[21] To calibrate the model, I use data on income, profits, and entrepreneurship from the 2007 Survey of Consumer Finances (SCF). I restrict the sample to household heads aged between 18 and 65 who are not unemployed/retired, and define the empirical counterpart of entrepreneurs in

[20] With fixed occupational choice and a share of entrepreneurs G_θ that is correlated with skills, the planner would want to redistribute between the two groups whenever the total welfare weight on entrepreneurs, given by $\int_\Theta G_\theta d\tilde{F}(\theta)$, is not equal to their population share $\int_\Theta G_\theta dF(\theta)$. Since such redistribution can be achieved without distortions through (unbounded) lump-sum taxes and transfers in this case, the optimal tax schedules would not be well-defined.

[21] I use an iterative numerical procedure that is adapted from Kleven, Kreiner, and Saez (2009) and specified in Appendix B.

my model as those individuals who (i) are self-employed, (ii) own a business, (iii) actively manage it, and (iv) employ at least two employees. This is a widely used empirical definition of the notion of entrepreneurship.[22] All other individuals in the sample are considered as workers.

Table 1 presents descriptive statistics of the resulting sample. A share of 6.9% of the sample ends up being classified as entrepreneurs according to the above criteria. Consistent with the theoretical findings so far, entrepreneurs have higher incomes than workers, even though the higher means come at the price of a higher income variability than for workers, as captured by the standard deviation. This suggests that entrepreneurship is more risky than employment, an aspect that will be accounted for explicitly in the next section. Entrepreneurs also work more than workers, as measured by yearly hours. Still, their wage, computed as the ratio of yearly income and hours, is higher than that of workers. This is consistent with the evidence on entrepreneurial incomes and wages in De Nardi, Doctor, and Krane (2007) and Berglann, Moen, Roed, and Skogstrom (2009).[23]

Table 1: Descriptive Statistics

	Entrepreneurs		Workers	
	Mean	St. Dev.	Mean	St. Dev.
Age	48.4	10.2	42.1	11.6
Yearly Income (in 1000$)	88.5	234.7	69.5	128.3
Hours per Week	48.3	14.1	43.4	10.5
Weeks per Year	50.2	6.0	50.4	5.7
Wage per Hour (in $)	55.5	243.8	34.6	124.9

For the baseline calibration, I work with the following parametric specifications: The disutility of effort takes the iso-elastic form $\psi(e) = e^{1+1/\varepsilon}/(1+1/\varepsilon)$, and, based on the empirical labor supply literature, the wage elasticity of effort is set to be $\varepsilon = .5$. The constant returns to scale technology used by entrepreneurs is captured by a Cobb-Douglas production function $Y(L, E) = L^{\alpha}E^{1-\alpha}$ with the parameter α set to equal the workers' share of income in the SCF data, so that $\alpha = .9$.

[22] See e.g. Cagetti and De Nardi (2006) for a discussion. Alternatively, Gentry and Hubbard (2000) only use business ownership to define entrepreneurs, whereas Evans and Jovanovic (1989), Hamilton (2000) and Blanchflower (2004) only focus on self-employment. Yet another distinction is chosen by Holtz-Eakin, Joulfaian, and Rosen (1994a) and Holtz-Eakin, Joulfaian, and Rosen (1994b), who use Schedule C in federal income tax returns to define entrepreneurs.

[23] In contrast, Hamilton (2000) and Blanchflower (2004) find lower returns to entrepreneurship than employment, but their definition of entrepreneurship is only based on self-employment and thus less restrictive than the concept used here.

To identify the skill distribution $F(\theta)$, I use the empirical income distributions of entrepreneurs and workers in the SCF. However, since the SCF does not include information on marginal tax rates faced by individuals, which is required to perform the identification step, I impute marginal tax rates as follows. I adopt the flexible functional form for average taxes $\tau(y)$ as a function of profits/income y suggested by Gouveia and Strauss (1994):

$$\tau(y) = b - b\left[sy^p + 1\right]^{-1/p}. \tag{18}$$

The parameters b, s and p are estimated by Cagetti and De Nardi (2009) using PSID data for entrepreneurs and workers separately, with point estimates $b = .26$, $s = .42$ and $p = 1.4$ for entrepreneurs and $b = .32$, $s = .22$ and $p = .76$ for workers. Then I obtain marginal tax rates from the average tax rates in (18). With this information, I am able to identify $w\theta$ for workers and $\tilde{w}\theta$ for entrepreneurs from the first order conditions of the individuals' utility maximization problem

$$1 - T'_\pi(\pi) = \frac{\pi^{1/\varepsilon}}{(\tilde{w}\theta)^{1+1/\varepsilon}} \quad \text{and} \quad 1 - T'_y(y) = \frac{y^{1/\varepsilon}}{(w\theta)^{1+1/\varepsilon}}$$

for entrepreneurs and workers, respectively. Finally, $\tilde{w}$ and w are found such that $\tilde{w}/w$ equals the ratio of the mean wages of entrepreneurs and workers in the SCF data, using the fact that $\tilde{w} = (1-\alpha)\left(\alpha/w\right)^{\frac{\alpha}{1-\alpha}}$ with Cobb-Douglas technology.

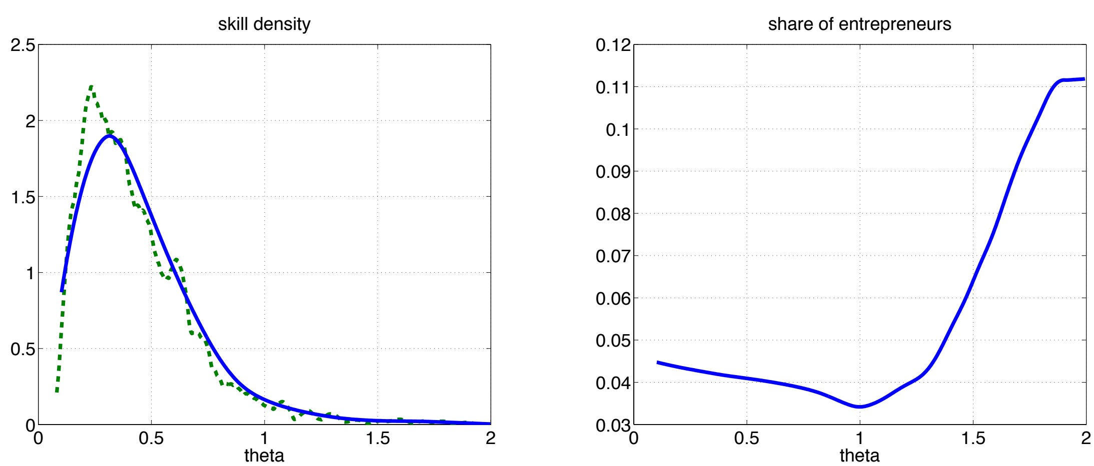

Figure 1: Skill density and share of entrepreneurs

The left panel of Figure 1 depicts a kernel estimate of the resulting inferred skill density. The smoothed approximation of it, also depicted, is used as $f(\theta)$ in the simulations to obtain smoother optimal tax schedules. The right panel in turn shows the share of entrepreneurs as a function of the skill level θ, which is the result of a locally weighted regression of the indicator variable for whether an individual is an entrepreneur on θ. As can be seen from the graph, the share of entrepreneurs is increasing in θ except for low skill levels.[24] I use this pattern to calibrate the cost distribution $G_\theta(\phi)$. In particular, I assume an iso-elastic specification with $G_\theta(\phi) = (\phi/\overline{\phi}_\theta)^\eta$ and $\eta = .5$, implying an elasticity of occupational choice of .5. Then the upper bound of the support $\overline{\phi}_\theta$ is adjusted to generate the pattern of the share of entrepreneurs in the right panel of Figure 1.

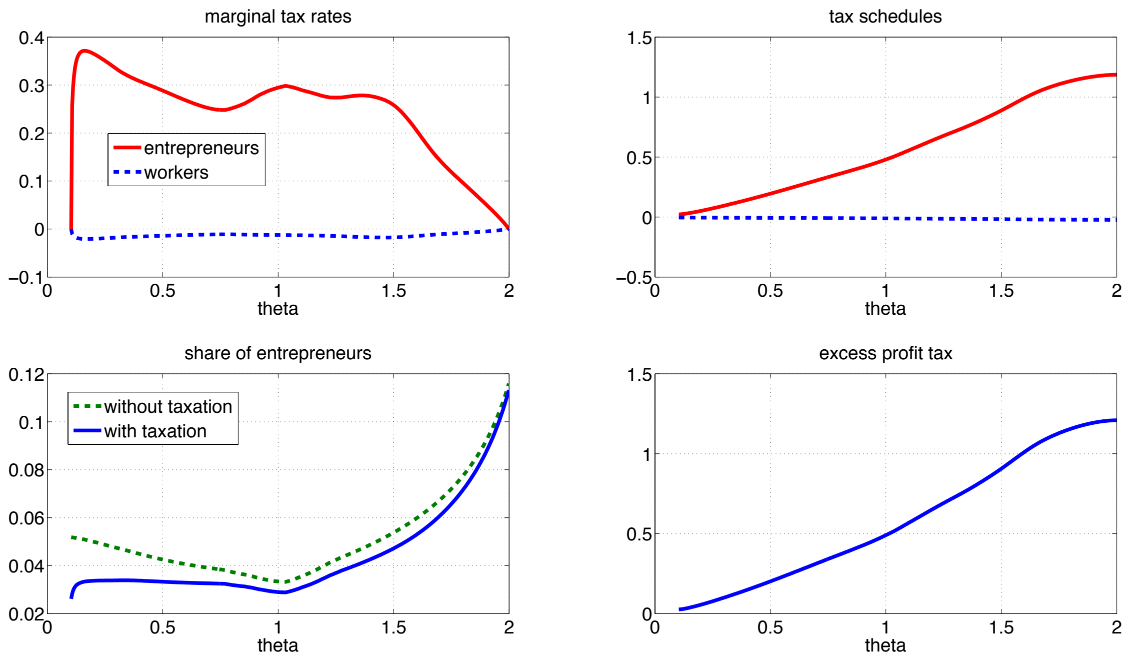

Figure 2: Pareto weights $\tilde{G}_\theta(\phi) = G_\theta(\phi)^{\rho_\Phi}$, $\rho_\Phi = 2$

