Entrepreneurial Taxation with Endogenous Entry

Florian Scheuer*
Stanford University and NBER

Revised Version: July 2013

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 low 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 thank Daron Acemoglu, Robert Barro, Peter Diamond, Roger Gordon, James Poterba, Emmanuel Saez, Karl Scholz, Iván Werning, and numerous seminar participants for valuable comments and suggestions. I owe special thanks to Stefania Albanesi and Christian Keuschnigg for discussions of this paper at the NBER TAPES conference 2012 (Oxford). All errors are my own. Some results in this paper were part of the earlier working paper circulated under the title "Entrepreneurial Taxation and Occupational Choice."

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, individuals are characterized by two-dimensional heterogeneity: Agents differ in their cost of (or preference for) 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 considering 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 a natural benchmark to start with 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 is often viewed as particularly appealing in light 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.[2] [3]

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.[4]

In characterizing optimal allocations, my work therefore shares a common goal with

[2]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. In the UK, for instance, social insurance contributions differ between employed and self-employed.

[3]In the longer working paper version (Scheuer, 2013b), I numerically compute 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.

[4]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.

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 optimal taxation in a static general equilibrium model that allows for endogenous wages and entry into entrepreneurship, providing a complementary contribution.

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 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. Related work by Rothschild and Scheuer (2011, 2013) considers optimal income taxation in occupational choice economies with multidimensional skill heterogeneity, where individuals can have different skills in different occupations. Their structure of private information therefore differs from the one considered here and their focus is on uniform taxation only. 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 provides a discussion of possible extensions and 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 $(\theta, \phi) \in [\underline{\theta}, \overline{\theta}] \times [\underline{\phi}_\theta, \overline{\phi}_\theta]$, where θ will be interpreted as an individual's skill, and ϕ as an individual's cost of or taste for becoming an entrepreneur, as explained in more detail below.[5] $F(\theta)$ 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.

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

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

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 non-pecuniary benefits from being an entrepreneur (e.g. a preference for being one's own boss).[6] 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 enter entrepreneurship and others who become workers due to their different ϕ-type. This is an empirically attractive implication of the present specification, since, in reality, the income distributions of workers and entrepreneurs do have overlapping supports.[7]

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

[6] This means that ϕ could in principle take positive or negative values for different individuals. Whether ϕ is positive or negative for most individuals determines the equilibrium return to effort of entrepreneurs relative to workers. See Section 5 for a discussion of the related evidence.

[7] 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)).

θ is determined by the critical cost value

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

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 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 $\tilde{w}(w)$:[8] 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:

[8]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. *Consider the no tax equilibrium as defined in Definition 1. If $\underline{\phi}_\theta \geq 0$ ($\overline{\phi}_\theta \leq 0$) for all $\theta \in \Theta$,*
(i) the entrepreneurs' wage is higher (lower) than 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 (decreasing) in θ, and
(iii) the share of entrepreneurs $G_\theta(\tilde{\phi}(\theta, w^*))$ is increasing (decreasing) in θ if $G_{\theta'}(\phi) \succeq_{FOSD} G_\theta(\phi)$ for $\theta' \leq (\geq) \theta$.

Proof. Consider the case with $\underline{\phi}_\theta \geq 0$ for all θ (the case $\overline{\phi}_\theta \leq 0$ is completely analogous). (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), $\tilde{\phi}(\theta, w^*) = \underline{\phi}_\theta$ 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. □

Proposition 1 summarizes intuitive properties of wages and occupational choice in equilibrium: First, if $\underline{\phi}_\theta \geq 0$, 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 $\phi \geq 0$ 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. Of course, the reversed results obtain if $\overline{\phi}_\theta \leq 0$, i.e. everyone has a preference for entrepreneurship relative to employment conditional on earning the same wage. Then the worker's wage must exceed the entrepreneurs' wage in equilibrium.

The second result in the proposition is that, if $\tilde{w}^* > w^*$, 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. Again, the opposite pattern results if entrepreneurship is associated with a benefit for everyone and hence $\tilde{w}^* < w^*$. More generally, 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 $G_\theta(\phi)$.[9]

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. 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"

[9] In Scheuer (2013b), I calibrate $G_\theta(\phi)$ to match the relationship between income and entrepreneurship found in the data.

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.[10] 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

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

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

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

[10] 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.

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

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

hold and

$$l(\theta) \quad \text{and} \quad E(\theta) \quad \text{are non-decreasing.}^{11} \tag{9}$$

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

$$\tilde{\phi}(\theta) = v_E(\theta) - v_W(\theta) \ \ \forall\theta\in\Theta.^{12} \tag{10}$$

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.

