Entrepreneurial Taxation, Occupational Choice, and Credit Market Frictions

Florian Scheuer*
MIT

May 2010

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 effects through wages to indirectly achieve redistribution between entrepreneurs and workers. As a result, constrained Pareto optimal policies can involve negative marginal tax rates at the top and, if available, input taxes that distort the firms' input choices. However, these properties disappear when a differential tax treatment of profits and labor income is possible, as for instance implemented by a corporate income tax. In this case, redistribution is achieved directly through the tax system rather than "trickle down" effects, and production efficiency is always optimal. When I extend the model to incorporate entrepreneurial borrowing in credit markets, I find that endogenous cross-subsidization in the credit market equilibrium results in excessive (insufficient) entry of low-skilled (high-skilled) agents into entrepreneurship. Even without redistributive objectives, this gives rise to an additional, corrective role for differential taxation of entrepreneurial profits and labor income. In particular, a regressive profit tax may restore the efficient occupational choice.

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

1 Introduction

The question at what rate business profits should be taxed – notably relative to the tax rates on other forms of income such as labor earnings – is a recurring and controversial theme in the public policy debate. On the one hand, it is often argued that individuals who receive business profits, such as entrepreneurs, tend to be better off than those who do not. Therefore, arguments based on direct redistribution, or "tagging," seem to justify the taxation of profits at a higher rate than other forms of income, as for instance implemented by a corporate income tax and the resulting double taxation of profits both at the firm and individual level. On the other hand, proponents of "supply side" or "trickle down economics" typically emphasize the general equilibrium effects of the tax treatment of businesses. In particular, they point out that a reduction in the entrepreneurs' tax burden encourages entrepreneurial activity and labor demand. It thereby increases wages and hence "trickles down" to medium or lower income workers, achieving redistribution indirectly. Moreover, entrepreneurs invest and therefore have to borrow funds in credit markets that are typically subject to imperfections. This raises concerns about a too low number of individuals setting up firms due to borrowing constraints from credit market frictions. From both of the latter perspectives, a reduced taxation of firm profits, or even a subsidization of entrepreneurial activities, appears optimal.

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

I characterize Pareto optimal allocations in this economy and demonstrate that the resulting multidimensional screening problem is tractable and allows for a transparent analysis of the issues raised above. The key result is that it crucially depends on the set of available tax instruments whether a Pareto optimal tax system uses general equilibrium effects to achieve redistribution indirectly through "trickle down." I start with characterizing constrained Pareto optimal allocations when the government imposes the same, non-linear tax schedule on both entrepreneurial profits and labor income. In fact, this appears particularly appealing in view of the general presumption that introducing wedges between different forms of income is distorting and should therefore be avoided.

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.

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 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.

I compare the optimal tax schedules for profits and labor earnings in an economy that is calibrated to match income distributions and occupational choice between entrepreneurship and employment in the 2007 Survey Consumer Finances. Under various assumptions on the government's redistributive objectives, there robustly emerges an "excess profit tax," i.e. a higher taxation of entrepreneurial profits compared to labor income for individuals of the same skill level, as for instance implemented by a corporate income tax or a separate tax schedule for self-employed persons. I also simulate the effects of optimal tax policy on wages and entrepreneurship for various parameter combinations, with the finding that wages decrease and entrepreneurship is discouraged for most skill levels even when the government only aims at redistributing across different ability types, not between entrepreneurs and workers directly.

Finally, I introduce entrepreneurial investment and borrowing into the analysis. Individuals are assumed to be wealth constrained and therefore have to borrow funds from banks in a competitive credit market in order to set up a firm. Since there is privately known heterogeneity, credit markets are affected by adverse selection. This raises the concern that, due to the credit market imperfections, an insufficient number of agents choose to set up a firm, calling for a lower taxation, or even a subsidization, of entrepreneurship compared to the preceding analysis. I characterize the credit market equilibrium and show that it takes the form of a pooling equilibrium that involves a single debt contract being offered to all entrepreneurs. The resulting cross-subsidization from high to low quality borrowers provides excessive incentives for low skilled individuals to enter entrepreneurship, but insufficient incentives for high ability agents.

Even without redistributive objectives, credit market frictions therefore give rise to an additional, corrective role for a differential tax treatment of entrepreneurial profits. In particular, the occupational misallocation can be removed by a regressive profit tax, which involves negative average tax payments for high-skilled entrepreneurs and positive ones for low ability entrepreneurs, thus counteracting the cross-subsidization in the credit market equilibrium. This demonstrates that credit market imperfections do not necessarily justify a general subsidization of entrepreneurship, as raised at the beginning.

Rather, they induce the wrong mix of individuals in the two occupations, with too many and too few entrepreneurs at the same time. More sophisticated tax interventions, such as the regressive profit tax suggested here, are therefore required to mitigate inefficiencies in occupational choice.

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.[1]

In characterizing optimal allocations, my work therefore shares a common goal with Albanesi (2006) and, more relatedly, Albanesi (2008), who has extended the framework of optimal dynamic taxation to account for entrepreneurial investment. More precisely, she considers a moral hazard model where entrepreneurs exert some hidden action that affects a stochastic return to capital. Her focus is on characterizing the optimal savings distortions that entrepreneurs should face when the government provides insurance for entrepreneurial investment risk. Similarly, Chari, Golosov, and Tsyvinski (2002) examine optimal intertemporal wedges in a dynamic economy with start-up firms and incomplete markets. In contrast to this literature, I focus on characterizing the optimal taxation of profits and labor income in a static general equilibrium model that emphasizes how taxes affect the effort-leisure wedge of entrepreneurs versus workers and thus wages. Moreover, when incorporating entrepreneurial borrowing, I explicitly consider private credit markets and how tax policy interacts with the endogenous credit market equilibrium.

In this regard, the present paper is also related the literature on the role of government

[1]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.

intervention in credit markets with adverse selection, e.g. in Stiglitz and Weiss (1976), De Meza and Webb (1987), Innes (1991), Innes (1992), and Parker (2003). While this research has pointed out efficiency properties of credit market equilibria and scope for government intervention, there has been no systematic treatment of optimal entrepreneurial taxation in the presence of such credit market frictions and occupational choice. Most related is the contribution by De Meza and Webb (1987), who point out the possibility that adverse selection in credit markets leads to excessive entry into entrepreneurship, quite in contrast the credit rationing emphasized by Stiglitz and Weiss (1976). They suggest a tax on bank profits to deal with this inefficiency. Their result is a special case of the present setting when the second dimension of heterogeneity is removed. With two-dimensional heterogeneity, however, it turns out that there is excessive and insufficient entry into entrepreneurship simultaneously, so that a simple tax on bank profits is not sufficient (nor necessary) to correct the occupational misallocation. The analysis here therefore points at the role of entrepreneurial tax policy in undoing cross-subsidization in credit markets and restoring efficiency of occupational choice.

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. Most related to the formal modelling approach used here is the recent contribution by Kleven, Kreiner, and Saez (2009) with an application to the optimal income taxation of couples. More generally, this paper builds on the large literature on optimal income taxation following the seminal contributions by Diamond and Mirrlees (1971) and Mirrlees (1971). However, rather than focusing on allocations that maximize some utilitarian social welfare criterion, I aim at characterizing the set of Pareto optimal tax policies, sharing the spirit of Werning (2007).

The structure of the paper is as follows. Section 2 introduces the baseline model and the equilibrium without taxation. In Section 3, I start with characterizing Pareto optimal tax policies when the same (non-linear) tax schedule is applied to both entrepreneurial profits and labor income. Properties of Pareto optimal tax schedules and the optimality of production distortions are discussed. As I show in Section 4, these properties disappear when profits and income can be made subject to different tax schedules. Section 4 also computes the two tax schedules for a calibrated economy. Section 5 then introduces entrepreneurial borrowing in credit markets, and shows that this gives rise to another, corrective role for entrepreneurial taxation. Finally, Section 6 concludes. Most of the proofs are relegated to the appendix.

2 The Baseline Model

2.1 Preference Heterogeneity and Occupational Choice

I consider a unit mass of heterogeneous individuals who are characterized by a two-dimensional type vector $(\theta, \phi) \in [\underline{\theta}, \overline{\theta}] \times [0, \overline{\phi}_\theta]$, where θ will be interpreted as an individual's skill, and ϕ as an individual's cost of becoming an entrepreneur, as explained in more detail below.[2] $F(\theta)$ 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

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

respect to the wage is constant and equal to ε. θ captures an individual's skill type in the sense that a higher value of θ implies that the individual has a lower disutility of providing a given amount of effective labor l.

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

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

and her utility is given by

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

ϕ is a heterogeneous utility cost of becoming an entrepreneur, which is distributed in the population as specified above, possibly depending on the skill type θ. Thus, θ determines an individual's skill in both occupations, but in addition, people differ in their idiosyncratic preferences for one of the two occupations, as captured by ϕ. The cost ϕ can therefore be interpreted as a shortcut for heterogeneity in the population that is not otherwise captured in the present model explicitly, such as a differences in setup costs, attitudes towards entrepreneurial risks, or access to entrepreneurial capital (see Section 5 for more on the latter).[3] 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 it is true that, in reality, the income distributions of workers and entrepreneurs have overlapping supports.[4]

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

[4]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)).

2.2 The Equilibrium without Taxes

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

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

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

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

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

In fact, the entrepreneurs' utility maximization problem can be decomposed as follows. Since their labor demand L only affects profits and not the other components of their utility, for given E and w, entrepreneurs of all types θ solve the same problem $\max_L Y(L, E) - wL$ with the conditional labor demand function $L^c(E, w)$ as solution such that $Y_L(L^c(E, w), E) = w$. Under the assumption of constant returns to scale, Euler's 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 $\tilde{w}(w)$:[5] The entrepreneurs' wage $\tilde{w}$ is high if the en-

[5] This is because, by linear homogeneity of Y, both Y_L and Y_E are homogeneous of degree zero and hence

trepreneurial effort to labor ratio used in production is low, which means that the marginal product of labor and thus the workers' wage is low.

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

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

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

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

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

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

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

Proposition 1 summarizes intuitive properties of wages and occupational choice in equilibrium: First, the entrepreneurs' wage $\tilde{w}^*$ must be higher than that of the workers w^* in equilibrium. The reason is that, when deciding whether to become a worker or an entrepreneur, an individual of a given skill type considers two variables: The different wage that she can earn when becoming an entrepreneur rather than a worker, and the cost ϕ she has to incur when doing so. Clearly, if the entrepreneurs' wage were lower than that of workers, there would be no trade-off and nobody would choose to enter entrepreneurship, which cannot be an equilibrium. The entrepreneurs' higher wage then immediately implies that they exert more effort and earn higher profits than workers of the same ability level. While this is a direct consequence of the assumption that $\phi \geq 0$, it is in line with empirical evidence on returns to entrepreneurship. For instance, De Nardi,

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).