Figure 2 starts with the case of redistribution across cost types only, and hence from entrepreneurs to workers, with Pareto weights $\tilde{F}(\theta) = F(\theta)$ and $\tilde{G}_\theta(\phi) = G_\theta(\phi)^{\rho_\Phi}$, $\rho_\Phi = 2$, so that $\tilde{G}_\theta(\phi) < G_\theta(\phi)$ for all $\phi \in (0, \overline{\phi}_\theta)$, $\theta \in \Theta$. It depicts the marginal tax schedules T'_π and T'_y, the tax schedules T_π and T_y, the excess profit tax ΔT, and the share of entrepreneurs $G(\tilde{\phi}(\theta))$ as a function of skill, both for the no tax equilibrium as well as for the case with taxation. The figure illustrates the results from Proposition 5: The marginal tax rates for entrepreneurs are positive, for workers negative (even though only slightly due to the large share of workers in the population and the fact that the average marginal tax rate must be zero for all skill levels), and there is a positive and increasing excess profit tax. Entry into entrepreneurship is discouraged for individuals of all skill levels compared to the no tax equilibrium. Moreover, the workers' wage falls by 2.6% from the no tax equilibrium to the equilibrium with taxation.

[24]This U-shaped pattern is in line with the evidence in Parker (1997), who finds that entrepreneurs are over-represented at both the highest and lowest ends of the overall income distribution in the UK.

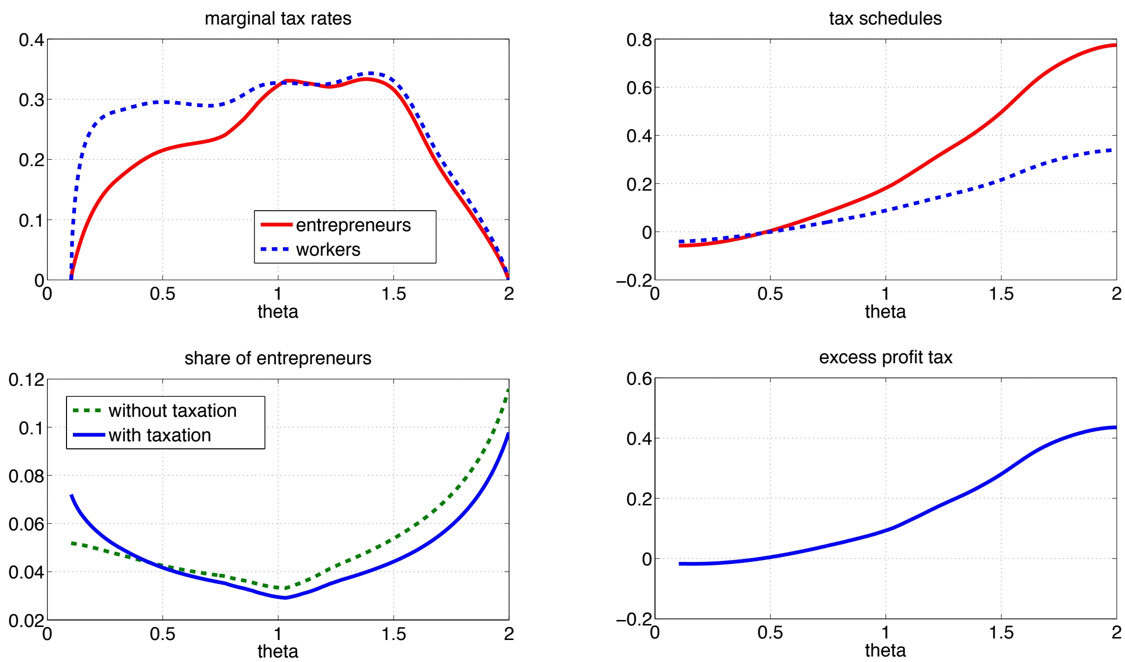

Figure 3: Pareto weights $\tilde{F}(\theta) = 1 - (1 - F(\theta))^{\rho_\Theta}$, $\rho_\Theta = 1.3$

Figure 3 illustrates the other benchmark case, when the Pareto weights are such that redistribution across skill types only is implied. In particular, it assumes $\tilde{F}(\theta) = 1 - (1 - F(\theta))^{\rho_\Theta}$ with $\rho_\Theta = 1.3$, so that $\tilde{F}(\theta) > F(\theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta})$. In this case, both marginal tax schedules are positive (as is the average marginal tax) and such that entrepreneurs and workers of most skill levels face similar marginal tax rates (see the benchmark case in equation (17)), except for lower skill levels, where entrepreneurs face lower marginal tax rates than workers. The excess profit tax remains increasing, but is considerably smaller than in the case where redistribution from entrepreneurs to workers directly is desired, and negative for low skill levels. The wage now remains unchanged compared to the no tax equilibrium and entry into entrepreneurship is now actually encouraged for low skill levels for whom the excess profit tax is negative.

Figure 4 depicts the solution when the planner aims at redistributing in both dimensions of heterogeneity, so that both $\tilde{G}_\theta(\phi) \leq G_\theta(\phi)$ and $\tilde{F}(\theta) \geq F(\theta)$.[25] Such redistributive objectives turn out to justify entrepreneurs facing both higher levels of taxation as well as higher marginal tax rates than workers for all skill levels, while all agents face positive marginal tax rates. Finally, Figure 5 shows a robustness check from reducing the elasticity of effort from $\varepsilon = .5$ to .25. This makes higher tax rates in both occupations optimal, holding Pareto-weights fixed. Again, the workers' wage falls (by 2.6%) as a result of the tax

[25] Such Pareto weights would correspond to the case where the government applies a concave social welfare function $W(.)$ to the total utility $U - \phi$ of individuals.

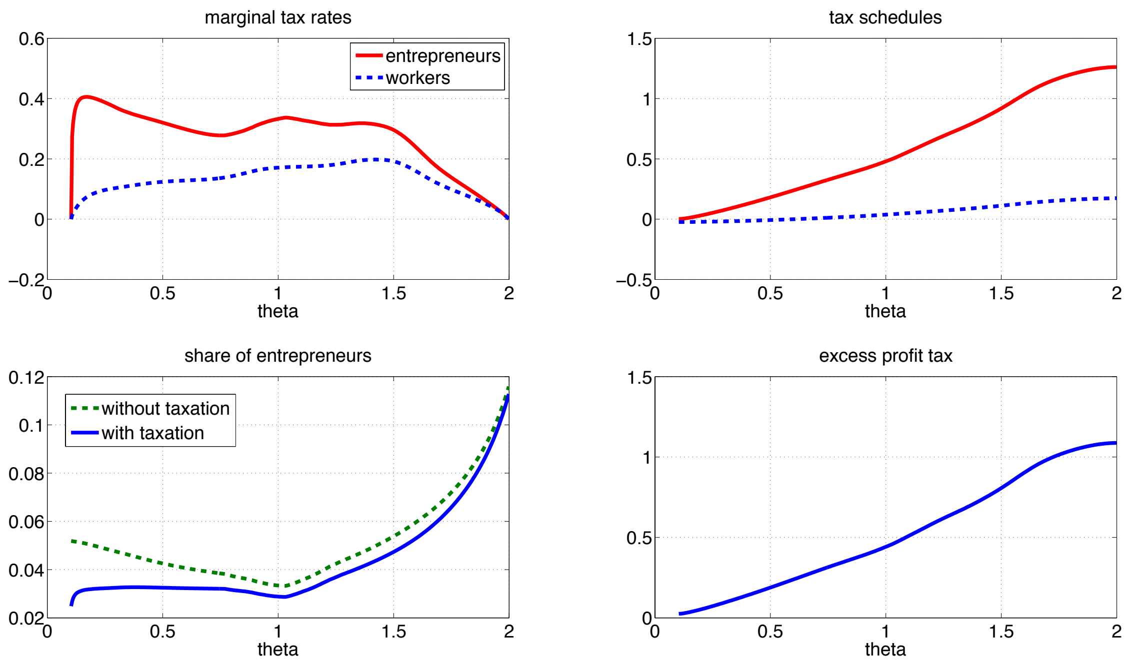

Figure 4: Pareto weights $\rho_\Theta = 1.1$, $\rho_\Phi = 2$

policy compared to the no tax equilibrium, entry into entrepreneurship is discouraged, and there emerges a positive and increasing excess profit tax ΔT.