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

Figure 1 illustrates these relationships.

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 $\tilde{G}_\theta(\phi)$.[13] Then the program is

$$\max_{\substack{\{E(\theta),L(\theta),l(\theta),v_E(\theta),\\ v_W(\theta),\tilde{\phi}(\theta),w,\tilde{w}\}}} \int_\Theta \left[\tilde{G}_\theta(\tilde{\phi}(\theta))v_E(\theta) - \int_{\underline{\phi}_\theta}^{\tilde{\phi}(\theta)} \phi d\tilde{G}_\theta(\phi) + (1-\tilde{G}_\theta(\tilde{\phi}(\theta)))v_W(\theta)\right] d\tilde{F}(\theta) \tag{12}$$

subject to
$$\int_\Theta G_\theta(\tilde{\phi}(\theta))L(\theta)dF(\theta) \le \int_\Theta (1-G_\theta(\tilde{\phi}(\theta)))l(\theta)dF(\theta), \tag{13}$$

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

[12] 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 $\phi \le \tilde{\phi}(\theta)$ become entrepreneurs and the others workers.

[13] Section 5 discusses how the results are affected when using a social welfare function rather than general Pareto weights.

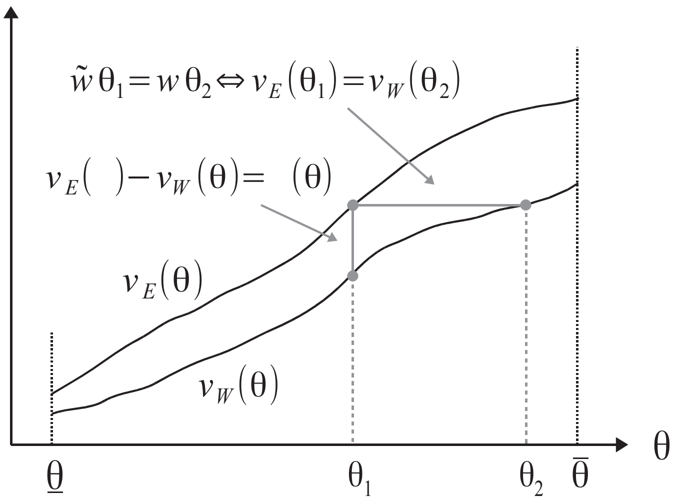

Figure 1: Relationship between $v_E(\theta)$ and $v_W(\theta)$ with no-discrimination constraints

$$\int_{\Theta} G_\theta(\tilde{\phi}(\theta)) \left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi(E(\theta)/\theta)\right] dF(\theta)$$
$$- \int_{\Theta} (1 - G_\theta(\tilde{\phi}(\theta))) \left[v_W(\theta) + \psi(l(\theta)/\theta)\right] dF(\theta) \quad \geq \quad 0, \tag{14}$$

and constraints (3), (6), (7), (8), (9), (10) and (11). Inequality (13) requires the total amount of labor demand assigned to entrepreneurs not to exceed the total amount of labor supply assigned to workers. Similarly, (14) 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.[14]

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

[14] 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 working paper version (Scheuer, 2013b), I check numerically whether the monotonicity constraint is satisfied at the optimum, and find that bunching does not arise.

non-standard properties. The following two propositions summarize characteristics of constrained Pareto optimal tax systems.

Proposition 2. *(i) If* $\underline{\phi}_\theta \geq 0$ *(*$\overline{\phi}_\theta \leq 0$*)* $\forall\, \theta \in \Theta$*, then at any Pareto-optimum,* $\tilde{w} > (<) w$*, and* $\tilde{w}E(\theta) > (<) wl(\theta)$ *for all* $\theta \in \Theta$*.*
(ii) If $Y(L, E)$ *is linear and* $\tilde{w} > (<) w$*,* $T'(wl(\underline{\theta})) = T'(\tilde{w}E(\overline{\theta})) = 0$ *(*$T'(wl(\overline{\theta})) = T'(\tilde{w}E(\underline{\theta})) = 0$*).*
(iii) If $Y(L, E)$ *is not linear and* $\tilde{w} > (<) w$*,* $T'(wl(\underline{\theta}))$ *and* $T'(\tilde{w}E(\overline{\theta}))$ *(*$T'(wl(\overline{\theta}))$ *and* $T'(\tilde{w}E(\underline{\theta}))$*) have opposite signs whenever (3) and (11) bind for some* $\theta \in \Theta$*.*
(iv) For instance, suppose the workers' effort $l(\theta)/\theta$ *is increasing in* θ *and* $\tilde{w} > w$*. 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, if becoming an entrepreneur is costly, 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 such Pareto optimum is an entrepreneur, and the bottom earner a worker.[15] On the other hand, if entrepreneurship provides a utility benefit and so $\phi \leq 0$ for everyone, the opposite pattern results, with the bottom earner an entrepreneur and the top earner a worker (and the no-discrimination constraints not binding at the bottom of the skill distribution).