Doctor, and Krane (2007) find that entrepreneurs have higher incomes than workers, and Berglann, Moen, Roed, and Skogstrom (2009) confirm this pattern for wages, controlling for hours. Moreover, based on data from the 2007 Survey of Consumer Finances (SCF), I find the same relationship between returns to entrepreneurship and employment, as will be discussed in Section 4.2.[6] In addition, since this is a static model, Proposition 1 can be interpreted in terms of lifetime incomes, or wealth. There is strong evidence that entrepreneurs have more wealth than workers, for instance in Quadrini (2000) and Cagetti and De Nardi (2006).

The second result in the proposition is that, the higher the skill type θ, the more the wage difference matters compared to the cost, which is why the critical cost value $\tilde{\phi}(\theta, w^*)$ increases with θ. Finally, the same holds for the share of entrepreneurs in equilibrium as a function of skill whenever skill and disutility from entrepreneurship are independent or such that higher skills tend to have a lower disutility from being an entrepreneur in the first-order stochastic dominance sense. More generally, while such a correlation between θ and ϕ may strike as plausible, the model is flexible enough to generate more complicated relationships between income and the share of entrepreneurs through the dependence of the cost distribution on θ, as captured by $G_\theta(\phi)$.[7] Proposition 1 thus demonstrates that, while the basic model is admittedly stylized and quite different from other models of entrepreneurship, it is able to produce reasonable predictions about empirical relationships, and to point out how they depend on the underlying heterogeneity in the population.

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 dis-

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

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

tort the occupational choice margin. 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' such that $\tilde{w}\theta = w\theta'$ are the same:

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

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 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.

By the revelation principle, any allocation that can be implemented with the single non-linear tax schedule $T(.)$ must therefore satisfy the no-discrimination constraints (3) and the incentive compatibility constraints as specified in the following. 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.[8] Then the incentive constraints can be written as

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

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

[8]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.

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

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

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

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

Finally, incentive compatibility requires that the critical cost values for occupational choice are given by

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

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

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

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

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

subject to

$$\int_{\Theta} G_{\theta}(\tilde{\phi}(\theta))L(\theta)dF(\theta) \leq \int_{\Theta}(1 - G_{\theta}(\tilde{\phi}(\theta)))l(\theta)dF(\theta), \tag{11}$$

$$\begin{aligned}\int_{\Theta} G_{\theta}(\tilde{\phi}(\theta))\left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi(E(\theta)/\theta)\right]dF(\theta)& \\ - \int_{\Theta}(1 - G_{\theta}(\tilde{\phi}(\theta)))\left[v_W(\theta) + \psi(l(\theta)/\theta)\right]dF(\theta) &\geq 0,\end{aligned} \tag{12}$$

and constraints (3), (6), (7), (8), (9) and (10). Inequality (11) requires the total amount of labor demand assigned to entrepreneurs not to exceed the total amount of labor supply assigned to workers. Similarly, (12) 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.[11]

3.2 Properties of Constrained Pareto Optimal Tax Systems

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

Proposition 2. *(i) At any Pareto-optimum, $\tilde{w} > w$, and $\tilde{w}E(\theta) > wl(\theta)$ for all $\theta \in \Theta$.*
(ii) $T'(wl(\underline{\theta})) = T'(\tilde{w}E(\overline{\theta})) = 0$ if $Y(L, E)$ is linear.
(iii) Otherwise, $T'(wl(\underline{\theta}))$ and $T'(\tilde{w}E(\overline{\theta}))$ have opposite signs whenever (3) binds for some $\theta \in \Theta$.

Proof. See Appendix A.1. □

The first part of Proposition 2 holds for the same reason as in the equilibrium without taxes: Since profits and labor income are subject to the same tax treatment, the en-

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

trepreneurs' marginal product must be higher than the workers', because otherwise nobody would choose to set up a firm. This implies that the top earner at any Pareto optimum is an entrepreneur, and the bottom earner a worker.[12]

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, and since part (i) has shown that the set of top earners is exclusively given by entrepreneurs and the lowest income is only earned by workers, the optimal tax schedule involves a negative marginal tax rate at the top and a positive marginal tax rate at the bottom in this case.[13] 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 pre-

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

[13] If, by contrast, the Pareto-weights are such that the no-discrimination constraints are relaxed by increasing $\tilde{w}/w$, the opposite pattern holds.

venting high skill agents from imitating low skill agents.[14] 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 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 skill individuals.[15]

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

Proposition 3. *(i) If, in addition to the non-linear tax $T(.)$ on profits and income, the government can impose a proportional tax on the firms' labor input, then a Pareto-optimal tax system satisfies $T'(wl(\underline{\theta})) = T'(\tilde{w}E(\bar{\theta})) = 0$.*
(ii) Moreover, if the government can distort $Y_L(L(\theta), E(\theta))$ across firms, e.g. through a non-linear tax on labor input, then it is optimal to do so whenever $Y(L, E)$ is not linear and the no-discrimination constraints (3) 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) from the marginal products of entrepreneurial effort and labor in constraints (6) and (7), so that there remains no need to affect them through the nonlinear tax schedule $T(.)$. As a result, the top and bottom marginal tax rates are again zero at any Pareto optimum, even if technology is not linear.

Whereas a pure profit tax, even when complemented by a proportional tax on labor inputs, always implies that marginal products of labor (and thus of entrepreneurial effort) are equalized across all firms, part (ii) shows that such production efficiency is not necessarily optimal in this framework. Intuitively, by distorting marginal products of labor and effort across firms, e.g. through a non-linear tax on the firms' labor input, the

[14] 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.

[15] In section 4, I provide conditions that pin down the direction in which the no-discrimination constraints bind, and the optimal top marginal tax rate is indeed negative. They essentially require Pareto-weights such that redistribution from low-ϕ agents to high-ϕ agents is desirable, and thus from entrepreneurs to workers.

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 at that skill level. Then the government faces a trade-off between production efficiency and relaxing the no-discrimination constraints, which generally involves some degree of production inefficiency at the optimum.[16]

4 Differential Tax Treatment of Profits and Income

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

4.1 A Theoretical Characterization

4.1.1 Pareto Optimal Tax Formulas

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

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

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

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

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

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

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

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.[17]

[17] 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.

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

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

(and analogously for income) and the mass of entrepreneurs $f(\theta)G_\theta(\tilde{\phi}(\theta))$ at θ (this mass is $f(\theta)(1 - G(\tilde{\phi}(\theta)))$ for workers). This accounts for the local effort (labor supply) distortion generated by the marginal tax. The first two terms in the integral, in turn, capture the redistributive effects of the tax schedule, comparing the mass of Pareto-weights $\tilde{G}_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\tilde{f}(\hat{\theta})$ for all skill types $\hat{\theta}$ below θ to that of the population densities $G_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})$ (and again equivalently for workers). The last term in the integrals, finally, captures the effect of differential profit and labor income taxation on occupational choice. Specifically, the mass of agents of skill θ driven out of entrepreneurship by an infinitesimal increase in profit taxation T_π is given by $g_\theta(\tilde{\phi}(\theta))f(\theta)$, i.e. those individuals who were just indifferent between entrepreneurship and employment before the change. The resulting effect on the government budget is captured by the excess entrepreneurial tax $\Delta T(\theta)$, which is the additional tax payment by an entrepreneur of type θ compared to a worker of the same skill. Of course, this budget effect appears with opposite signs in the optimality formulas for the entrepreneurial profit and labor income tax schedule.[18]

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

[18] 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.

taxation must be such that the bottom and top marginal tax rates for both workers and entrepreneurs are zero, any allocation that does not satisfy these properties must be Pareto inefficient. But Proposition 2 has shown that, whenever uniform taxation leads to binding no-discrimination constraints, the bottom and top marginal tax rates are not zero. Hence, starting from such an allocation, there must exist a Pareto improvement using differential taxation.

The following result is an immediate corollary of Proposition 4.

Corollary 1. *With a constant elasticity ε,[19] the average marginal tax across occupations satisfies*

$$G_\theta(\tilde{\phi}(\theta))\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))}+(1-G_\theta(\tilde{\phi}))\frac{T'_y(y(\theta))}{1-T'_y(y(\theta))}=\frac{1+1/\varepsilon}{\theta f(\theta)}\left(\tilde{F}(\theta)-F(\theta)\right). \qquad (13)$$

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 (13): 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 (13) 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.[20]

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

[19] Even without a constant elasticity, a modified version of (13) holds, with 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.

[20] 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 (13) cannot be obtained.

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

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

$$\frac{\theta f_E(\theta)}{1+1/\varepsilon}\,\frac{T'_\pi(\pi(\theta))}{1-T'_\pi(\pi(\theta))} + F_E(\theta) - \int_{\underline{\theta}}^{\theta} \frac{g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})}{\overline{G}} \Delta T(\hat{\theta})d\hat{\theta} \quad \textit{and} \tag{14}$$

$$\frac{\theta f_W(\theta)}{1+1/\varepsilon}\,\frac{T'_y(y(\theta))}{1-T'_y(y(\theta))} + F_W(\theta) + \int_{\underline{\theta}}^{\theta} \frac{g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})}{1-\overline{G}} \Delta T(\hat{\theta})d\hat{\theta} \tag{15}$$

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

Two remarks on Corollary 2 are in order. First, adding conditions (14) and (15) 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 (14) and (15) occurs when there is no occupational choice, so that whether an individual is an entrepreneur or a worker is a fixed characteristic. This can be thought of as a special case of the general formulation considered so far, where the cost ϕ has a degenerate distribution with only two mass points, at 0 and $\overline{\phi}$, and $\overline{\phi}$ is sufficiently high. Then T_π must be such that

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

is non-decreasing, and analogously for T_y replacing the (fixed) skill distribution for entrepreneurs by that for workers, $F_W(\theta)$. This coincides with the integral version of the condition derived in Werning (2007) for a standard Mirrlees model. Hence, the key difference arising from the present framework are the terms $-\int_{\underline{\theta}}^{\theta} g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})\Delta T(\hat{\theta})d\hat{\theta}/\overline{G}$ and $\int_{\underline{\theta}}^{\theta} g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})\Delta T(\hat{\theta})d\hat{\theta}/(1-\overline{G})$, reflecting the effects of taxation on occupational choice and thus on the resource constraint. Note that, since these terms enter conditions (14) and (15) 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 $\tilde{\phi}(\theta)$).