All these effects appear as robust properties of optimal tax schedules from these quantitative explorations. Notably, even when the Pareto weights imply that workers should be favored, their wage declines, in contrast to "trickle down" based arguments. Moreover, as a general pattern, the numerical simulations suggest that the difference between the optimal marginal tax rates faced by entrepreneurs and workers increases as more redistribution from entrepreneurs to workers directly is aimed at, compared to redistribution across skill types only. In particular, whenever the redistributive preferences of the government put a sufficiently high weight on redistribution from entrepreneurs to workers, in addition to redistribution across skill types, higher marginal tax rates on entrepreneurial incomes compared to labor earnings are justified, as well as higher tax payments as measured by the excess profit tax.[26]

[26] It should also be noted that in all numerical simulations, the monotonicity constraint is satisfied, so that bunching does not occur at the optimum.

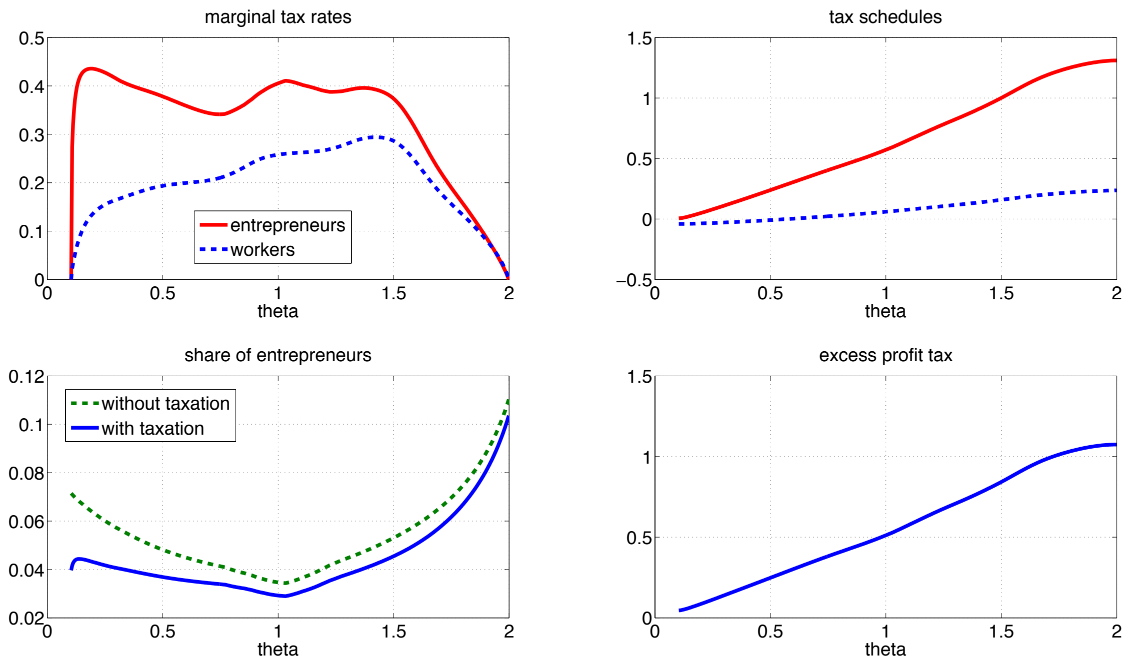

Figure 5: Pareto weights $\rho_\Theta = 1.1$, $\rho_\Phi = 2$, elasticity $\varepsilon = .25$

5 Conclusion

This paper has analyzed the optimal non-linear taxation of profits and labor income in a private information economy with endogenous firm formation. I have demonstrated that it is optimal to apply different non-linear tax schedules on these two forms of income, removing the need for redistribution through indirect, general equilibrium effects and production distortions. In addition, the quantitative importance of differential taxation has been explored in a calibrated model economy.

While these points have been made in a particular even though flexible model, many of the results do not depend on the specific assumptions made. For instance, the model could be interpreted more generally as the optimal income taxation in an economy with different occupations, whose effective wages depend on the relative employment in the sectors through some aggregate production function. In this interpretation, the uniform taxation analysis may be particularly relevant. As Rothschild and Scheuer (2011a) and (2011b) show in ongoing work, rather than assuming that one dimension of heterogeneity captures the individuals' skill level in both sectors while the second enters preferences additively, similar results can be obtained in a standard Roy model where each individual is characterized by a two-dimensional skill type, one for each sector. While the present paper abstracts from the need for entrepreneurs to borrow funds in order to set up a firm, Scheuer (2011) extends the present model to incorporate entrepreneurial borrowing and demonstrates that this provides yet another argument for differential taxation of entrepreneurial profits, by mitigating occupational misallocation that results from credit market frictions.

Both the theoretical and numerical analysis have abstracted from several other aspects of entrepreneurship and its implications for tax policy. Notably, income effects and risk aversion, capital accumulation and additional choices available to entrepreneurs, such as the decision whether to incorporate or not, have been neglected in this paper. In addition, the role of entrepreneurs in fostering technological innovations and economic growth may generate yet other roles for entrepreneurial taxation, given that these activities are typically associated with externalities. Extensions of the present results to a more comprehensive exploration of these issues are left for the future.

References

ALBANESI, S. (2006): "Optimal Taxation of Entrepreneurial Capital with Private Information," *NBER Working Paper No. 12419*.

——— (2008): "Selection and Incentives: Optimal Taxation with Occupational Choice and Private Information," *Working Paper, Columbia University*.

ALLEN, F. (1982): "Optimal Linear Income Taxation with General Equilibrium Effects on Wages," *Journal of Public Economics*, 17, 135–143.

BERGLANN, H., E. MOEN, K. ROED, AND J. SKOGSTROM (2009): "Entrepreneurship: Origins and Returns," *IZA Discussion Paper No. 4250*.

BLANCHFLOWER, D. (2004): "Self-Employment: More May Not Be Better," *NBER Working Paper No. 10286*.

BOADWAY, R., N. MARCEAU, AND P. PESTIEAU (1991): "Optimal Linear Income Taxation in Models with Occupational Choice," *Journal of Public Economics*, 46, 133–162.

CAGETTI, M., AND M. DE NARDI (2006): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 114, 835–870.

——— (2009): "Estate Taxation, Entrepreneurship, and Wealth," *American Economic Review*, 99, 85–111.

CHARI, V., M. GOLOSOV, AND A. TSYVINSKI (2002): "Business Start-ups, the Lock-in Effect, and Capital Gains Taxation," *Mimeo*.

CHRISTIANSEN, V. (1990): "Choice of Occupation, Tax Incidence and Piecemal Tax Revision," *Scandinavian Journal of Economics*, 90, 141–160.

CRAWFORD, C., AND J. FREEDMAN (2010): "Small Business Taxation," in *The Mirrlees Review – Dimensions of Tax Design*. Institute for Fiscal Studies.

CULLEN, J. B., AND R. GORDON (2007): "Taxes and Entrepreneurial Risk Taking: Theory and Evidence for the US," *Journal of Public Economics*, 91, 1479–1505.

DE NARDI, M., P. DOCTOR, AND S. KRANE (2007): "Evidence on Entrepreneurs in the United States: Data from the 1989 to 2004 Survey of Consumer Finances," *Economic Perspectives, Federal Reserve Bank of Chicago*, Q IV, 18–36.

DIAMOND, P. (1998): "Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Tax Rates," *American Economic Review*, 88, 83–95.

DIAMOND, P., AND J. MIRRLEES (1971): "Optimal Taxation in Public Production, I: Production Efficiency," *American Economic Review*, 61, 8–27.

EVANS, D., AND B. JOVANOVIC (1989): "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy*, 97, 808–827.

FELDSTEIN, M. (1973): "On the Optimal Progressivity of the Income Tax," *Journal of Public Economics*, 2, 356–376.

FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. MIT Press, Cambridge.

GENTRY, W., AND R. HUBBARD (2000): "Tax policy and entrepreneurial entry," *American Economic Review*, 90, 283–287.

GORDON, R. (1985): "Taxation of Corporate Capital Income: Tax Revenues Versus Tax Distortions.," *Quarterly Journal of Economics*, 100, 1–26.

GOUVEIA, M., AND R. STRAUSS (1994): "Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis," *National Tax Journal*, 47, 317–339.

GRAVELLE, J., AND L. KOTLIKOFF (1989): "The Incidence and Efficiency Costs of Corporate Taxation when Corporate and Noncorporate Firms Produce the same Good," *Journal of Political Economy*, 97, 749–780.

HAMILTON, B. (2000): "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment," *Journal of Political Economy*, 108, 604–631.

HOLTZ-EAKIN, D., D. JOULFAIAN, AND H. ROSEN (1994a): "Entrepreneurial Decisions and Liquidity Constraints," *RAND Journal of Economics*, 25, 334–347.

——— (1994b): "Sticking It Out: Entrepreneurial Survival and Liquidity Constraints," *Journal of Political Economy*, 102, 53–75.

KANBUR, S. (1981): "Risk Taking and Taxation: An Alternative Perspective," *Journal of Public Economics*, 15, 163–184.

KIHLSTROM, R., AND J. LAFFONT (1979): "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy*, 87, 719–748.

——— (1983): "Taxation and Risk Taking in General Equilibrium Models with Free Entry," *Journal of Public Economics*, 21, 159–181.

KITAO, S. (2008): "Entrepreneurship, Taxation, and Capital Investment," *Review of Economic Dynamics*, 11, 44–69.

KLEVEN, H. J., C. T. KREINER, AND E. SAEZ (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77, 537–560.

MEH, C. (2005): "Entrepreneurship, Wealth Inequality, and Taxation," *Review of Economic Dynamics*, 8, 688–719.