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 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, if $\tilde{w} > w$ and hence the set of top earners is exclusively given by entrepreneurs and

[15] It also implies that the no-discrimination constraints (3) and (11) do not bind at the top of the skill distribution in this case: 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 $b = (w/\tilde{w})\overline{\theta}$.

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.[16] In the present occupational 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

[16] Allen (1982) analyzes optimal *linear* taxation with endogenous wages. In this case, the incentive effects of taxes on wages through the labor supply of different income groups are less clear, since all agents face the same marginal tax rate.

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.[17]

The next results demonstrate 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 $T'(wl(\underline{\theta})) = T'(\tilde{w}E(\bar{\theta})) = 0$ ($T'(wl(\bar{\theta})) = T'(\tilde{w}E(\underline{\theta})) = 0$) for $\tilde{w} > (<)w$.*
(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 profit tax, even when complemented by a proportional tax on labor inputs, implies that marginal products of labor (and thus of entrepreneurial effort) are equalized across 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 on how much (and in which direction) the no-discrimination constraint binds locally. 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.

[17] Rothschild and Scheuer (2013) extend these results to a general Roy model of occupational choice where individuals can have different skills for different sectors and preferences allow for income effects. They also numerically compute Pareto optimal uniform tax schedules in this setting.

This is in contrast to the Diamond-Mirrlees Theorem (Diamond and Mirrlees, 1971) in settings without pecuniary externalities.[18]

4 Differential Tax Treatment of Profits and Income

In this section, I relax the assumption that the government can only impose a single nonlinear 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 are that (i) the 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, this analysis 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.

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:

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

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

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

[18]See 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.

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

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

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) $T'_\pi(\pi(\underline{\theta})) = T'_\pi(\pi(\overline{\theta})) = T'_y(y(\underline{\theta})) = T'_y(y(\overline{\theta})) = 0$.

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.[19]

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

[19] 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.

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.[20]

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 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

[20]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 (see also Section 5).

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,[21] the average marginal tax across occupations satisfies*

$$G_\theta(\tilde{\phi}(\theta))\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))}+(1-G_\theta(\tilde{\phi}))\,\frac{T'_y(y(\theta))}{1-T'_y(y(\theta))}=\frac{1+1/\varepsilon}{\theta f(\theta)}\left(\tilde{F}(\theta)-F(\theta)\right). \qquad (15)$$

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 (15): 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 (15) is the same as the one that would be obtained in a standard quasi-linear Mirrlees-model without occupational choice and with only one-dimensional heterogeneity in θ.[22]

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

[21] Even without a constant elasticity, a modified version of (15) 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.

[22] 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 (15) cannot be obtained.

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*

$$\frac{\theta f_E(\theta)}{1+1/\varepsilon}\,\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))} + F_E(\theta) - \int_{\underline{\theta}}^{\theta} \frac{g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})}{\overline{G}}\Delta T(\hat{\theta})d\hat{\theta} \quad \textit{and} \tag{16}$$

$$\frac{\theta f_W(\theta)}{1+1/\varepsilon}\,\frac{T'_y(y(\theta))}{1-T'_y(y(\theta))} + F_W(\theta) + \int_{\underline{\theta}}^{\theta} \frac{g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})}{1-\overline{G}}\Delta T(\hat{\theta})d\hat{\theta} \tag{17}$$

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 (16) and (17) 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 has been extended in the working paper version (Scheuer, 2013b) to the setting with two-dimensional heterogeneity and occupational choice considered here.

Two remarks on Corollary 2 are in order. First, adding conditions (16) and (17) yields another test for Pareto optimality, which is weaker but requires less information to be 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 (16) and (17) 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 $\underline{\phi}$ and $\overline{\phi}$, with $\underline{\phi}$ ($\overline{\phi}$) sufficiently low (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 (16) and (17) 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.