4.1.3 Comparing Optimal Profit and Income Tax Schedules

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

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

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

Proposition 5. *Suppose that $\tilde{F}(\theta)=F(\theta)$, $\tilde{g}(\phi)<g(\phi)$ for all $\phi\leq\tilde{\phi}(\overline{\theta})$ and Assumption 1 holds. Then*
(i) $T'_y(y(\theta))<0$, $T'_\pi(\pi(\theta))>0$ for all $\theta\in(\underline{\theta},\overline{\theta})$,
(ii) $\Delta T(\theta)>0$ and $\Delta T'(\theta)>0$ for all $\theta\in\Theta$, and
(iii) compared to the no tax equilibrium, w decreases, $\tilde{w}$ increases and $L(\theta)/E(\theta)$ rises for all $\theta\in\Theta$.

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.[21] 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.

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

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

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

[21] This implies that $E(\theta) < (w/\tilde{w})l((\tilde{w}/w)\theta)$ for all $\theta \in \Theta$. Hence, under the assumptions in Proposition 5 and Assumption 1, the no-discrimination constraints (3) in the previous Section 3 all bind in the same direction and such that the optimal restricted tax schedule involves a positive bottom and a negative top marginal tax rate.

occupations. This makes clear that any difference in the optimal tax schedules for profits and income must be the result of an active occupational choice margin or a non-zero correlation between ability and occupational choice, which will be further explored in the subsequent numerical simulations.

4.2 A Quantitative Exploration

To further explore the importance of a differential tax treatment of profits and income, I provide a quantitative illustration of the analysis so far by computing optimal tax systems under various redistributive objectives. Notably, the formulas in Proposition 4 can be used to compute the tax schedules T_π and T_y once distributions for θ and ϕ, Pareto-weights, a production function and preferences are specified.[22] To calibrate the model, I use data on income, profits, and entrepreneurship from the 2007 Survey of Consumer Finances (SCF). I restrict the sample to household heads aged between 18 and 65 who are not unemployed/retired, and define the empirical counterpart of entrepreneurs in my model as those individuals who (i) are self-employed, (ii) own a business, (iii) actively manage it, and (iv) employ at least two employees. This is a widely used empirical definition of the notion of entrepreneurship.[23] All other individuals in the sample are considered as workers.

Table 1 presents descriptive statistics of the resulting sample. A share of 7.4% of the sample ends up being classified as entrepreneurs according to the above criteria. Consistent with the theoretical findings so far, entrepreneurs have higher incomes than workers, even though the higher means come at the price of a higher income variability than for workers, as captured by the standard deviation. This suggests that entrepreneurship is more risky than employment, an aspect that will be accounted for explicitly in the next section. Entrepreneurs also work more than workers, as measured by yearly hours. Still, their wage, computed as the ratio of yearly income and hours, is higher than that of workers. This is consistent with the evidence on entrepreneurial incomes and wages in De Nardi, Doctor, and Krane (2007) and Berglann, Moen, Roed, and Skogstrom (2009).[24]

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

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

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

Table 1: Descriptive Statistics

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

For the baseline calibration, I work with the following parametric specifications: The disutility of effort takes the iso-elastic form $\psi(e) = e^{1+1/\varepsilon}/(1+1/\varepsilon)$, and, based on the empirical labor supply literature, the wage elasticity of effort is set to be $\varepsilon = .25$. The constant returns to scale technology used by entrepreneurs is captured by a Cobb-Douglas production function $Y(L, E) = L^{\alpha}E^{1-\alpha}$ with the parameter α set to equal the workers' share of income in the SCF data, so that $\alpha = .63$.

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

$$\tau(y) = b - b\left[sy^{p} + 1\right]^{-1/p}. \tag{16}$$

The parameters b, s and p are estimated by Cagetti and De Nardi (2009) using PSID data for entrepreneurs and workers separately, with point estimates $b = .26$, $s = .42$ and $p = 1.4$ for entrepreneurs and $b = .32$, $s = .22$ and $p = .76$ for workers. Then I obtain marginal tax rates from the average tax rates in (16). With this information, I am able to identify $w\theta$ for workers and $\tilde{w}\theta$ for entrepreneurs from the first order conditions of the individuals' utility maximization problem

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

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

The left panel of Figure 1 depicts a kernel estimate of the resulting inferred skill density, truncated at the 99 percentile. The smoothed approximation of it, also depicted, is

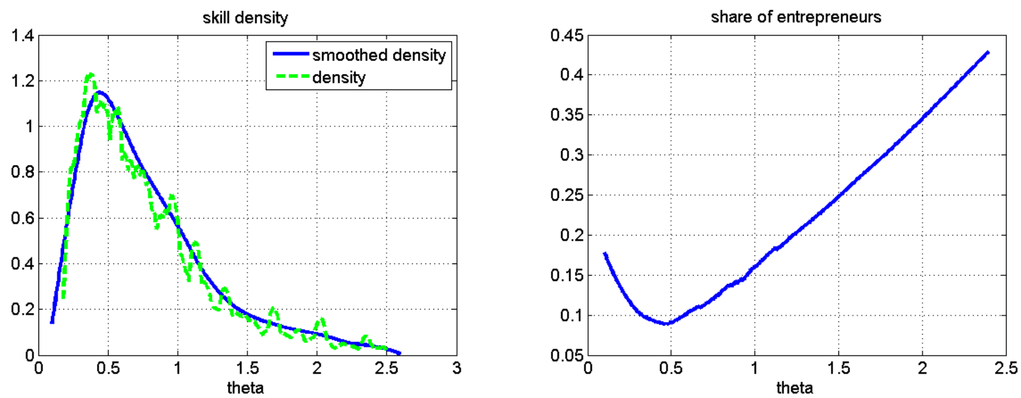

Figure 1: Skill density and share of entrepreneurs

used as $f(\theta)$ in the simulations to obtain smoother optimal tax schedules. The right panel in turn shows the share of entrepreneurs as a function of the skill level θ, which is the result of a locally weighted regression of the indicator variable for whether an individual is an entrepreneur on θ. As can be seen from the graph, the share of entrepreneurs is increasing in θ, except for the lowest skill levels.[25] I use this pattern to calibrate the cost distribution $G_\theta(\phi)$. In particular, I assume an iso-elastic specification with $G_\theta(\phi) = (\phi/\overline{\phi}_\theta)^\eta$ and $\eta = .5$, implying an elasticity of occupational choice of .5. Then the upper bound of the support $\overline{\phi}_\theta$ is adjusted to generate the pattern of the share of entrepreneurs in the right panel of Figure 1.

Figure 2 starts with the case of redistribution across cost types only, and hence from entrepreneurs to workers, with Pareto weights $\tilde{F}(\theta) = F(\theta)$ and $\tilde{G}_\theta(\phi) = G_\theta(\phi)^{\rho_\Phi}$, $\rho_\Phi = 2$. It depicts the marginal tax schedules T'_π and T'_y, the tax schedules T_π and T_y, the excess profit tax ΔT, and the share of entrepreneurs $G(\tilde{\phi}(\theta))$ as a function of skill, both for the no tax equilibrium as well as for the case with taxation. The figure illustrates the results from Proposition 5: The marginal tax rates for entrepreneurs are positive, for workers negative, and there is a positive and increasing excess profit tax. Entry into entrepreneurship is discouraged for individuals of all skill levels compared to the no tax equilibrium. Moreover, the workers' wage falls by 11% from the no tax equilibrium to the equilibrium with taxation.

Figure 3 illustrates the other benchmark case, when the Pareto weights are such that redistribution across skill types only is implied. In particular, it assumes $\tilde{F}(\theta) = F(\theta)^{1/\rho_\Theta}$ with $\rho_\Theta = 2$. In this case, both marginal tax schedules are positive (as is the average

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

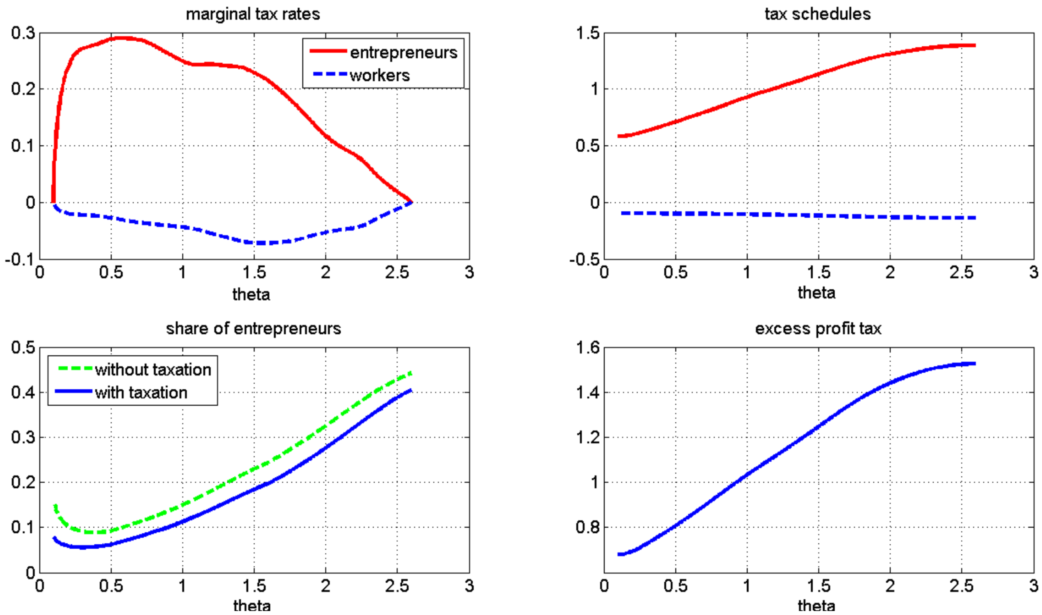

Figure 2: Pareto weights $\tilde{G}_\theta(\phi) = G_\theta(\phi)^{\rho_\Phi}$, $\rho_\Phi = 2$