——— (2008): "Business Risk, Credit Constraints, and Corporate Taxation," *Journal of Economic Dynamics and Control*, 32, 2971–3008.

MEH, C., AND V. QUADRINI (2004): "Endogenous Market Incompleteness with Investment Risks," *Working Paper, University of Southern California*.

MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.

MORESI, S. (1997): "Optimal Taxation and Firm Formation: A Model of Asymmetric Information," *European Economic Review*, 42, 1525–1551.

NAITO, H. (1999): "Re-Examination of Uniform Commodity Taxes Under A Non-Linear Income Tax System And Its Implication for Production Efficiency," *Journal of Public Economics*, 71, 165–188.

PANOUSI, V. (2008): "Capital Taxation with Entrepreneurial Risk," *Mimeo, MIT*.

PARKER, S. C. (1997): "The Distribution of Self-Employment Income in the United Kingdom, 1976-1991," *Economic Journal*, 107, 455–466.

——— (1999): "The Optimal Linear Taxation of Employment and Self-Employment Incomes," *Journal of Public Economics*, 73, 107–123.

QUADRINI, V. (2000): "Entrepreneurship, Saving, and Social Mobility," *Review of Economic Dynamics*, 3, 1–40.

ROTHSCHILD, C., AND F. SCHEUER (2011a): "Optimal Taxation with Rent-Seeking," *NBER Working Paper 17035*.

——— (2011b): "Redistributive Taxation in a Roy Model," *Mimeo, Wellesley College and Stanford University*.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Tax Rates," *Review of Economic Studies*, 68, 205–229.

——— (2004): "Direct or Indirect Tax Instruments for Redistributiopn: Short-run versus Long-run," *Journal of Public Economics*, 88, 503–518.

SCHEUER, F. (2011): "Adverse Selection in Credit Markets and Regressive Profit Taxation," *Mimeo, Stanford University*.

SHOURIDEH, A. (2010): "Optimal Taxation of Entrepreneurial Income," *Mimeo, University of Minnesota*.

STIGLITZ, J. (1982): "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 17, 213–240.

TOPKIS, D. (1998): *Supermodularity and Complementarity*. Princeton University Press, Princeton, USA.

WERNING, I. (2007): "Pareto-Efficient Income Taxation," *Mimeo, MIT*.

ZECKHAUSER, R. (1977): "Taxes in Fantasy, or Most Any Tax on Labor Can Turn Out to Help the Laborers," *Journal of Public Economics*, 8, 133–150.

ZUBRICKY, J. (2007): "Flat Tax Reform in an Economy with Occupational Choice and Financial Frictions," *Mimeo*.

A Proofs for Sections 3 and 4

A.1 Proof of Proposition 2

(i) Analogously to the proof of Proposition 1, if $\tilde{w} \leq w$, then $\tilde{\phi}(\theta) = 0$ for all $\theta \in \Theta$ (with the only additional argument that, since both occupations face the same tax schedule on their profits (resp. income), there is also no tax advantage from entering entrepreneurship). This, together with the fact that (12) and (13) must hold as equalities at an optimum, implies $l(\theta) = E(\theta) = v_E(\theta) = v_W(\theta) = 0$ for all $\theta \in \Theta$. Clearly, the no tax equilibrium characterized in Proposition 1 is Pareto-superior, demonstrating that $\tilde{w} \leq w$ cannot be part of a Pareto-optimum.

To see that $\tilde{w}E(\theta) > wl(\theta)$ $\forall \theta \in \Theta$, define $\pi(\theta) \equiv \tilde{w}E(\theta)$ and $y(\theta) \equiv wl(\theta)$. $\pi(\theta)$ solves $\max_\pi \pi - T(\pi) - \psi(\pi/(\tilde{w}\theta))$, and analogously for $y(\theta)$, replacing $\tilde{w}$ by w. Note that $-\psi(x/(w\theta))$ is supermodular in (x, w). Then the result follows from Topkis' theorem and $\tilde{w} > w$.

(ii) Using the result from (i) that $\tilde{w} > w$, let me recapitulate the Pareto problem as follows:

$$\max_{\substack{\{E(\theta), L(\theta), l(\theta), \\ v_E(\theta), v_W(\theta), \tilde{\phi}(\theta), w, \tilde{w}\}}} \int_\Theta \left[\tilde{G}_\theta(\tilde{\phi}(\theta)) v_E(\theta) - \int_{\underline{\phi}}^{\tilde{\phi}(\theta)} \phi d\tilde{G}_\theta(\phi) + (1 - \tilde{G}_\theta(\tilde{\phi}(\theta))) v_w(\theta) \right] d\tilde{F}(\theta)$$

$$\text{s.t.} \quad \tilde{\phi}(\theta) = v_E(\theta) - v_W(\theta) \;\; \forall \theta \in \Theta$$

$$v_E'(\theta) = E(\theta)\psi'\left(E(\theta)/\theta\right)/\theta^2, \quad v_W'(\theta) = l(\theta)\psi'\left(l(\theta)/\theta\right)/\theta^2 \;\; \forall \theta \in \Theta \tag{IC}$$

$$\int_\Theta G_\theta(\tilde{\phi}(\theta)) L(\theta) dF(\theta) \leq \int_\Theta (1 - G_\theta(\tilde{\phi}(\theta))) l(\theta) dF(\theta) \tag{LM}$$

$$\begin{aligned} \int_\Theta G_\theta(\tilde{\phi}(\theta)) \left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi(E(\theta)/\theta) \right] dF(\theta) & \\ -(1 - G(\tilde{\phi}(\theta))) \left[v_W(\theta) + \psi(l(\theta)/\theta) \right] dF(\theta) \geq & \quad 0 \end{aligned} \tag{RC}$$

$$v_E(\theta) = v_W\left((\tilde{w}/w)\theta\right), \quad E(\theta) = (w/\tilde{w}) l\left((\tilde{w}/w)\theta\right) \;\; \forall \theta \in \left[\underline{\theta}, (w/\tilde{w})\overline{\theta}\right] \tag{ND}$$

$$w = Y_L(L(\theta), E(\theta)), \quad \tilde{w} = Y_E(L(\theta), E(\theta)) \;\; \forall \theta \in \Theta. \tag{MP}$$

Note that I have dropped the monotonicity constraint (9), assuming that it will not bind at the optimum (and thus ignoring problems of bunching). Attaching multipliers $\mu_E(\theta)$ and $\mu_W(\theta)$ to the incentive constraints (IC), λ_{LM} to the labor market clearing constraint (LM), λ_{RC} to the resource constraint (RC), $\xi_v(\theta)$ and $\xi_E(\theta)$ to the no-discrimination constraints (ND) and $\kappa_L(\theta)$ and $\kappa_E(\theta)$ to the marginal product constraints (MP), the corresponding Lagrangian, after integrating by parts, can be written as

$$\begin{aligned}
\mathcal{L} \quad = \quad & \int_{\Theta}\Bigg[\tilde{G}_{\theta}(\tilde{\phi}(\theta))v_E(\theta) - \int_{\underline{\phi}}^{\tilde{\phi}(\theta)} \phi d\tilde{G}_{\theta}(\phi) + (1-\tilde{G}_{\theta}(\tilde{\phi}(\theta)))v_W(\theta)\Bigg] d\tilde{F}(\theta) \\
& - \int_{\Theta}\left[\mu_E'(\theta)v_E(\theta) + \mu_E(\theta)\psi'\left(\frac{E(\theta)}{\theta}\right)\frac{E(\theta)}{\theta^2}\right]d\theta - \int_{\Theta}\left[\mu_W'(\theta)v_W(\theta) + \mu_W(\theta)\psi'\left(\frac{l(\theta)}{\theta}\right)\frac{l(\theta)}{\theta^2}\right]d\theta \\
& + \lambda_{LM}\left[\int_{\Theta}(1 - G_{\theta}(\tilde{\phi}(\theta)))l(\theta)dF(\theta) - \int_{\Theta} G_{\theta}(\tilde{\phi}(\theta))L(\theta)dF(\theta)\right] \\
& + \lambda_{RC}\Bigg[\int_{\Theta} G_{\theta}(\tilde{\phi}(\theta))\left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi\left(\frac{E(\theta)}{\theta}\right)\right] - (1 - G(\tilde{\phi}(\theta)))\left[v_W(\theta) + \psi\left(\frac{l(\theta)}{\theta}\right)\right] dF(\theta)\Bigg] \\
& + \int_{\underline{\theta}}^{(w/\tilde{w})\bar{\theta}} \xi_v(\theta)\left[v_E(\theta) - v_W\left(\frac{\tilde{w}}{w}\theta\right)\right]d\theta + \int_{\underline{\theta}}^{(w/\tilde{w})\bar{\theta}} \xi_E(\theta)\left[E(\theta) - \frac{w}{\tilde{w}} l\left(\frac{\tilde{w}}{w}\theta\right)\right]d\theta \\
& + \int_{\Theta} \kappa_L(\theta)\left[w - Y_L(L(\theta), E(\theta))\right]d\theta + \int_{\Theta} \kappa_E(\theta)\left[\tilde{w} - Y_E(L(\theta), E(\theta))\right]d\theta. \qquad (19)
\end{aligned}$$