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 are relaxed in the numerical explorations in the working paper version (Scheuer, 2013b). 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 $\phi \leq \tilde{\phi}(\overline{\theta})$, $\tilde{w} > w$ 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

$$\frac{T'_\pi(\pi(\theta))}{1 - T'_\pi(\pi(\theta))} = \frac{T'_y(y(\theta))}{1 - T'_y(y(\theta))} = \frac{1 + 1/\varepsilon}{\theta f(\theta)} \left(\tilde{F}(\theta) - F(\theta)\right) \tag{18}$$

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 is further explored in the numerical simulations in Scheuer (2013b).[23]

5 Discussion and 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. I conclude with an informal discussion of possible re-interpretations and extensions.[24]

5.1 Social Welfare Function

Instead of tracing out the entire constrained Pareto frontier using general Pareto weights in the two dimensions of heterogeneity $\tilde{F}(\theta)$ and $\tilde{G}_\theta(\phi)$, much of the literature on optimal taxation following Mirrlees (1971) has confined attention to maximizing a social welfare function. This typically takes the utilitarian form using a concave transformation $W(.)$ of

[23] 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.

[24] I thank Roger Gordon for suggesting this discussion.

individual utilities, so that the objective function would become

$$\int_{\Theta}\left[\int_{\underline{\phi}_\theta}^{\tilde{\phi}(\theta)} W(v_E(\theta)-\phi)dG_\theta(\phi)+(1-G_\theta(\tilde{\phi}(\theta)))W(v_W(\theta))\right]dF(\theta)$$

instead of (12). The resulting solution would select a particular point on the Pareto frontier, which is why all the properties derived here for the entire set of constrained Pareto optimal tax policies would continue to apply. In particular, since both the entrepreneurs' and workers' utilities are increasing in θ and the entrepreneurs' utilities are decreasing in ϕ, a concave welfare function $W(.)$ would put more marginal welfare weight on individuals with low θ and high ϕ. Hence, the optimum would correspond to a point on the Pareto frontier with Pareto weights $\tilde{F}(\theta) > F(\theta)$ and $\tilde{G}_\theta(\phi) < G_\theta(\phi)$ for all θ and ϕ, i.e. simultaneous redistributive motives from high to low skills and from entrepreneurs to workers, as considered in the numerical analysis in Scheuer (2013b).

The advantage of using exogenous Pareto weights is that it makes the redistributive motives in the two dimensions of heterogeneity fully transparent, and allows to separate their implications for properties of optimal tax schedules, as shown in Section 4. It also allows for the closed form marginal tax rate formulas in Proposition 4 and Corollary 1. With a social welfare function $W(.)$, the structure of the tax formulas would be exactly the same, but the marginal social welfare weights on entrepreneurs of skill type θ would become $\int_{\underline{\phi}_\theta}^{\tilde{\phi}(\theta)} W'(v_E(\theta)-\phi)dG_\theta(\phi)f(\theta)/\lambda$ rather than simply $\tilde{G}_\theta(\tilde{\phi}(\theta))\tilde{f}(\theta)$, where λ would be the (endogenous) marginal social cost of public funds, and analogously for workers (see Diamond, 1998, and Kleven, Kreiner and Saez, 2009).

5.2 Determinants of Entry into Entrepreneurship

I have introduced the second dimension of heterogeneity ϕ to capture different idiosyncratic preferences for or costs of entering entrepreneurship, which drive occupational choice in addition to an individual's skill type θ. This could include disutilities with $\phi \geq 0$ (for example the fact that entrepreneurship is riskier than employment or setup costs) and benefits with $\phi \leq 0$ (e.g. the value derived from being one's own boss as a self-employed). As formalized in Propositions 1 and 2 in the context of this model, whether ϕ is mostly positive or negative in the population determines whether the entrepreneurs' wage is higher or lower than the workers' in equilibrium. In particular, if entering entrepreneurship is perceived as costly, equilibrium returns to entrepreneurship have to exceed the workers' wage as a compensation, and vice versa.

As shown based on evidence from the 2007 Survey of Consumer Finances in the longer working paper version (Scheuer, 2013b), the average hourly wage of entrepreneurs is indeed higher than that of workers, suggesting that the support of ϕ is mostly positive. However, this evidence needs to be interpreted carefully. While De Nardi, Doctor, and Krane (2007) also find that entrepreneurs have higher incomes than workers, and Berglann, Moen, Roed, and Skogstrom (2010) confirm this pattern for wages, controlling for hours, Hamilton (2000) and Blanchflower (2004) find lower returns to entrepreneurship than to employment. One difference that can explain this is the definition of entrepreneurship. The latter studies set entrepreneurship equal to self-employment, whereas the former consider individuals as entrepreneurs only if they are not just self-employed, but also own and actively manage a business and hire employees.