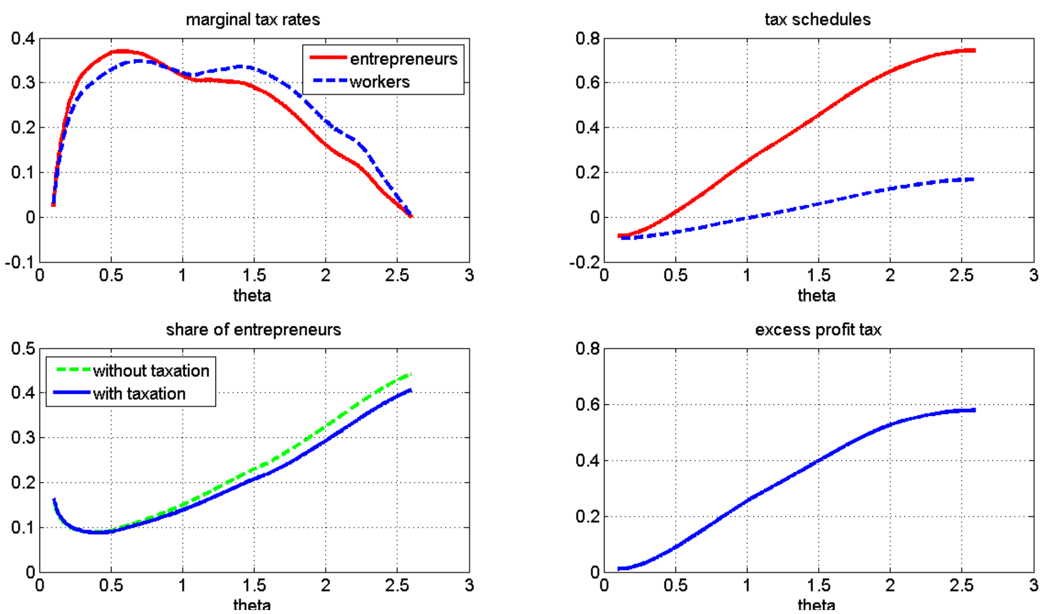

Figure 3: Pareto weights $\tilde{F}(\theta) = F(\theta)^{1/\rho_\Theta}$, $\rho_\Theta = 2$

marginal tax) and such that entrepreneurs of low skill levels face a higher marginal tax rate than workers of the same skill, and the opposite relation at high skill levels. The excess profit tax remains positive and increasing, but is considerably smaller than in the case where redistribution from entrepreneurs to workers directly is desired. Again, the wage falls (by 3% compared to the no tax equilibrium) and entry into entrepreneurship is discouraged slightly, notably for higher skill levels for whom the excess profit tax is higher.

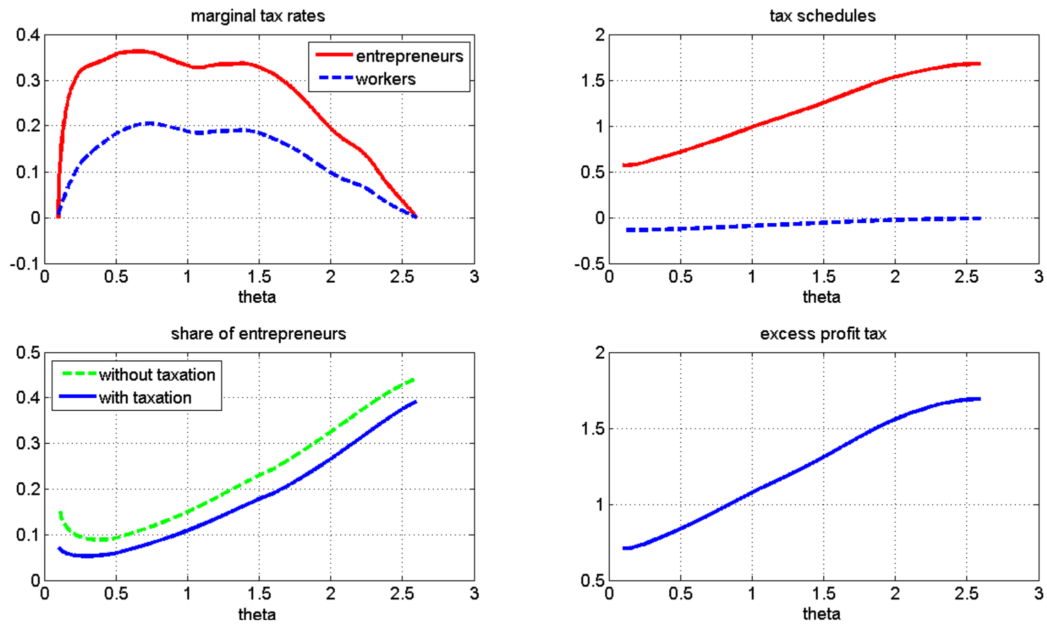

Figure 4: Pareto weights $\rho_\Theta = 2$, $\rho_\Phi = 2$

Figure 4 depicts the solution when the planner aims at redistributing in both dimensions of heterogeneity, so that both $\tilde{G}_\theta(\phi) \leq G_\theta(\phi)$ and $\tilde{F}(\theta) \geq F(\theta)$. Such redistributive objectives turn out to justify entrepreneurs facing both higher levels of taxation as well as higher marginal tax rates than workers for all skill levels, while all agents face positive marginal tax rates. Finally, Figure 5 shows a robustness check from increasing the elasticity of effort from $\varepsilon = .25$ to .5. This makes lower marginal tax rates in both occupations optimal, holding Pareto-weights fixed. Again, the workers' wage falls (by 10%) as a result of the tax policy compared to the no tax equilibrium, entry into entrepreneurship is discouraged, and there emerges a positive and increasing excess profit tax ΔT. All these effects appear as robust properties of optimal tax schedules from these quantitative explorations. Notably, even when the Pareto weights imply that workers should be favored, their wage declines, in contrast to "trickle down" based arguments. Moreover, as a general pattern, the numerical simulations suggest that the difference between the optimal marginal tax rates faced by entrepreneurs and workers increases as more redistribution from entrepreneurs to workers directly is aimed at, compared to redistribution across skill types only.

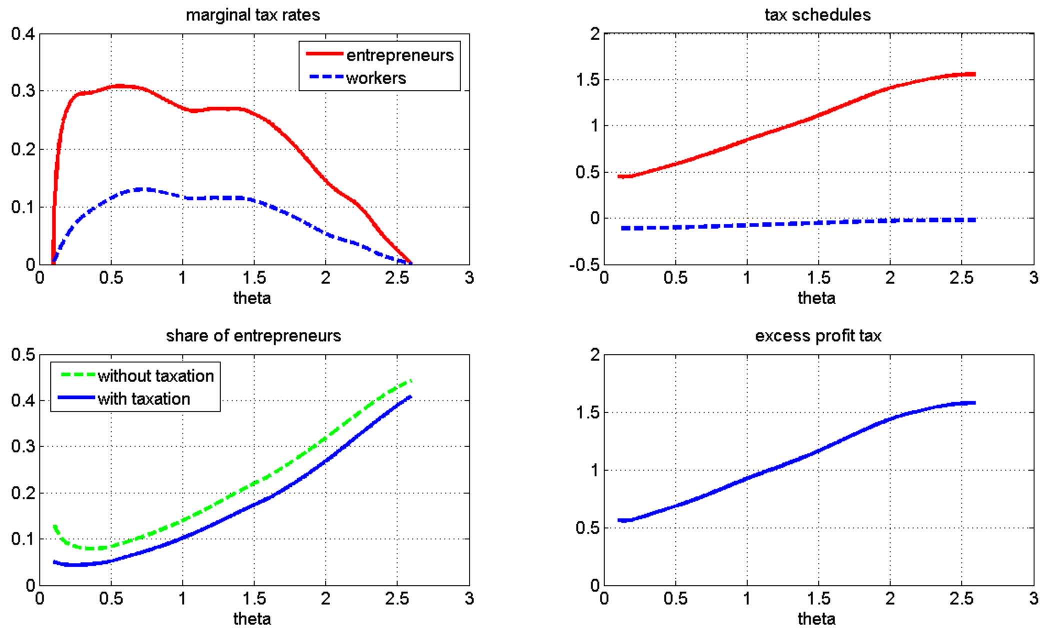

Figure 5: Pareto weights $\rho_\Theta = 2$, $\rho_\Phi = 2$, elasticity $\varepsilon = .5$

5 Entrepreneurial Taxation with Credit Market Frictions

The previous sections have shown that entrepreneurial taxation, in the form of a differential tax treatment of entrepreneurial profits and labor income, plays a role for redistribution and for dealing with pecuniary externalities that arise without such discrimination. However, the analysis so far has ignored the entrepreneurs' investment and borrowing decisions, which are important aspects of entrepreneurial activity. The concern is that, when entrepreneurs have to borrow the funds required for investment in credit markets that are subject to imperfections, then the aggregate level of entrepreneurship may be inefficiently low.[26] This may then generate an economic force that pushes towards a lower taxation of entrepreneurial profits compared to what the previous analysis has implied. However, it turns out that credit market frictions do not necessarily discourage entry into entrepreneurship across the board. While I do find that the no tax equilibrium, which involves adverse selection, is no longer efficient, in contrast to the no tax equilibrium characterized in Section 2.2, it provides incentives to enter entrepreneurship to the wrong mix of individuals, with too many and too few entrepreneurs at the same time. Even without redistributive objectives, this generates scope for a corrective role of entrepreneurial taxation, but it does not take the form of a general subsidization of entrepreneurship, as I

[26] See e.g. Evans and Leighton (1989), Hurst and Lusardi (2004) and De Nardi, Doctor, and Krane (2007) for evidence on the importance of borrowing frictions for entrepreneurship.

show in the following.

5.1 Entrepreneurial Investment and Borrowing

Suppose that, to set up a firm, each entrepreneur has to make a fixed investment I. Yet, agents are born without wealth, and hence have to borrow these funds from banks in a competitive credit market. Since I am mainly interested in the efficiency of occupational choice in the following, I will simplify the model slightly in another dimension and assume that there is no intensive margin, i.e. entrepreneurial effort and labor supply are fixed. More precisely, let all workers supply some fixed amount of labor l and receive utility $v_W = wl$. As before, entrepreneurs hire labor, but now produce stochastic profits

$$\pi = Y(L) - wL + \epsilon,$$

where ϵ has some cdf $H_\epsilon(\epsilon|\theta)$ that depends on the entrepreneurs' skill θ. In particular, I assume that $H_\epsilon(\epsilon|\theta) \succeq_{MLRP} H_\epsilon(\epsilon|\theta')$ for $\theta > \theta'$, i.e. a higher skilled entrepreneur has a distribution of ϵ that is better in the sense of the monotone likelihood ratio property.[27]

There is a large number of risk-neutral banks, offering credit contracts that supply funding I in return for a repayment schedule $R_\theta(\pi)$.[28] The expected utility of an entrepreneur of ability type θ and cost type ϕ from such a contract is then

$$\int_{\mathcal{E}} \left[Y(L) - wL + \epsilon - R_\theta(Y(L) - wL + \epsilon)\right] dH_\epsilon(\epsilon|\theta) - \phi,$$

where $\mathcal{E}$ is the support of H_ϵ. Both θ and ϕ are private information as before.

Let me first observe that, given this specification, for any given set of contracts $\{R_\theta(\pi)\}$, all entrepreneurs hire the same amount of labor such that $Y'(L) = w$, and hence I can work directly with the resulting distribution of profits $\pi \sim H(\pi|\theta)$, with the support denoted by Π. I drop the additional dependence of H and Π on the wage w for notational simplicity. This leads to the following definition of a credit market equilibrium, taking a wage w as given.