The transversality conditions are $\mu_E(\underline{\theta}) = \mu_E(\bar{\theta}) = \mu_W(\underline{\theta}) = \mu_W(\bar{\theta}) = 0$. Note first that, due to quasi-linear preferences, $\lambda_{RC} = 1$. Then the necessary condition for $L(\theta)$ is

$$G_{\theta}(\tilde{\phi}(\theta))f(\theta)\left[Y_L(L(\theta), E(\theta)) - \lambda_{LM}\right] - \left[\kappa_L(\theta)Y_{LL}(\theta) + \kappa_E(\theta)Y_{EL}(\theta)\right] = 0 \;\; \forall\theta \in \Theta. \qquad (20)$$

Using the transversality conditions, the necessary conditions for $E(\bar{\theta})$ and $l(\underline{\theta})$ are

$$G_{\bar{\theta}}(\tilde{\phi}(\bar{\theta}))f(\bar{\theta})\left[\tilde{w} - \frac{1}{\bar{\theta}}\psi'\left(\frac{E(\bar{\theta})}{\bar{\theta}}\right)\right] - \left[\kappa_L(\bar{\theta})Y_{LE}(\bar{\theta}) + \kappa_E(\bar{\theta})Y_{EE}(\bar{\theta})\right] = 0 \qquad (21)$$

and

$$\lambda_{LM} - \frac{1}{\underline{\theta}}\psi'\left(\frac{l(\underline{\theta})}{\underline{\theta}}\right) = 0. \qquad (22)$$

If $Y(L, E)$ is linear, then $Y_{LL} = Y_{LE} = Y_{EE} = 0$ and thus (20) and (MP) imply $\lambda_{LM} = Y_L(\theta) = w$ for all θ. Therefore, by (21) and (22),

$$\tilde{w} = \frac{1}{\bar{\theta}}\psi'\left(\frac{E(\bar{\theta})}{\bar{\theta}}\right) \quad \text{and} \quad w = \frac{1}{\underline{\theta}}\psi'\left(\frac{l(\underline{\theta})}{\underline{\theta}}\right).$$

Note that the first-order condition for the entrepreneurs' and workers' problem is

$$\tilde{w}(1 - T'(\tilde{w}E)) = \frac{1}{\theta}\psi'\left(\frac{E}{\theta}\right) \quad \text{and} \quad w(1 - T'(wl)) = \frac{1}{\theta}\psi'\left(\frac{l}{\theta}\right),$$

so I obtain $T'(\tilde{w}E(\bar{\theta})) = T'(wl(\underline{\theta})) = 0$ at any Pareto-optimum if technology is linear.

(iii) There are 3 cases to be considered:

Case 1: $\lambda_{LM} = w$.

In this case, (20) together with (MP) implies

$$\kappa_L(\theta)Y_{LL}(\theta) + \kappa_E(\theta)Y_{EL}(\theta) = 0 \;\; \forall\theta \in \Theta.$$

Note that, with constant returns to scale,

$$Y_{LL}(\theta) = -xY_{EL}(\theta) \quad \text{and} \quad Y_{EL}(\theta) = -xY_{EE}(\theta) \;\; \forall \theta \in \Theta, \tag{23}$$

where $x = E(\theta)/L(\theta)$ is independent of θ by (MP). Thus

$$\kappa_L(\theta)Y_{EL}(\theta) + \kappa_E(\theta)Y_{EE}(\theta) = 0 \;\; \forall \theta \in \Theta,$$

and then (21) and (22) imply $T'(\tilde{w}E(\overline{\theta})) = T'(wl(\underline{\theta})) = 0$.
Case 2: $\lambda_{LM} < w$.
Now (20) and (MP) yield

$$\kappa_L(\theta)Y_{LL}(\theta) + \kappa_E(\theta)Y_{EL}(\theta) > 0 \;\; \forall \theta \in \Theta$$

and hence by (23)

$$\kappa_L(\theta)Y_{EL}(\theta) + \kappa_E(\theta)Y_{EE}(\theta) < 0 \;\; \forall \theta \in \Theta.$$

Then (21) and (22) yield $T'(\tilde{w}E(\overline{\theta})) < 0$ and $T'(wl(\underline{\theta})) > 0$.
Case 3: $\lambda_{LM} > w$.
This case is completely analogous to case 2 with all signs reversed.

(iv) To prove the last part of the proposition, it is useful to rewrite the constraints (ND) and (MP) as follows. First, with constant returns to scale, (MP) is equivalent to requiring that

$$x = \frac{E(\theta)}{L(\theta)} \;\Leftrightarrow\; xL(\theta) - E(\theta) = 0 \;\; \forall \theta \in \Theta. \tag{MP'}$$

Using this, we can write (ND) as

$$v_E(\theta) = v_W\left((Y_E(x)/Y_L(x))\theta\right),\; E(\theta) = (Y_L(x)/Y_E(x))l\left((Y_E(x)/Y_L(x))\theta\right) \;\forall \theta \in \left[\underline{\theta}, (Y_L(x)/Y_E(x))\overline{\theta}\right]. \tag{ND'}$$

Then the Pareto problem is as in the proof of part (ii), with the only difference that maximization is over x rather than w and $\tilde{w}$, and (ND) and (MP) are replaced by (ND') and (MP'). Denote the multipliers on (MP') by $\kappa(\theta)$. The necessary conditions for $L(\theta)$ and $E(\overline{\theta})$ become

$$G_\theta(\tilde{\phi}(\theta))f(\theta)[Y_L(x) - \lambda_{LM}] + x\kappa(\theta) = 0 \;\forall \theta \in \Theta, \tag{24}$$

$$G_{\overline{\theta}}(\tilde{\phi}(\overline{\theta}))f(\overline{\theta})\left[\tilde{w} - \frac{1}{\overline{\theta}}\psi'\left(\frac{E(\overline{\theta})}{\overline{\theta}}\right)\right] - \kappa(\overline{\theta}) = 0 \tag{25}$$

and the first-order condition for $l(\underline{\theta})$ remains as in (22). Moreover, the necessary condition for x is

$$\begin{aligned} &\int_a^b \kappa(\theta)L(\theta)d\theta - \frac{d(Y_E(x)/Y_L(x))}{dx}\int_a^b \xi_v(\theta)v_W'\left(\frac{Y_E(x)}{Y_L(x)}\theta\right)\theta d\theta \\ -&\int_a^b \xi_E(\theta)\left[\frac{d(Y_L(x)/Y_E(x))}{dx}l\left(\frac{Y_E(x)}{Y_L(x)}\theta\right) + \frac{d(Y_E(x)/Y_L(x))}{dx}\frac{Y_L(x)}{Y_E(x)}l'\left(\frac{Y_E(x)}{Y_L(x)}\theta\right)\theta\right]d\theta \;=\; 0 \end{aligned} \tag{26}$$

with $a = \underline{\theta}$ and $b = (Y_L(x)/Y_E(x))\overline{\theta}$. Note that $d(Y_L(x)/Y_E(x))/dx = -(Y_L(x)/Y_E(x))^2 d(Y_E(x)/Y_L(x))/dx$, so that (26) can be written as

$$\int_a^b \kappa(\theta)L(\theta)d\theta = \frac{d(Y_E(x)/Y_L(x))}{dx}\left[\int_a^b \xi_v(\theta)v'_W\left(\tilde{\theta}(x,\theta)\right)\theta d\theta \right. \\ \left. +\frac{Y_L(x)^2}{Y_E(x)^2}\int_a^b \xi_E(\theta)\left[\tilde{\theta}(x,\theta)l'\left(\tilde{\theta}(x,\theta)\right) - l\left(\tilde{\theta}(x,\theta)\right)\right]d\theta\right],$$

where I wrote $\tilde{\theta}(x,\theta) \equiv (Y_E(x)/Y_L(x))\theta$ to simplify notation. The RHS can be signed as follows. First, $d(Y_E(x)/Y_L(x))/dx < 0$ by the concavity of Y and constant returns to scale. Next, $v'_W(.) > 0$ by the incentive constraints (IC). Finally, $\tilde{\theta}l'(\tilde{\theta}) - l(\tilde{\theta}) > 0$ if the workers' effort $l(\theta)/\theta$ is increasing in θ, as assumed in the proposition. Hence, $\int_a^b \kappa(\theta)L(\theta)d\theta$ is positive (negative) whenever $\xi_v(\theta)$ and $\xi_E(\theta)$ are negative (positive) for all $\theta \in [a,b]$. Moreover, $\kappa(\theta)$ must have the same sign for all θ by (24), so that the same result must hold for all $\kappa(\theta)$. To determine the sign of the multipliers $\xi_v(\theta)$ and $\xi_E(\theta)$, note that the no-discrimination constraints (ND') are equivalent to imposing both inequality constraints

$$v_E(\theta) \geq v_W\left((Y_E(x)/Y_L(x))\theta\right),\ E(\theta) \geq (Y_L(x)/Y_E(x))l\left((Y_E(x)/Y_L(x))\theta\right) \tag{ND'$_\geq$}$$

and

$$v_E(\theta) \leq v_W\left((Y_E(x)/Y_L(x))\theta\right),\ E(\theta) \leq (Y_L(x)/Y_E(x))l\left((Y_E(x)/Y_L(x))\theta\right). \tag{ND'$_\leq$}$$

We have $\xi_v(\theta), \xi_E(\theta) \geq 0$ whenever (ND'$_\geq$) binds and (ND'$_\leq$) is slack, whereas $\xi_v(\theta), \xi_E(\theta) \leq 0$ if (ND'$_\leq$) is binding and (ND'$_\geq$) is slack. In the first case, all $\kappa(\theta)$ are negative, so that $Y_L(x) > \lambda_{LM}$ by (24). Then the top and bottom marginal tax rate results immediately follow from (22) and (25). In the second case, all signs are reversed.