The latter definition, even though common, is therefore likely to select the most successful subset of the self-employed, whose business survives for long enough and grows to the point that they end up hiring employees. But many self-employed leave entrepreneurship shortly after entering, and panel evidence suggests that even those who stay report lower incomes than what they previously earned as workers (see e.g. Hamilton, 2000). These findings have been associated with non-pecuniary benefits to entrepreneurship, corresponding to $\phi \leq 0$.[25] On the other hand, since the present model is static, it is best interpreted in terms of lifetime incomes, or wealth. Tergiman (2013) argues that returns to entrepreneurship can be positive from a life-cycle perspective, and Quadrini (2000) and Cagetti and De Nardi (2006) provide evidence that entrepreneurs have more wealth than workers (although the direction of causation is unclear).

A standard justification for $\phi \geq 0$ is that it serves as a shortcut for the disutility from the higher riskiness of entrepreneurship, which is not otherwise captured here. Scheuer (2013a) accounts for this difference in risk between entrepreneurship and employment explicitly and also demonstrates how the results extend to a framework where entrepreneurship requires investment and borrowing in frictional credit markets. This turns out to provide an additional argument for differential taxation of entrepreneurial profits: it is able to mitigate occupational misallocation that results from adverse selection in credit markets.

An alternative way of generating heterogeneity in entry into entrepreneurship conditional on income would be through a Roy (1951) model, where individuals have a two-dimensional skill type, with one (latent) skill for entrepreneurship and another for employed work. Rothschild and Scheuer (2013) have analyzed optimal income taxation in such a framework, generalizing it to an economy with different occupations, whose

[25] Another possibility would be that actual incomes do not in fact drop, and reported incomes drop because of easier tax avoidance by the self-employed.

effective wages depend on the relative employment in the sectors through some aggregate production function.[26] They derive similar results for the case of uniform taxation, but do not consider differential taxation of different occupations. An individual's skill as worker and entrepreneur need not be comparable in their extension, relaxing the assumption made here that θ captures an individual's skill in both occupations and the second dimension of heterogeneity enters preferences only additively.

5.3 General Equilibrium and Spillover Effects

As the analysis in Section 3 has emphasized, in the present model, the optimal shape of a uniform tax schedule crucially depends on the substitution elasticity between labor and entrepreneurial effort. With linear technology, wages do not respond to the relative magnitudes of entrepreneurial effort and labor, and hence the standard properties of the tax schedule obtain. However, if there are complementarities between the two, even with the constant returns to scale technology assumed here throughout, the optimal tax policy explicitly manipulates incentives for effort and entry into entrepreneurship to affect relative wages and achieve "trickle down" effects.

I do not attempt to quantify the importance of such complementarities in this paper. However, I suspect that similar effects would obtain in alternative settings. For instance, if different firms produce different goods and there is monopolistic competition (as for instance in Dixit and Stiglitz, 1977), then entering firms would introduce new products and thereby increase employment and the workers' real wage, leading to the same effect as here even though through a different channel. However, an analysis of optimal income taxation would be more involved in such a setting due to the endogenous number of different consumption goods and inefficiencies from imperfect competition.

Another complication that would arise in such a framework is the effect of externalities from entrepreneurship on optimal tax policy. Indeed, if there are positive spillover effects from new businesses, e.g. through innovation, so that entrepreneurs are unable to capture the entire social benefits they generate, there is yet another reason for a differential tax treatment of entrepreneurs and workers: it has to trade off the Pigouvian motives for subsidizing entrepreneurial effort with the redistributive motives for taxing it. There is little work on Pigouvian taxation in Mirrleesian settings with privately observed skill heterogeneity, let alone the multidimensional heterogeneity considered here. The contribution by Rothschild in Scheuer (2011) has addressed these issues in a Roy-type model with two types of activities, a traditional one that exerts no externalities and an extractive

[26] They also allow for general preferences with income effects.

one (interpreted as rent-seeking) that generates negative externalities. How their results extend to settings where some activities generate positive externalities, as would be natural to assume for some entrepreneurs, is subject to ongoing research.