[27] A stochastic component of profits is introduced to obtain non-trivial credit market equilibria. Otherwise, competitive banks will have entrepreneurs just repay I after profits have been earned.

[28] I introduce the index θ since, in a separating equilibrium, banks may offer different credit contracts to entrepreneurs of different quality levels θ. Of course, any such assignment has to be incentive compatible, as specified below.

5.2 Credit Market Equilibrium

Definition 2. *A credit market equilibrium is a set of contracts* $\{R_\theta(\pi)\}$ *such that*
(i)

$$\int_\Pi (\pi - R_\theta(\pi))\, dH(\pi|\theta) \geq \int_\Pi (\pi - R_{\theta'}(\pi))\, dH(\pi|\theta) \ \ \forall \theta, \theta' \in \Theta, \tag{17}$$

(ii)

$$\int_\Theta G(\tilde{\phi}(\theta)) \left[\int_\Pi R_\theta(\pi) dH(\pi|\theta) - I \right] dF(\theta) \geq 0 \tag{18}$$

with

$$\tilde{\phi}(\theta) = \int_\Pi (\pi - R_\theta(\pi))\, dH(\pi|\theta) - v_W \ \ \forall \theta \in \Theta, \ \ \textit{and}$$

(iii) there exists no other set of contracts $\{\tilde{R}_\theta(\pi)\}$ *that earns strictly positive profits when offered in addition to* $\{R_\theta(\pi)\}$ *and all individuals select their preferred occupation and preferred contract from* $\{\tilde{R}_\theta(\pi)\} \cup \{R_\theta(\pi)\}$.

(17) is the set of incentive constraints, which require that each entrepreneur is willing to select the credit contract $R_\theta(\pi)$ intended for her. Constraint (18) makes sure that the set of equilibrium credit contracts make non-negative profits *in aggregate* when taken up by the agents who select into entrepreneurship, as given by the critical cost values $\tilde{\phi}(\theta)$ for all $\theta \in \Theta$. Finally, the last part of the definition rules out profitable sets of deviating contract offers. Note that, although the structure of this definition is similar to Rothschild and Stiglitz (1976), it is considerably more general by letting banks offer *sets* of contracts, thus allowing for cross-subsidization between different contracts in equilibrium.

Following Innes (1993), I restrict attention to contracts $R_\theta(\pi)$ that satisfy the following two properties: First, $0 \leq R_\theta(\pi) \leq \pi$ for all $\theta \in \Theta$ and $\pi \in \Pi$, which is a standard limited liability constraint. Second, $R_\theta(\pi)$ is non-decreasing in π so that, when the entrepreneur earns higher profits, the repayment received by the bank $R_\theta(\pi)$ is also higher. This monotonicity constraint can be motivated by noting that banks may have means to reduce firm profits, if they want to. For instance, they may compel the entrepreneur to prove her output level to an impartial third party and to bear the related audit costs, thus reducing the amount of profits available for distribution.[29]

Under these assumptions, the following characterization of a credit market equilibrium as defined in Definition 2 can be obtained.

[29] In addition, it is straightforward to see that whenever the stochastic profit component ϵ has only two possible realizations, any repayment scheme that guarantees the bank non-negative profits in equilibrium must be weakly increasing in π.

Proposition 6. *Under Assumption 1, the credit market equilibrium is such that only the single contract $R_{z^*}(\pi) = \min\{\pi, z^*\}$ is offered and z^* solves*

$$\int_\Theta G(\tilde{\phi}_{z^*}(\theta)) \left[\int_\Pi \min\{\pi, z^*\} dH(\pi|\theta) - I \right] dF(\theta) = 0 \tag{19}$$

with

$$\tilde{\phi}_{z^*}(\theta) = \int_\Pi (\pi - \min\{\pi, z^*\}) \, dH(\pi|\theta) - v_W \quad \forall \theta \in \Theta. \tag{20}$$

Proof. See Appendix C. □

Proposition 6 shows first that an equilibrium always exists. Second, it is such that entrepreneurs of all ability types are pooled in the same contract. Both of these results are very different from the canonical competitive screening model by Rothschild and Stiglitz (1976). In particular, in their model, even when restricting each bank to offer a single contract only, an equilibrium may fail to exist, and it can never take the pooling form. Moreover, an equilibrium as specified in Definition 2, allowing for cross-subsidization, would fail to exist under an even larger set of parameters in the Rothschild-Stiglitz model, as shown by Wilson (1977) and Miyazaki (1977). However, the present model is quite different due to risk-neutrality of all parties and endogenous entry into the credit market.

The second key result in the proposition is that the equilibrium credit contract takes the very simple form of a debt contract: It specifies a fixed repayment level z^*, which the entrepreneur has to return to the bank whenever she can, i.e. whenever $\pi \geq z^*$. Otherwise, the firm goes bankrupt, and the entire amount of profits goes to the bank, with the entrepreneur hitting her liability limit and thus being left with zero consumption. This results in the contract $R_{z^*}(\pi) = \min\{\pi, z^*\}$, where z^* is such that the banks' expected profit is zero given the set of agents who enter the credit market when anticipating the equilibrium contract $R_{z^*}(\pi)$. The intuition for this debt contracting result is as follows:[30] By the monotone likelihood ratio property, low-skill entrepreneurs have a larger probability weight in low-profit states. Among all contracts satisfying the monotonicity constraint, debt contracts in turn are the ones that put the maximal repayment weight in these low-profit states. As a result, debt contracts are least attractive to low-skill borrowers, and hence any set of deviation contracts that do not take the debt contract form would attract a lower quality borrower pool and generate lower profits for banks.

[30] See also Innes (1990) and Innes (1993) for related results in models with moral hazard and one-dimensional private heterogeneity.

5.3 Inefficiency of Occupational Choice and Entrepreneurial Tax Policy

The preceding subsection has characterized credit market equilibria for any given wage w. Of course, an equilibrium in the entire economy is then given by a wage w^* such that the labor market clears, i.e.

$$\int_\Theta G(\tilde{\phi}_{z^*(w^*)}(\theta))L(w^*)dF(\theta) = \int_\Theta (1 - G(\tilde{\phi}_{z^*(w^*)}(\theta)))ldF(\theta), \tag{21}$$

where I have now written $z^*(w)$ to clarify the dependency of the credit market equilibrium on the wage, and where $L(w)$ comes from the entrepreneurs solving $Y'(L) = w$.

I now ask whether the no tax equilibrium in this economy involves the efficient occupational choice. In fact, efficiency would require that a type (θ, ϕ) becomes an entrepreneur if and only if

$$\int_\Pi \pi dH(\pi|\theta) - I - \phi \geq v_W,$$

i.e. her expected profits minus the investment outlays minus the utility cost ϕ exceed the utility from being a worker v_W.[31] This can be solved for the efficient critical cost value

$$\tilde{\phi}_e(\theta) \equiv \int_\Pi \pi dH(\pi|\theta) - I - v_W \tag{22}$$

for any $\theta \in \Theta$. Then the following result is a corollary of Proposition 6:

Corollary 3. *There exists a skill-type $\tilde{\theta}$ s.t. $\int_\Pi \min\{\pi, z^*\}dH(\pi|\tilde{\theta}) = I$ and*

$$\tilde{\phi}_{z^*}(\theta) > \tilde{\phi}_e(\theta) \ \ \forall \theta < \tilde{\theta} \quad \textit{and} \quad \tilde{\phi}_{z^*}(\theta) < \tilde{\phi}_e(\theta) \ \ \forall \theta > \tilde{\theta}.$$

Proof. First, $\int_\Pi \min\{\pi, z^*\}dH(\pi|\theta)$ is increasing in θ by the monotone likelihood ratio property. Second, $\tilde{\theta}$ exists by the aggregate zero profit constraint (19). Third, by (20) and (22), $\tilde{\phi}_e(\theta) \gtreqless \tilde{\phi}_{z^*}$ if and only if $\int_\Pi \min\{\pi, z^*\}dH(\pi|\theta) \gtreqless I$. □

Since the credit market equilibrium is a pooling equilibrium, it involves cross-subsidization across entrepreneurs of different quality θ. In particular, by the monotone likelihood ratio property, banks make higher profits with higher ability entrepreneurs, and thus by the zero profit condition (19), there exists some critical skill level $\tilde{\theta}$ such that banks make profits with all higher quality entrepreneurs and negative profits with all the others. But this cross-subsidization implies that, compared to the efficient occupational choice defined in (22), low skilled agents have too strong incentives to set up a firm, and too many high skill agents stay in the workforce. In other words, the credit market equilib-

[31] This definition holds for any given wage, as does all of the following analysis.

rium generates occupational misallocation such that there is excessive entry of low ability types into entrepreneurship, but insufficient entry of high-skilled types. This can be seen most easily by substituting the equilibrium zero profit condition (19) into equation (22), solving the former for I:

$$\tilde{\phi}_e(\theta) = \int_{\Pi} \pi dH(\pi|\theta) - \frac{\int_{\Theta} G(\tilde{\phi}_{z^*}(\theta)) \int_{\Pi} \min\{\pi, z^*\} \, dH(\pi|\theta) dF(\theta)}{\int_{\Theta} G(\tilde{\phi}_{z^*}(\theta)) dF(\theta)} - v_W$$

and comparing it with the equilibrium critical values for occupational choice in (20):

$$\tilde{\phi}_{z^*}(\theta) = \int_{\Pi} \pi dH(\pi|\theta) - \int_{\Pi} \min\{\pi, z^*\} dH(\pi|\theta) - v_W.$$

Notably, this comparison clearly identifies the importance of cross-subsidization as the source of the inefficiency: If Θ is singleton, for instance, then $\tilde{\phi}_{z^*}(\theta) = \tilde{\phi}_e(\theta)$.

A special case of this misallocation obtains when ϕ is the same for all agents, so that there only remains one-dimensional heterogeneity in ability. Then the equilibrium involves an excessive entry into entrepreneurship, with too many low-skill types receiving funding in the credit market. This observation has been made first by De Meza and Webb (1987) in a model where agents choose between a safe investment and a risky project (entrepreneurship) with binary output. This extreme case is quite in contrast to the seminal analysis by Stiglitz and Weiss (1976), who emphasized credit rationing and thus insufficient entry into entrepreneurship in a model where entrepreneurs differ in the riskiness of their projects rather than expected returns. The present model demonstrates that, with two-dimensional heterogeneity, the occupational inefficiency can take both forms simultaneously, as there are too many and too few entrepreneurs of different skill types. It makes clear that, most generally, if different occupations are affected by different degrees of cross-subsidization, this makes the equilibrium occupational choice decisions inefficient.