A.2 Proof of Proposition 3

(i) With a proportional tax on the labor input of firms, entrepreneurs effectively face a wage τw rather than the wage w that workers receive, and hence the Pareto problem is the same as in Proposition 2 with the only difference that maximization is also performed over τ and (MP) is replaced by

$$\tau w = Y_L(L(\theta), E(\theta)), \quad \tilde{w} = Y_E(L(\theta), E(\theta))\ \ \forall \theta \in \Theta. \tag{MP''}$$

The necessary condition for τ yields $\int_\Theta \kappa_L(\theta)d\theta = 0$, and the necessary conditions for w and $\tilde{w}$ are

$$\int_\Theta \kappa_L(\theta)d\theta = \frac{\tilde{w}}{w^2}\int_{\underline{\theta}}^{(w/\tilde{w})\bar{\theta}} \xi_v(\theta)v'_W\left(\frac{\tilde{w}}{w}\theta\right)\theta d\theta + \int_{\underline{\theta}}^{(w/\tilde{w})\bar{\theta}} \xi_E(\theta)\left[\frac{1}{w}l'\left(\frac{\tilde{w}}{w}\theta\right)\theta - \frac{1}{\tilde{w}}l\left(\frac{\tilde{w}}{w}\theta\right)\right]d\theta$$

$$\int_\Theta \kappa_E(\theta)d\theta = -\frac{1}{w}\int_{\underline{\theta}}^{(w/\tilde{w})\bar{\theta}} \xi_v(\theta)v'_W\left(\frac{\tilde{w}}{w}\theta\right)\theta d\theta - \int_{\underline{\theta}}^{(w/\tilde{w})\bar{\theta}} \xi_E(\theta)\left[\frac{1}{\tilde{w}}l'\left(\frac{\tilde{w}}{w}\theta\right)\theta - \frac{w}{\tilde{w}^2}l\left(\frac{\tilde{w}}{w}\theta\right)\right]d\theta,$$

which implies

$$\int_\Theta \kappa_E(\theta)d\theta = -\frac{w}{\tilde{w}}\int_\Theta \kappa_L(\theta)d\theta \tag{27}$$

and hence $\int_\Theta \kappa_E(\theta)d\theta = 0$. To obtain a contradiction, suppose $\lambda_{LM} < \tau w$. Then (20) and (MP'') imply

$$\kappa_L(\theta)Y_{LL}(\theta) + \kappa_E(\theta)Y_{EL}(\theta) > 0\ \ \forall \theta \in \Theta$$

and rearranging yields (since $Y_{LL} < 0$)

$$\kappa_L(\theta) < -\frac{Y_{EL}(\theta)}{Y_{LL}(\theta)}\kappa_E(\theta) = x\kappa_E(\theta) \ \ \forall\theta \in \Theta.$$

Yet this contradicts the above result that $\int_\Theta \kappa_L(\theta)d\theta = \int_\Theta \kappa_E(\theta)d\theta = 0$. Similarly, if $\lambda_{LM} > \tau w$, then $\kappa_L(\theta) > x\kappa_E(\theta)\ \forall\theta \in \Theta$, also yielding a contradiction. Hence, $\lambda_{LM} = \tau w$ must hold at a Pareto optimum. Then $Y_L - \psi'(wl(\underline{\theta}))/\underline{\theta} = Y_E - \psi'(\tilde{w}E(\overline{\theta}))/\overline{\theta} = 0$ and thus $T'(\tilde{w}E(\overline{\theta})) = T'(wl(\underline{\theta})) = 0$ follows from the proof of part (iii) of Proposition 2, case 1.

(ii) If the government can distort the marginal products of labor across firms, e.g. through a non-linear tax on labor inputs, then an entrepreneur of skill θ effectively faces a wage $\tau(\theta)w$, and (MP) is to be replaced by

$$\tau(\theta)w = Y_L(L(\theta), E(\theta)), \quad \tilde{w}(\theta) = Y_E(L(\theta), E(\theta)) \ \ \forall\theta \in \Theta, \tag{MP'''}$$

and (ND) becomes

$$v_E(\theta) = v_W\left((\tilde{w}(\theta)/w)\theta\right), \quad E(\theta) = (w/\tilde{w}(\theta))l\left((\tilde{w}(\theta)/w)\theta\right) \ \ \forall\theta \in \left[\underline{\theta}, (w/\tilde{w}(\theta))\overline{\theta}\right]. \tag{ND''}$$

Now the necessary condition for $\tau(\theta)$ is $\kappa_L(\theta) = 0$ for all $\theta \in \Theta$, and for $\tilde{w}(\theta)$

$$\kappa_E(\theta) = -\frac{1}{w}\xi_v(\theta)v_W'\left(\frac{\tilde{w}}{w}\theta\right)\theta - \xi_E(\theta)\left[\frac{1}{\tilde{w}}l'\left(\frac{\tilde{w}}{w}\theta\right)\theta - \frac{w}{\tilde{w}^2}l\left(\frac{\tilde{w}}{w}\theta\right)\right] \ \ \forall\theta \in \Theta. \tag{28}$$

Note that it must hold that $\tilde{w}(\overline{\theta}) > w$ (since otherwise $\tilde{\phi}(\overline{\theta}) = 0$), and therefore (ND'') does not bind at $\overline{\theta}$, which yields $\kappa_E(\overline{\theta}) = 0$. Then (20) implies (together with $\kappa_L(\overline{\theta}) = 0$) that $Y_L(\overline{\theta}) = \lambda_{LM}$. However, whenever there exists some $\theta < \overline{\theta}$ such that (ND'') binds, then (28) implies $\kappa_E(\theta) \neq 0$ and thus, by (20), $Y_L(\theta) \neq \lambda_{LM}$.

A.3 Proof of Proposition 4

(i) After integrating by parts, the Lagrangian corresponding to the Pareto problem now becomes

$$\begin{aligned}\mathcal{L} \;=\; & \int_\Theta\left[\tilde{G}_\theta(\tilde{\phi}(\theta))v_E(\theta) - \int_{\underline{\phi}}^{\tilde{\phi}(\theta)}\phi d\tilde{G}_\theta(\phi) + (1-\tilde{G}_\theta(\tilde{\phi}(\theta)))v_W(\theta)\right]d\tilde{F}(\theta) \\ & - \int_\Theta\left[\mu_E'(\theta)v_E(\theta) + \mu_E(\theta)\psi'\left(\frac{E(\theta)}{\theta}\right)\frac{E(\theta)}{\theta^2}\right]d\theta - \int_\Theta\left[\mu_W'(\theta)v_W(\theta) + \mu_W(\theta)\psi'\left(\frac{l(\theta)}{\theta}\right)\frac{l(\theta)}{\theta^2}\right]d\theta \\ & + \lambda_{LM}\left[\int_\Theta(1 - G_\theta(\tilde{\phi}(\theta)))l(\theta)dF(\theta) - \int_\Theta G_\theta(\tilde{\phi}(\theta))L(\theta)dF(\theta)\right] \\ & + \lambda_{RC}\left[\int_\Theta G_\theta(\tilde{\phi}(\theta))\left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi\left(\frac{E(\theta)}{\theta}\right)\right] - (1-G(\tilde{\phi}(\theta)))\left[v_W(\theta) + \psi\left(\frac{l(\theta)}{\theta}\right)\right]dF(\theta)\right].\end{aligned} \tag{29}$$

The necessary condition for $L(\theta)$ immediately implies

$$Y_L(L(\theta), E(\theta)) = \lambda_{LM}/\lambda_{RC} \ \ \forall\theta \in \Theta \tag{30}$$

and hence the result.

(ii) Note that (30) together with constant returns to scale implies that both $Y_L(\theta)$ and $Y_E(\theta)$ are equalized across all θ, and I can therefore again write $\tilde{w} \equiv Y_E$ and $w \equiv Y_L$. Hence $w = \lambda_{LM}/\lambda_{RC}$ and the necessary condition for $v_E(\theta)$ can be rearranged to

$$\mu_E'(\theta) = \tilde{G}(\tilde{\phi}(\theta))\tilde{f}(\theta) - \lambda_{RC}G(\tilde{\phi}(\theta))f(\theta) + g(\tilde{\phi}(\theta))f(\theta)\lambda_{RC}\left[Y(\theta) - c_E(\theta) + c_W(\theta) - w\left(L(\theta) + l(\theta)\right)\right], \tag{31}$$

where $c_E(\theta) \equiv v_E(\theta) + \psi(E(\theta)/\theta)$ and $c_W(\theta) \equiv v_W(\theta) + \psi(l(\theta)/\theta)$. Note first that, by Euler's theorem, $Y(\theta) - wL(\theta) = \tilde{w}E(\theta)$. Next, let me define the excess entrepreneurial tax (i.e. the additional tax payment by an entrepreneur of type θ compared to a worker of type θ) as

$$\Delta T(\theta) \equiv T_\pi(\pi(\theta)) - T_y(y(\theta)) = \tilde{w}E(\theta) - c_E(\theta) - (wl(\theta) - c_W(\theta)).$$

Then using the transversality conditions $\mu_E(\underline{\theta}) = \mu_E(\overline{\theta}) = 0$, I obtain

$$0 = \int_\Theta \left[\tilde{G}(\tilde{\phi}(\theta))\tilde{f}(\theta) - \lambda_{RC}G(\tilde{\phi}(\theta))f(\theta) + g(\tilde{\phi}(\theta))f(\theta)\lambda_{RC}\Delta T(\theta)\right] d\theta.$$