5.4 Further Issues

The analysis in this paper has ignored several other aspects of entrepreneurship and its implications for tax policy. Notably, capital accumulation and additional choices available to entrepreneurs, such as the decision whether to incorporate or not, have been abstracted from. Accounting for these aspects would give rise to issues such as the reclassification of entrepreneurial labor income into (favorably taxed) capital income, and the choice of compensation through salaries, dividends or shares. These margins may be as important as the decision whether to set up a new business, especially for large firms where ownership and management/entrepreneurship are separated. Even though I expect the key mechanisms for optimal tax policy emphasized here to extend to such richer settings, a more comprehensive exploration of these issues is 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 (2010): "Entrepreneurship: Origins and Returns," *Labor Economics*, 18, 180–193.

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.

DIXIT, A., AND J. STIGLITZ (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297–308.

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.

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.

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. (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 (2011): "Optimal Taxation with Rent-Seeking," *NBER Working Paper 17035*.

——— (2013): "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 128, 623–668.

ROY, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

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

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

SCHEUER, F. (2013a): "Adverse Selection in Credit Markets and Regressive Profit Taxation," *Journal of Economic Theory*, 148, 1333–1360.

——— (2013b): "Entrepreneurial Taxation with Endogenous Entry," *NBER Working paper No. 19235*.

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.

TERGIMAN, C. (2013): "Entrepreneurship Does Pay," *Mimeo, Univeristy of British Columbia*.

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) Consider first the case $\underline{\phi}_\theta \geq 0$ for all θ. Analogously to the proof of Proposition 1, if $\tilde{w} \leq w$, then $\tilde{\phi}(\theta) = \underline{\phi}_\theta$ 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 (13) and (14) 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. The case $\overline{\phi}_\theta \leq 0$ is analogous.

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) Let me recapitulate the Pareto problem as follows:

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

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

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

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

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

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

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

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

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

The transversality conditions are $\mu_E(\underline{\theta})=\mu_E(\overline{\theta})=\mu_W(\underline{\theta})=\mu_W(\overline{\theta})=0$. 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 and considering the case $\tilde{w}>w$, the necessary conditions for $E(\overline{\theta})$ and $l(\underline{\theta})$ are

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

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),

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

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 $T'(\tilde{w}E(\overline{\theta}))=T'(wl(\underline{\theta}))=0$ at any Pareto-optimum if technology is linear. In the case $\tilde{w}<w$, the necessary conditions for $E(\underline{\theta})$ and $l(\overline{\theta})$ are analogous to (21) and (22), replacing $\overline{\theta}$ by $\underline{\theta}$ in (21) and vice versa in (22). This then yields $T'(\tilde{w}E(\underline{\theta}))=T'(wl(\overline{\theta}))=0$.

(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 $T'(\tilde{w}E(\bar{\theta})) = T'(wl(\underline{\theta})) = 0$ for $\tilde{w} < w$. Otherwise, $T'(\tilde{w}E(\underline{\theta})) = T'(wl(\bar{\theta})) = 0$ by an analogous argument.

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 $T'(\tilde{w}E(\bar{\theta})) < 0$ and $T'(wl(\underline{\theta})) > 0$ for $\tilde{w} > w$. Analogously, if $\tilde{w} < w$, $T'(\tilde{w}E(\underline{\theta})) < 0$ and $T'(wl(\bar{\theta})) > 0$.

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

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

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 $E(\bar{\theta})$ become (using $\tilde{w} > w$)

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

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

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

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

with $a = \underline{\theta}$ and $b = (Y_L(x)/Y_E(x))\bar{\theta}$. 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

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

and

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

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

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

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

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

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

which implies

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

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. The case with $\tilde{w} < w$ is completely analogous and yields $T'(\tilde{w}E(\underline{\theta})) = T'(wl(\overline{\theta})) = 0$.

(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

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

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 (ND”) is slack either at $\overline{\theta}$ or $\underline{\theta}$, which yields $\kappa_E(\overline{\theta}) = 0$ or $\kappa_E(\underline{\theta}) = 0$. Then (20) implies (together with $\kappa_L(\overline{\theta}) = 0$) that $Y_L(\overline{\theta}) = \lambda_{LM}$ or $Y_L(\underline{\theta}) = \lambda_{LM}$. However, whenever there exists some $\theta \notin \{\underline{\theta}, \overline{\theta}\}$ 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

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

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), $\tilde{w}E(\theta) - \psi\left(E(\theta)/\theta\right)$ falls and $wl(\theta) - \psi\left(l(\theta)/\theta\right)$ 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 (13) 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 (14) 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

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

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 $T_\pi(z) > T_y(z)\ \forall z$.