In the following, I show that there is a simple entrepreneurial tax policy that may eliminate this occupational misallocation. Suppose the government introduces a (possibly non-linear) entrepreneurial profit tax $T(\pi)$, so that an entrepreneur's after-tax profits are given by $\hat{\pi} \equiv \pi - T(\pi)$. Banks and entrepreneurs, taking the tax schedule $T(\pi)$ as given, then write contracts contingent on these after-tax profits $\hat{\pi}$, and I can define the resulting credit market equilibrium for any given tax policy just as in Definition 2, replacing π by $\hat{\pi}$. Moreover, I keep assuming that contracts $R_\theta(\hat{\pi})$ must satisfy the limited liability constraint $0 \leq R_\theta(\hat{\pi}) \leq \hat{\pi}$ and the monotonicity constraint that $R_\theta(\hat{\pi})$ is non-decreasing in $\hat{\pi}$.[32] Under these conditions, it is known from Proposition 6 that the credit market equilibrium is a pooling equilibrium with only a debt contract being offered if the after-tax profits $\hat{\pi}$ satisfy the monotone likelihood ratio property with respect to θ, so that $\hat{H}(\hat{\pi}|\theta) \succeq_{MLRP} \hat{H}(\hat{\pi}|\theta')$ for $\theta > \theta'$, where $\hat{H}(\hat{\pi}|\theta)$ is the cdf of after-tax profits for type θ. The following lemma provides a condition on the tax schedule $T(\pi)$ for this to hold.

Lemma 1. *Suppose $H(\pi|\theta) \succeq_{MLRP} H(\pi|\theta')$ for $\theta > \theta'$, $\theta, \theta' \in \Theta$, and $T(\pi)$ is such that $\hat{\pi} = \pi - T(\pi)$ is increasing. Then $\hat{H}(\hat{\pi}|\theta) \succeq_{MLRP} \hat{H}(\hat{\pi}|\theta')$.*

Proof. Since $\hat{\pi} \equiv \Gamma(\pi) \equiv \pi - T(\pi)$ and $\Gamma(\pi)$ is increasing, the following relation between $\hat{H}(\hat{\pi}|\theta)$ and $H(\pi|\theta)$ is true:

$$\hat{H}(\Gamma(\pi)|\theta) = H(\pi|\theta) \;\; \forall \pi \in \Pi, \theta \in \Theta$$

and equivalently

$$\hat{H}(\hat{\pi}|\theta) = H(\Gamma^{-1}(\hat{\pi})|\theta) \;\; \forall \hat{\pi} \in \hat{\Pi}, \theta \in \Theta.$$

Differentiating with respect to $\hat{\pi}$, I therefore obtain

$$\hat{h}(\hat{\pi}|\theta) = \frac{h(\Gamma^{-1}(\hat{\pi})|\theta)}{\Gamma'(\Gamma^{-1}(\hat{\pi}))} \;\; \forall \pi \in \hat{\Pi}, \theta \in \Theta. \tag{23}$$

By assumption, $H(\pi|\theta)$ satisfies MLRP, which means that $h(\pi|\theta)/h(\pi|\theta')$ is increasing in π for $\theta > \theta'$. Equation (23) yields

$$\frac{\hat{h}(\hat{\pi}|\theta)}{\hat{h}(\hat{\pi}|\theta')} = \frac{h(\Gamma^{-1}(\hat{\pi})|\theta)/\Gamma'(\Gamma^{-1}(\hat{\pi}))}{h(\Gamma^{-1}(\hat{\pi})|\theta')/\Gamma'(\Gamma^{-1}(\hat{\pi}))} = \frac{h(\Gamma^{-1}(\hat{\pi})|\theta)}{h(\Gamma^{-1}(\hat{\pi})|\theta')},$$

which is increasing in $\Gamma^{-1}(\hat{\pi})$ by the assumption that $H(\pi|\theta)$ satisfies MLRP. Then the result follows from the fact that $\Gamma^{-1}(\hat{\pi})$ is an increasing function since $\hat{\pi} = \Gamma(\pi) = \pi - T(\pi)$ is increasing. □

Lemma 1 considers entrepreneurial profit tax schedules that involve marginal tax rates uniformly less than one, so that after-tax profits are increasing in before-tax profits. This is a weak restriction on tax policy that I assume to be satisfied in the following. The lemma shows that, under this condition, the fact that higher θ-types have better before-tax profit distributions in the sense of the monotone likelihood ratio property translates into the same ordering of after-tax profit distributions. This is intuitive since such a profit tax preserves the ranking of before-tax profit levels and applies to all θ-types equally. Combined with Proposition 6, Lemma 1 then implies that, whenever the government imposes a tax on entrepreneurial profits $T(\pi)$ that involves marginal tax rates less than one, the resulting credit market equilibrium with this tax will be a single debt contract

[32] The idea behind this is that the government has no superior ability to extract tax payments from a firm in case of bankruptcy compared to banks, so that is must always hold that $\pi - T(\pi) - R \geq 0$, where R is the repayment to the bank. In addition, when $T(\pi)$ is negative, it is assumed that banks can capture this transfer from the government in case of bankruptcy, so that $R \leq \pi - T(\pi)$. In other words, tax payments are fully pledgeable.

$R_{z_T^*}(\hat{\pi}) = \min\{\hat{\pi}, z_T^*\}$, where z_T^* is such that banks make zero profits in aggregate:

$$\int_\Theta G(\tilde{\phi}_{z_T^*,T}(\theta)) \left[\int_\Pi \min\{\pi - T(\pi), z_T^*\} dH(\pi|\theta) - I\right] dF(\theta) = 0. \tag{24}$$

Here,

$$\tilde{\phi}_{z_T^*,T}(\theta) = \int_\Pi (\pi - T(\pi) - \min\{\pi - T(\pi), z_T^*\})\, dH(\pi|\theta) - v_W \tag{25}$$

denotes the critical cost value for entry into entrepreneurship at θ when the tax policy $T(\pi)$ is in place.

Now suppose the government sets the profit tax schedule $T(\pi)$ such that, for all $\theta \in \Theta$,

$$\int_\Pi T(\pi) dH(\pi|\theta) = -\left(\int_\Pi \min\{\pi - T(\pi), z_T^*\} dH(\pi|\theta) - I\right). \tag{26}$$

Obviously, substituting equation (26) into (25) yields $\tilde{\phi}_{z_T^*,T}(\theta) = \tilde{\phi}_e(\theta)$ for all θ, so that the policy is exactly counteracting the cross-subsidization in the credit market, providing the efficient incentives for entry into entrepreneurship to all agents.[33] Note that equation (26) is a fixed point condition, since for any given profit tax schedule $T(\pi)$, we can compute the equilibrium debt contract z_T^* from solving equations (24) and (25), and given z_T^*, the tax policy must satisfy equation (26). Total revenue from the profit tax is then given by

$$\begin{aligned} & \int_\Theta G(\tilde{\phi}_{z_T^*,T}(\theta)) \int_\Pi T(\pi) dH(\pi|\theta) dF(\theta) \\ = & -\int_\Theta G(\tilde{\phi}_{z_T^*,T}(\theta)) \left(\int_\Pi \min\{\pi - T(\pi), z_T^*\} dH(\pi|\theta) - I\right) dF(\theta) = 0, \end{aligned}$$

where the first equality follows from equation (26) and the second from the zero profit condition (24). Hence, the government budget constraint is automatically satisfied with equality and the tax policy $T(\pi)$ is feasible. The following proposition summarizes these insights:

Proposition 7. *Suppose that an entrepreneurial tax policy $T(\pi)$ is introduced that is such that $\pi - T(\pi)$ is increasing and equation (26) is satisfied. Then*
(i) the resulting credit market equilibrium is such that $\tilde{\phi}_{z_T^*,T}(\theta) = \tilde{\phi}_e(\theta)$ for all $\theta \in \Theta$, where $\tilde{\phi}_{z_T^*,T}(\theta)$ and $\tilde{\phi}_e(\theta)$ are given by (25) and (22), respectively, and
(ii) the government budget is balanced.

By (26), the efficient entrepreneurial tax policy is regressive in the sense that higher ability entrepreneurs face a lower expected tax payment. In fact, for all $\theta > \tilde{\theta}_T$ with

[33] From a bank's perspective, of course, there is still cross-subsidization as net profits vary with θ.

$\tilde{\theta}_T$ such that $\int_{\Pi} \min\{\pi - T(\pi), z_T^*\}\, dH(\pi|\tilde{\theta}_T) = I$, the expected tax payment is negative. This is because the entrepreneurial tax has to counteract the equilibrium cross-subsidization in the credit market, which is decreasing in θ as argued above. By the monotone likelihood ratio property of $H(\pi|\theta)$, this pushes towards a tax schedule $T(\pi)$ that is itself decreasing in π and in that sense regressive as well. This makes the assumption in Lemma 1 that $\pi - T(\pi)$ is increasing even less restrictive.

To see how the system of equation (24) to (26) can be solved for the efficient tax policy, it is useful to consider the following simple example.

Example 1. *Suppose there are two ability types θ^k, $k = g, b$, where the good type g has the better profit distribution in the sense of MLRP than the bad type b. Also, suppose there are three possible profit levels $\pi_l < \pi_m < \pi_h$ with probabilities h_l^k, h_m^k and h_h^k for each type $k = g, b$. Let me fix some debt repayment level z_T^* and assume that only the realization of the lowest profit level π_l leads to bankruptcy. Then equation (26) becomes*

$$h_h^k T(\pi_h) + h_m^k T(\pi_m) = -\left[(h_h^k + h_m^k)(z_T^* - I) + h_l^k(\pi_l - I)\right]$$

for $k = g, b$. This is a system of two linear equations that can be solved for the two unknowns $T(\pi_h)$, $T(\pi_m)$. Next, the critical cost values for occupational choice in equation (25)

$$\tilde{\phi}_{z_T^*,T}(\theta^k) = h_h^k(\pi_h - T(\pi_h) - z_T^*) + h_m^k(\pi_m - T(\pi_m) - z_T^*) - v_W,$$

$k = g, b$, are entirely pinned down by z_T^*, $T(\pi_h)$ and $T(\pi_m)$. Hence, the zero profit condition (24) holds with equality when finding $T(\pi_l)$ such that

$$\sum_{k=g,b} G(\tilde{\phi}_{z_T^*,T}(\theta^k))[(h_h^k + h_m^k)z_T^* + h_l^k(\pi_l - T(\pi_l)) - I] f^k = 0.$$

Finally, z_T^* is adjusted to make $\pi_h - T(\pi_h) > \pi_m - T(\pi_m) > z_T^* > \pi_l - T(\pi_l)$ hold.