By the same steps, the necessary condition for $v_W(\theta)$ can be transformed to

$$0 = \int_\Theta \left[(1 - \tilde{G}(\tilde{\phi}(\theta)))\tilde{f}(\theta) - \lambda_{RC}(1 - G(\tilde{\phi}(\theta)))f(\theta) - g(\tilde{\phi}(\theta))f(\theta)\lambda_{RC}\Delta T(\theta)\right] d\theta.$$

Adding the two equations yields $\lambda_{RC} = 1$. With this, I find that, for all $\theta \in \Theta$,

$$\mu_E(\theta) = \int_{\underline{\theta}}^{\theta} \left[\tilde{G}(\tilde{\phi}(\hat{\theta}))\tilde{f}(\hat{\theta}) - G(\tilde{\phi}(\hat{\theta}))f(\hat{\theta}) + g(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})\Delta T(\hat{\theta})\right] d\hat{\theta} \tag{32}$$

and

$$\mu_W(\theta) = \int_{\underline{\theta}}^{\theta} \left[(1 - \tilde{G}(\tilde{\phi}(\hat{\theta})))\tilde{f}(\hat{\theta}) - (1 - G(\tilde{\phi}(\hat{\theta})))f(\hat{\theta}) - g(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})\Delta T(\hat{\theta})\right] d\hat{\theta}. \tag{33}$$

Next, consider the necessary condition for $E(\theta)$, which is given by

$$G(\tilde{\phi}(\theta))f(\theta)\left[\tilde{w} - \frac{1}{\theta}\psi'\left(\frac{E(\theta)}{\theta}\right)\right] = \frac{\mu_E(\theta)}{\theta}\left[\psi'\left(\frac{E(\theta)}{\theta}\right)\frac{1}{\theta} + \psi''\left(\frac{E(\theta)}{\theta}\right)\frac{E(\theta)}{\theta^2}\right].$$

Dividing through by $\psi'(E(\theta)/\theta)/\theta$ and rearranging yields

$$\frac{\tilde{w} - \psi'(E(\theta)/\theta)/\theta}{\psi'(E(\theta)/\theta)/\theta} = \frac{\mu_E(\theta)}{\theta f(\theta)G(\tilde{\phi}(\theta))}\left(1 + \frac{\psi''(E(\theta)/\theta)E(\theta)/\theta^2}{\psi'(E(\theta)/\theta)/\theta}\right). \tag{34}$$

Note that the entrepreneur's first order condition from $\max_E \tilde{w}E - T_\pi(\tilde{w}E) - \psi(E/\theta)$ is

$$\tilde{w}(1 - T_\pi'(\pi(\theta))) = \psi'\left(\frac{E(\theta)}{\theta}\right)\frac{1}{\theta},$$

where $\pi(\theta) \equiv \tilde{w}E(\theta)$, and hence the elasticity of entrepreneurial effort $E(\theta)$ with respect to the after-tax wage $\tilde{w}(1 - T_\pi'(\pi(\theta)))$ is

$$\varepsilon_\pi(\theta) = \frac{\psi'(E(\theta)/\theta)/\theta}{\psi''(E(\theta)/\theta)E(\theta)/\theta^2}.$$

After substituting (32), this allows me to rewrite (34) as

$$\frac{T_\pi'(\pi(\theta))}{1-T_\pi'(\pi(\theta))}=\frac{1+1/\varepsilon_\pi(\theta)}{\theta f(\theta)G_\theta(\tilde{\phi}(\theta))}\int_{\underline{\theta}}^{\theta}[\tilde{G}_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\tilde{f}(\hat{\theta})-G_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})+g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\Delta T(\hat{\theta})f(\hat{\theta})]d\hat{\theta}, \tag{35}$$

which is the result in Proposition 4. The derivation for $T_y'(y(\theta))$ proceeds completely analogously from the necessary condition for $l(\theta)$ and using (33).

(iii) $T_\pi'(\pi(\underline{\theta})) = T_\pi'(\pi(\overline{\theta})) = 0$ immediately follows from (34) evaluated at $\underline{\theta}$ and $\overline{\theta}$ and the transversality conditions $\mu_E(\underline{\theta}) = \mu_E(\overline{\theta}) = 0$. Analogously, $T_y'(y(\underline{\theta})) = T_y'(y(\overline{\theta})) = 0$ is implied by the first order conditions for $l(\underline{\theta})$ and $l(\overline{\theta})$ and the transversality conditions for $\mu_W(\theta)$.

A.4 Proof of Proposition 5

(i) By way of contradiction, suppose there exists some $\theta \in (\underline{\theta}, \overline{\theta})$ such that $T_\pi'(\pi(\theta)) \leq 0$ and $T_y'(y(\theta)) \geq 0$. By continuity of the marginal tax rates (from ignoring bunching issues), and the result that marginal tax rates are zero at the top and bottom, this implies that there must exist a subinterval $[\theta_a, \theta_b]$ of Θ such that $T_\pi'(\pi(\theta)) \leq 0$ and $T_y'(y(\theta)) \geq 0$ for all $\theta \in (\theta_a, \theta_b)$ and $T_\pi'(\pi(\theta)) = T_y'(y(\theta)) = 0$ at θ_a and θ_b. Using $\tilde{F}(\theta) = F(\theta)$, independence of θ and ϕ and the optimality formulas in Proposition 5, this implies

$$\int_{\underline{\theta}}^{\theta} \left[\tilde{G}(\tilde{\phi}(\hat{\theta})) - G(\tilde{\phi}(\hat{\theta})) + g(\tilde{\phi}(\hat{\theta}))\Delta T(\hat{\theta})\right] dF(\hat{\theta}) \leq 0$$

on (θ_a, θ_b), with equality at θ_a and θ_b. Taking derivatives at θ_a and θ_b, I must therefore have

$$\tilde{G}(\tilde{\phi}(\theta_a)) - G(\tilde{\phi}(\theta_a)) + g(\tilde{\phi}(\theta_a))\Delta T(\theta_a) \leq 0 \quad \text{and} \quad \tilde{G}(\tilde{\phi}(\theta_b)) - G(\tilde{\phi}(\theta_b)) + g(\tilde{\phi}(\theta_b))\Delta T(\theta_b) \geq 0,$$

which can be rearranged to

$$\Delta T(\theta_a) \leq \frac{G(\tilde{\phi}(\theta_a)) - \tilde{G}(\tilde{\phi}(\theta_a))}{g(\tilde{\phi}(\theta_a))} \quad \text{and} \quad \Delta T(\theta_b) \geq \frac{G(\tilde{\phi}(\theta_b)) - \tilde{G}(\tilde{\phi}(\theta_b))}{g(\tilde{\phi}(\theta_b))}. \tag{36}$$

The assumption that $T_\pi'(\pi(\theta)) \leq 0$ and $T_y'(y(\theta)) \geq 0$ for all $\theta \in (\theta_a, \theta_b)$ and $\tilde{w} \geq w$ imply by the agents' first-order conditions

$$\tilde{w}(1 - T_\pi'(\tilde{w}E(\theta))) = \frac{1}{\theta}\psi'\left(\frac{E(\theta)}{\theta}\right) \quad \text{and} \quad w(1 - T_y'(wl(\theta))) = \frac{1}{\theta}\psi'\left(\frac{l(\theta)}{\theta}\right)$$

that $E(\theta) > l(\theta)$ for all $\theta \in [\theta_a, \theta_b]$ and hence that

$$\tilde{\phi}'(\theta) = v_E'(\theta) - v_W'(\theta) = \frac{E(\theta)}{\theta^2}\psi'\left(\frac{E(\theta)}{\theta}\right) - \frac{l(\theta)}{\theta^2}\psi'\left(\frac{l(\theta)}{\theta}\right) \geq 0 \;\; \forall \theta \in (\theta_a, \theta_b),$$

where I have used the local incentive constraints (8). Hence, I obtain $\tilde{\phi}(\theta_a) \leq \tilde{\phi}(\theta_b)$. Next, note that by the assumption in the proposition that $\tilde{g}(\phi) \leq g(\phi)$ for all $\phi \leq \tilde{\phi}(\overline{\theta})$ and by the second part of Assumption 1, $\left(G(\tilde{\phi}) - \tilde{G}(\tilde{\phi})\right)/g(\tilde{\phi})$ is non-decreasing in $\tilde{\phi}$. With this, equation (36) yields $\Delta T(\theta_a) \leq \Delta T(\theta_b)$. But recall that I assumed $T_\pi'(\pi(\theta)) \leq 0$ and $T_y'(y(\theta)) \geq 0$ for all $\theta \in (\theta_a, \theta_b)$. Therefore,

$$\Delta T'(\theta) = T_\pi'(\tilde{w}E(\theta))\tilde{w}E'(\theta) - T_y'(wl(\theta))wl'(\theta) < 0 \;\; \forall \theta \in (\theta_a, \theta_b),$$

where I have used (9) and thus $E'(\theta), l'(\theta) \geq 0$. This implies $\Delta T(\theta_a) > \Delta T(\theta_b)$ and hence the desired contradiction.