This example shows that it is easiest to find an entrepreneurial tax policy that solves equation (26) if the space of possible profit realizations is relatively rich compared to the type space Θ, for example when Π is an interval but there are only two groups of entrepreneurial abilities. Then the profit tax schedule $T(\pi)$ provides a high degree of flexibility, while only a small number of restrictions need to be satisfied. In the other extreme case, where there is a continuum of types but only a small number of possible profit realizations, it may not be possible to find an entrepreneurial tax policy that satisfies (26) and thus restores efficient occupational choice for all ability types $\theta \in \Theta$. However, in this case, one may think of grouping entrepreneurs in a finite set of ability levels, and providing the correct incentives for entry into entrepreneurship for those.

It is worth emphasizing that the entrepreneurial tax policy in Proposition 7 is quite different from the general subsidization of entrepreneurship that one may think of at first glance in view of credit market frictions. Even if there is only one-dimensional heterogeneity in entrepreneurial abilities, the resulting excessive entry into entrepreneurship in this case would require a lump sum tax on entrepreneurial profits, rather than a subsidy. Since the inefficiency with two-dimensional heterogeneity is more complicated, however, such a uniform tax turns out not to be optimal in general. The policy also differs from the tax on bank profits that De Meza and Webb (1987) propose in order to deal with the excessive entry into entrepreneurship that they find in their model with one-dimensional heterogeneity. As can be seen from the zero profit condition (24) and most clearly Example 1, a tax on bank profits and a lump-sum tax of entrepreneurial profits have equivalent effects. Thus, a tax on bank profits is not able (nor necessary) to restore occupational efficiency in the present setting. Instead, Proposition 7 points out the importance of entrepreneurial profit taxation as a more flexible corrective instrument.

6 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. Moreover, I have pointed out that a differential tax treatment of profits can also be justified based on corrective arguments, mitigating occupational misallocation that results from credit market frictions. In addition, the quantitative importance of differential taxation has been explored in a calibrated model economy.

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

References

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

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

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

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

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

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

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

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

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

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

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 MEZA, D., AND D. WEBB (1987): "Too Much Investment: A Problem of Asymmetric Information," *Quarterly Journal of Economics*, 102, 281–292.

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

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

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

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

EVANS, D., AND L. LEIGHTON (1989): "Some Empirical Aspects of Entrepreneurship," *American Economic Review*, 79, 519–535.

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

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

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

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

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

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

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

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

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

HURST, E., AND A. LUSARDI (2004): "Liquidity Constraints, Wealth Accumulation, and Entrepreneurship," *Journal of Political Economy*, 112, 319–347.

INNES, R. (1990): "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory*, 52, 45–67.

——— (1991): "Investment and Government Intervention in Credit Markets When There Is Asymmetric Information," *Journal of Public Economics*, 46, 347–381.

——— (1992): "Adverse Selection, Investment, and Profit Taxation," *European Economic Review*, 36, 1427–1452.

——— (1993): "Financial Contracting Under Risk Neutrality, Limited Liability and Ex Ante Asymmetric Information," *Economica*, 237, 27–40.

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.

MIYAZAKI, H. (1977): "The Rat Race and Internal Labor Markets," *Bell Journal of Economics*, 8, 394–418.

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

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

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

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

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

——— (2003): "Asymmetric Information, Occupational Choice and Government Policy," *Economic Journal*, 113, 861–882.

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

ROTHSCHILD, M., AND J. STIGLITZ (1976): "Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Incomplete Information," *Quarterly Journal of Economics*, 90, 629–649.

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

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

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

STIGLITZ, J., AND A. WEISS (1976): "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 71, 393–410.

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

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

WILSON, C. (1977): "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, 12, 167–207.

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

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

A Proofs for Sections 3 and 4

A.1 Proof of Proposition 2

(i) Analogously to the proof of Proposition 1, if $\tilde{w} \leq w$, then $\tilde{\phi}(\theta) = 0$ for all $\theta \in \Theta$ (with the only additional argument that, since both occupations face the same tax schedule on their profits (resp. income), there is also no tax advantage from entering entrepreneurship). This, together with the fact that (11) and (12) must hold as equalities at an optimum, implies $l(\theta) = E(\theta) = v_E(\theta) = v_W(\theta) = 0$ for all $\theta \in \Theta$. Clearly, the no tax equilibrium characterized in Proposition 1 is Pareto-superior, demonstrating that $\tilde{w} \leq w$ cannot be part of a Pareto-optimum.

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}\mathcal{L} \;=\; & \int_\Theta \Bigg[\tilde{G}_\theta(\tilde{\phi}(\theta))v_E(\theta) - \int_{\underline{\phi}}^{\tilde{\phi}(\theta)} \phi d\tilde{G}_\theta(\phi) + (1-\tilde{G}_\theta(\tilde{\phi}(\theta)))v_W(\theta)\Bigg] d\tilde{F}(\theta) \\ & - \int_\Theta \left[\mu_E'(\theta)v_E(\theta)+\mu_E(\theta)\psi'\left(\frac{E(\theta)}{\theta}\right)\frac{E(\theta)}{\theta^2}\right]d\theta - \int_\Theta \left[\mu_W'(\theta)v_W(\theta)+\mu_W(\theta)\psi'\left(\frac{l(\theta)}{\theta}\right)\frac{l(\theta)}{\theta^2}\right]d\theta \\ & +\lambda_{LM}\left[\int_\Theta (1-G_\theta(\tilde{\phi}(\theta)))l(\theta)dF(\theta) - \int_\Theta G_\theta(\tilde{\phi}(\theta))L(\theta)dF(\theta)\right] \\ & +\lambda_{RC}\Bigg[\int_\Theta G_\theta(\tilde{\phi}(\theta))\left[Y(L(\theta),E(\theta))-v_E(\theta)-\psi\left(\frac{E(\theta)}{\theta}\right)\right]-(1-G(\tilde{\phi}(\theta)))\left[v_W(\theta)+\psi\left(\frac{l(\theta)}{\theta}\right)\right]dF(\theta)\Bigg] \\ & + \int_{\underline{\theta}}^{(w/\tilde{w})\overline{\theta}} \xi_v(\theta)\left[v_E(\theta) - v_W\left(\frac{\tilde{w}}{w}\theta\right)\right]d\theta + \int_{\underline{\theta}}^{(w/\tilde{w})\overline{\theta}} \xi_E(\theta)\left[E(\theta) - \frac{w}{\tilde{w}}l\left(\frac{\tilde{w}}{w}\theta\right)\right]d\theta \\ & + \int_\Theta \kappa_L(\theta)\left[w - Y_L(L(\theta),E(\theta))\right]d\theta + \int_\Theta \kappa_E(\theta)\left[\tilde{w} - Y_E(L(\theta),E(\theta))\right]d\theta. \end{aligned} \tag{27}$$

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

$$G_\theta(\tilde{\phi}(\theta))f(\theta)\left[Y_L(L(\theta),E(\theta)) - \lambda_{LM}\right] - \left[\kappa_L(\theta)Y_{LL}(\theta) + \kappa_E(\theta)Y_{EL}(\theta)\right] = 0 \;\; \forall\theta\in\Theta. \tag{28}$$

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

$$G_{\overline{\theta}}(\tilde{\phi}(\overline{\theta}))f(\overline{\theta})\left[\tilde{w} - \frac{1}{\overline{\theta}}\psi'\left(\frac{E(\overline{\theta})}{\overline{\theta}}\right)\right] - \left[\kappa_L(\overline{\theta})Y_{LE}(\overline{\theta}) + \kappa_E(\overline{\theta})Y_{EE}(\overline{\theta})\right] = 0 \tag{29}$$

and

$$\lambda_{LM} - \frac{1}{\underline{\theta}}\psi'\left(\frac{l(\underline{\theta})}{\underline{\theta}}\right) = 0. \tag{30}$$

If $Y(L,E)$ is linear, then $Y_{LL} = Y_{LE} = Y_{EE} = 0$ and thus (28) and (MP) imply $\lambda_{LM} = Y_L(\theta) = w$ for all θ. Therefore, by (29) and (30),

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

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.

(iii) There are 3 cases to be considered:

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

In this case, (28) 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,

$$Y_{LL}(\theta) = -xY_{EL}(\theta) \quad \text{and} \quad Y_{EL}(\theta) = -xY_{EE}(\theta) \ \ \forall\theta \in \Theta, \tag{31}$$

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 (29) and (30) imply $T'(\tilde{w}E(\overline{\theta})) = T'(wl(\underline{\theta})) = 0$.
Case 2: $\lambda_{LM} < w$.
Now (28) and (MP) yield

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

and hence by (31)

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

Then (29) and (30) yield $T'(\tilde{w}E(\overline{\theta})) < 0$ and $T'(wl(\underline{\theta})) > 0$.
Case 3: $\lambda_{LM} > w$.
This case is completely analogous to case 2 with all signs reversed.

A.2 Proof of Proposition 3

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

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

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

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

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

which implies

$$\int_\Theta \kappa_E(\theta)d\theta = -\frac{w}{\tilde{w}}\int_\Theta \kappa_L(\theta)d\theta \tag{32}$$

and hence $\int_\Theta \kappa_E(\theta)d\theta = 0$. To obtain a contradiction, suppose $\lambda_{LM} < \tau w$. Then (28) and (MP') imply

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

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

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

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

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

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

and (ND) becomes

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

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

$$\kappa_E(\theta) = -\frac{1}{w}\xi_v(\theta)v_W'\left(\frac{\tilde{w}}{w}\theta\right) - \xi_E(\theta)\left[\frac{1}{\tilde{w}}l'\left(\frac{\tilde{w}}{w}\theta\right) - \frac{w}{\tilde{w}^2}l\left(\frac{\tilde{w}}{w}\theta\right)\right] \;\; \forall \theta \in \Theta. \tag{33}$$

Note that it must hold that $\tilde{w}(\overline{\theta}) > w$ (since otherwise $\tilde{\phi}(\overline{\theta}) = 0$), and therefore (ND') does not bind at $\overline{\theta}$, which yields $\kappa_E(\overline{\theta}) = 0$. Then (28) implies (together with $\kappa_L(\overline{\theta}) = 0$) that $Y_L(\overline{\theta}) = \lambda_{LM}$. However, whenever there exists some $\theta < \overline{\theta}$ such that (ND') binds, then (33) implies $\kappa_E(\theta) \neq 0$ and thus, by (28), $Y_L(\theta) \neq \lambda_{LM}$.