(ii) Note first that part (i) immediately implies

$$\Delta T'(\theta) = T'_\pi(\tilde{w}E(\theta))\tilde{w}E'(\theta) - T'_y(wl(\theta))wl'(\theta) > 0 \;\; \forall \theta \in \Theta.$$

Next, at $\underline{\theta}$, I must have

$$\tilde{G}(\tilde{\phi}(\underline{\theta})) - G(\tilde{\phi}(\underline{\theta})) + g(\tilde{\phi}(\underline{\theta}))\Delta T(\underline{\theta}) \geq 0$$

by the same arguments as in the proof for part (i). Since $\tilde{G}(\tilde{\phi}(\underline{\theta})) < G(\tilde{\phi}(\underline{\theta}))$ by the assumption in the proposition, I obtain $\Delta T(\underline{\theta}) > 0$ and therefore $\Delta T(\theta) > 0$ for all $\theta \in \Theta$.

(iii) Suppose $w = Y_L$ increases and thus $\tilde{w} = Y_E$ falls compared to the no-tax equilibrium. Then part (i) implies that $E(\theta)$ falls and $l(\theta)$ increases for all $\theta \in \Theta$ compared to the no-tax equilibrium. Moreover, by constant returns to scale, an increase in Y_L implies an increase in $E(\theta)/L(\theta)$, and hence $L(\theta)$ must fall for all $\theta \in \Theta$. Finally, note that

$$\begin{aligned}
\tilde{\phi}(\theta) &= v_E(\theta) - v_W(\theta) = \left(\tilde{w}E(\theta) - T_\pi(\tilde{w}E(\theta)) - \psi\left(\frac{E(\theta)}{\theta}\right)\right) - \left(wl(\theta) - T_y(wl(\theta)) - \psi\left(\frac{l(\theta)}{\theta}\right)\right) \\
&= \left(\tilde{w}E(\theta) - \psi\left(\frac{E(\theta)}{\theta}\right)\right) - \left(wl(\theta) - \psi\left(\frac{l(\theta)}{\theta}\right)\right) - \Delta T(\theta) \;\; \forall \theta \in \Theta.
\end{aligned}$$

Since w increases and $\tilde{w}$ falls by assumption, and because of part (i), $\tilde{w}E(\theta) - \psi(E(\theta)/\theta)$ falls and $wl(\theta) - \psi(l(\theta)/\theta)$ increases compared to the no-tax equilibrium. Moreover, since $\Delta T(\theta) = 0$ in the no-tax equilibrium and $\Delta T(\theta) > 0$ by part (ii) in the Pareto optimum with redistribution, I conclude that $\tilde{\phi}(\theta)$ falls for all $\theta \in \Theta$. Putting this together with the above results for $E(\theta)$, $L(\theta)$ and $l(\theta)$, this means that the labor market clearing constraint (12) is strictly slack in the Pareto optimum. This cannot be part of a Pareto optimum, however, since increasing $L(\theta)$ for some θ increases production and thus relaxes the resource constraint (13) without affecting any other constraint nor the objective of the Pareto problem. A slack resource constraint in turn cannot be Pareto optimal since consumption could be increased uniformly without affecting incentives nor occupational choice, increasing the objective for any set of Pareto weights. This completes the proof.

(iv) Since both $E(\theta) < (w/\tilde{w})l((\tilde{w}/w)\theta)$ and $v_E(\theta) < v_W((\tilde{w}/w)\theta)$ compare individuals who earn the same "total" wage but in different occupations, it is useful to define this total return to effort as $\omega \equiv \tilde{w}\theta = w\theta'$, where $\theta' = (\tilde{w}/w)\theta$. Writing allocations in terms of these total wages, the inequalities can be expressed as $\pi(\omega) < y(\omega)$ and $v_E(\omega) < v_W(\omega)$ with $\pi \equiv \tilde{w}E$ and $y \equiv wl$. Since $\pi(\omega)$ and $y(\omega)$ solve the individual first-order conditions

$$1 - T'_\pi(\omega) = \frac{\psi'(\pi/\omega)}{\omega} \text{ and } 1 - T'_y(\omega) = \frac{\psi'(y/\omega)}{\omega}$$

and $T'_\pi > 0$, $T'_y < 0$ by (i), the inequalities $\pi(\omega) < y(\omega)$ for all ω immediately follow. To see the second result, first define $\Delta T(\omega) \equiv T_\pi(\pi(\omega)) - T_y(y(\omega))$ and note that

$$\Delta T(\underline{\omega}) = T_\pi(\tilde{w}E(\underline{\theta})) - T_y\left(wl\left(\frac{\tilde{w}}{w}\underline{\theta}\right)\right) > T_\pi(\tilde{w}E(\underline{\theta})) - T_y\left(wl\left(\underline{\theta}\right)\right) = \Delta T(\underline{\theta}) > 0,$$

where the first inequality follows from the fact that $T_y(.)$ is decreasing by (i) and $l(.)$ is increasing, and the second from (ii). Moreover,

$$\Delta T'(\omega) = T'_\pi(\pi(\omega))\pi'(\omega) - T'_y(y(\omega))y'(\omega) > 0$$

by (i), which implies $\Delta T(\omega) > 0$ for all ω. As a consequence, $T_\pi(y(\omega)) > T_\pi(\pi(\omega)) > T_y(y(\omega))$ for all ω, where the first inequality holds because $T_\pi(.)$ is increasing and $y(\omega) > \pi(\omega)$, and the second because $\Delta T(\omega) > 0$. This implies $T_\pi(z) > T_y(z)$ for all z. Finally, note that

$$\begin{aligned} v_W(\omega) = y(\omega) - T_y(y(\omega)) - \psi\left(\frac{y(\omega)}{\omega}\right) \quad &\geq \quad \pi(\omega) - T_y(\pi(\omega)) - \psi\left(\frac{\pi(\omega)}{\omega}\right) \\ &> \quad \pi(\omega) - T_\pi(\pi(\omega)) - \psi\left(\frac{\pi(\omega)}{\omega}\right) = v_E(\omega), \end{aligned}$$

where the first inequality follows from the fact that $y(\omega)$ is optimal for a worker of total wage ω faced with tax schedule T_y, and the second inequality follows from $T_\pi(z) > T_y(z)\ \forall z$.

B Computational Procedure for Section 4.2

To compute the optimal schedules T_π and T_y for any set of Pareto weights, I first fix some $x \equiv E(\theta)/L(\theta)$, equal for all θ, which implies wages $\tilde{w} = Y_E(x)$ and $w = Y_L(x)$ for entrepreneurs and workers. Then I proceed as outlined in the following steps:

1. Start with an initial guess for the marginal tax schedules $T'_\pi(\pi(\theta))$ and $T'_y(y(\theta))$.
2. Given this, compute $E(\theta)$ and $l(\theta)$ from the individual first-order conditions

$$\tilde{w}(1 - T'_\pi(\pi(\theta))) = \frac{1}{\theta}\psi'\left(\frac{E(\theta)}{\theta}\right) \quad \text{and} \quad w(1 - T'_y(y(\theta))) = \frac{1}{\theta}\psi'\left(\frac{l(\theta)}{\theta}\right).$$

   Also, $L(\theta)$ is obtained from $x = E(\theta)/L(\theta)$ and $E(\theta)$.
3. Note that the marginal tax schedules $T'_\pi(\pi(\theta))$ and $T'_y(y(\theta))$ pin down the actual tax schedules $T_\pi(\pi(\theta))$ and $T_y(y(\theta))$, except for the two intercepts, which in turn are given by $T_y(y(\underline{\theta}))$ and $\Delta T(\underline{\theta})$. To find these two, proceed as follows:

   (a) First, find $\Delta T(\underline{\theta})$ from solving the transversality condition

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[\tilde{G}_\theta(\tilde{\phi}(\theta))\tilde{f}(\theta) - G_\theta(\tilde{\phi}(\theta))f(\theta) + g_\theta(\tilde{\phi}(\theta))\Delta T(\theta)f(\theta)\right] d\theta = 0,$$

   using the fact that

$$\tilde{\phi}(\theta) = \left(\tilde{w}E(\theta) - \psi\left(\frac{E(\theta)}{\theta}\right)\right) - \left(wl(\theta) - \psi\left(\frac{l(\theta)}{\theta}\right)\right) - \Delta T(\theta)$$

   and

$$\Delta T(\theta) = \Delta T(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \left[T'_\pi(\pi(\hat{\theta}))\tilde{w}E'(\hat{\theta}) - T'_y(y(\hat{\theta}))wl'(\hat{\theta})\right] d\hat{\theta}.$$

(b) Then find $T_y(y(\underline{\theta}))$ from solving the resource constraint (RC)

$$\int_{\Theta} G_\theta(\tilde{\phi}(\theta)) \left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi(E(\theta)/\theta)\right] dF(\theta)$$
$$-(1 - G(\tilde{\phi}(\theta))) \left[v_w(\theta) + \psi(l(\theta)/\theta)\right] dF(\theta) \quad = \quad 0,$$

using $v_E(\theta) = \tilde{w}E(\theta) - T_\pi(\pi(\theta)) - \psi(E(\theta)/\theta)$ and $v_W(\theta) = wl(\theta) - T_y(y(\theta)) - \psi(l(\theta)/\theta)$.

4. Use the optimality formulas in Proposition 4 to compute updated marginal tax schedules $T'_\pi(\pi(\theta))$ and $T'_y(y(\theta))$. Repeat steps 2. to 4. until convergence.

For any given x and hence wages $\tilde{w}$ and w, iterating on 1. to 4. yields tax schedules and an allocation that satisfy the optimality formulas as well as the transversality conditions and the resource constraint. Finally, I adjust x until the labor market clearing condition (LM) holds with equality.