A.3 Proof of Proposition 4

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

$$\begin{aligned}
\mathcal{L} \;\; = \;\; & \int_\Theta \left[\tilde{G}_\theta(\tilde{\phi}(\theta))v_E(\theta) - \int_{\underline{\phi}}^{\tilde{\phi}(\theta)} \phi d\tilde{G}_\theta(\phi) + (1-\tilde{G}_\theta(\tilde{\phi}(\theta)))v_W(\theta)\right]d\tilde{F}(\theta) \\
& - \int_\Theta \left[\mu_E'(\theta)v_E(\theta) + \mu_E(\theta)\psi'\left(\frac{E(\theta)}{\theta}\right)\frac{E(\theta)}{\theta^2}\right]d\theta - \int_\Theta \left[\mu_W'(\theta)v_W(\theta) + \mu_W(\theta)\psi'\left(\frac{l(\theta)}{\theta}\right)\frac{l(\theta)}{\theta^2}\right]d\theta \\
& + \lambda_{LM}\left[\int_\Theta (1 - G_\theta(\tilde{\phi}(\theta)))l(\theta)dF(\theta) - \int_\Theta G_\theta(\tilde{\phi}(\theta))L(\theta)dF(\theta)\right] \\
& + \lambda_{RC}\left[\int_\Theta G_\theta(\tilde{\phi}(\theta))\left[Y(L(\theta), E(\theta)) - v_E(\theta) - \psi\left(\frac{E(\theta)}{\theta}\right)\right] - (1 - G(\tilde{\phi}(\theta)))\left[v_W(\theta) + \psi\left(\frac{l(\theta)}{\theta}\right)\right]dF(\theta)\right].
\end{aligned} \tag{34}$$

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

$$Y_L(L(\theta), E(\theta)) = \lambda_{LM}/\lambda_{RC} \;\; \forall \theta \in \Theta \tag{35}$$

and hence the result.

(ii) Note that (35) 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

$$\mu_E'(\theta) = \tilde{G}(\tilde{\phi}(\theta))\tilde{f}(\theta) - \lambda_{RC} G(\tilde{\phi}(\theta)) f(\theta) + g(\tilde{\phi}(\theta)) f(\theta) \lambda_{RC} \left[Y(\theta) - c_E(\theta) + c_W(\theta) - w\left(L(\theta) + l(\theta)\right)\right], \tag{36}$$

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

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

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

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

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

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

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

$$\mu_E(\theta) = \int_{\underline{\theta}}^{\theta} \left[\tilde{G}(\tilde{\phi}(\hat{\theta}))\tilde{f}(\hat{\theta}) - G(\tilde{\phi}(\hat{\theta})) f(\hat{\theta}) + g(\tilde{\phi}(\hat{\theta})) f(\hat{\theta}) \Delta T(\hat{\theta}) \right] d\hat{\theta} \tag{37}$$

and

$$\mu_W(\theta) = \int_{\underline{\theta}}^{\theta} \left[(1 - \tilde{G}(\tilde{\phi}(\hat{\theta})))\tilde{f}(\hat{\theta}) - (1 - G(\tilde{\phi}(\hat{\theta}))) f(\hat{\theta}) - g(\tilde{\phi}(\hat{\theta})) f(\hat{\theta}) \Delta T(\hat{\theta}) \right] d\hat{\theta}. \tag{38}$$

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

$$\frac{\tilde{w} - \psi'(E(\theta)/\theta)/\theta}{\psi'(E(\theta)/\theta)/\theta} = \frac{\mu_E(\theta)}{\theta f(\theta) G(\tilde{\phi}(\theta))}\left(1 + \frac{\psi''(E(\theta)/\theta)E(\theta)/\theta^2}{\psi'(E(\theta)/\theta)/\theta}\right). \tag{39}$$

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 (37), this allows me to rewrite (39) as

$$\frac{T_\pi'(\pi(\theta))}{1-T_\pi'(\pi(\theta))}=\frac{1+1/\varepsilon_\pi(\theta)}{\theta f(\theta)G_\theta(\tilde{\phi}(\theta))}\int_{\underline{\theta}}^{\theta}[\tilde{G}_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\tilde{f}(\hat{\theta})-G_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))f(\hat{\theta})+g_{\hat{\theta}}(\tilde{\phi}(\hat{\theta}))\Delta T(\hat{\theta})f(\hat{\theta})]d\hat{\theta}, \tag{40}$$

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 (38).

(iii) $T_\pi'(\pi(\underline{\theta})) = T_\pi'(\pi(\overline{\theta})) = 0$ immediately follows from (39) 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

$$\Delta T(\theta_a) \leq \frac{G(\tilde{\phi}(\theta_a)) - \tilde{G}(\tilde{\phi}(\theta_a))}{g(\tilde{\phi}(\theta_a))} \quad \text{and} \quad \Delta T(\theta_b) \geq \frac{G(\tilde{\phi}(\theta_b)) - \tilde{G}(\tilde{\phi}(\theta_b))}{g(\tilde{\phi}(\theta_b))}. \tag{41}$$

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 ϕ. With this, equation (41) 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 (11) 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 (12) 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.

B Computational Procedure for Section 4.2

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

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

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

   Also, $L(\theta)$ is obtained from $x = E(\theta)/L(\theta)$ and $E(\theta)$.

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

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

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

   using the fact that

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

   and

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

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

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

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

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

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

C Proof of Proposition 6

By construction, the proposed equilibrium contract $R_{z^*}(\pi)$ satisfies conditions (i) and (ii) of Definition 2, and thus only requirement (iii) remains to be checked. I will do so by proving a series of lemmas, starting with the following result due to Innes (1993).

Lemma 2. *Consider an arbitrary non-debt contract $R(\pi)$ that satisfies the limited liability and monotonicity constraints, and let $R_{z_\theta}(\pi) \equiv \min\{\pi, z_\theta\}$ denote the debt contract such that*

$$\int_\Pi R(\pi)dH(\pi|\theta) = \int_\Pi R_{z_\theta}(\pi)dH(\pi|\theta)$$

for some $\theta \in \Theta$. Then

$$\int_\Pi R(\pi)dH(\pi|\theta') \leq \int_\Pi R_{z_\theta}(\pi)dH(\pi|\theta') \;\; \forall\theta' \leq \theta.$$

In words, whenever banks offer a debt contract $R_{z_\theta}(\pi)$ that involves the same expected repayment for entrepreneurs of ability θ as the non-debt contract $R(\pi)$, then the expected repayment from the debt contract $R_{z_\theta}(\pi)$ is at least as high as from the non-debt contract $R(\pi)$ for all entrepreneurs of a lower skill $\theta' \leq \theta$. This result immediately follows from the fact that the entrepreneurs' profit distributions are ranked by MLRP and that, among the contracts that satisfy the limited liability and monotonicity constraints, debt contracts put the maximal repayment in low profit states. Note that Lemma 2 also immediately implies

$$\int_\Pi [\pi - R(\pi)]\, dH(\pi|\theta') \geq \int_\Pi [\pi - R_{z_\theta}(\pi)]\, dH(\pi|\theta') \;\; \forall \theta' \leq \theta,$$

i.e. all entrepreneurs of quality less than θ prefer the non-debt contract $R(\pi)$ to the debt contract $R_{z_\theta}(\pi)$. Clearly, this is independent of the cost type ϕ.

Suppose that, in the presence of the equilibrium contract $R_{z^*}(\pi)$, a bank offers an arbitrary, incentive compatible set of deviation contracts $\{R^d_\theta(\pi)\}$. Let me denote the resulting critical cost values for occupational choice by $\tilde{\phi}_d(\theta)$, i.e. for all $\theta \in \Theta$,

$$\tilde{\phi}_d(\theta) \equiv \max\left\{\int_\Pi [\pi - R_{z^*}(\pi)]\, dH(\pi|\theta), \int_\Pi \left[\pi - R^d_\theta(\pi)\right] dH(\pi|\theta)\right\} - v_W. \tag{42}$$

Next, the following auxiliary result is useful.

Lemma 3. *For all $\theta \in \Theta$, let $\Delta\tilde{\phi}(\theta) \equiv \tilde{\phi}_d(\theta) - \tilde{\phi}_{z^*}(\theta)$ denote the change in critical cost values for occupational choice due to the deviation. Then $\Delta\tilde{\phi}(\theta)$ is decreasing in θ.*

Proof. Showing that $\Delta\tilde{\phi}(\theta) \equiv \tilde{\phi}_d(\theta) - \tilde{\phi}_{z^*}(\theta)$ is decreasing in θ is, by (20) and (42), equivalent to showing that

$$\int_\Pi R^d_\theta(\pi) dH(\pi|\theta) - \int_\Pi R_{z^*}(\pi) dH(\pi|\theta)$$

is increasing in θ. To see this, note that, by Lemma 2, if $\int_\Pi R^d_\theta(\pi)dH(\pi|\theta) = \int_\Pi R_{z^*}(\pi)dH(\pi|\theta)$ for some θ, then $\int_\Pi R^d_\theta(\pi)dH(\pi|\theta') \leq \int_\Pi R_{z^*}(\pi)dH(\pi|\theta')$ for all $\theta' \leq \theta$, which implies that

$$\int_\Pi R^d_\theta(\pi)dH(\pi|\theta) - \int_\Pi R_{z^*}(\pi)dH(\pi|\theta) \geq \int_\Pi R^d_\theta(\pi)dH(\pi|\theta') - \int_\Pi R_{z^*}(\pi)dH(\pi|\theta') \tag{43}$$

whenever $\theta \geq \theta'$. Moreover, by incentive compatibility of $\{R^d_\theta(\pi)\}$,

$$\int_\Pi \left[\pi - R^d_{\theta'}(\pi)\right] dH(\pi|\theta') \geq \int_\Pi \left[\pi - R^d_\theta(\pi)\right] dH(\pi|\theta')$$

and hence $\int_\Pi R^d_{\theta'}(\pi)dH(\pi|\theta') \leq \int_\Pi R^d_\theta(\pi)dH(\pi|\theta')$, which, when combined with (43), completes the argument. □

This allows me to prove the following lemma:

Lemma 4. *Let $\Delta G(\theta) \equiv G(\tilde{\phi}_d(\theta)) - G(\tilde{\phi}_{z^*}(\theta))$ for all $\theta \in \Theta$. Under Assumption 1, $\Delta G(\theta)$ is decreasing in θ.*

Proof. First, observe that $\tilde{\phi}_{z^*}(\theta)$ is increasing in θ by MLRP. Moreover, $G(\phi)$ is concave by Assumption 1. Therefore, the result from Lemma 3 that $\Delta\tilde{\phi}(\theta) \equiv \tilde{\phi}_d(\theta) - \tilde{\phi}_{z^*}(\theta)$ is decreasing in θ implies that

$$\Delta G(\theta) \equiv G(\tilde{\phi}_d(\theta)) - G(\tilde{\phi}_{z^*}(\theta))$$

is also decreasing in θ, proving the lemma. □

The deviating bank's expected profits from offering $\{R^d_\theta(\pi)\}$ are given by

$$\begin{aligned}\Pi^d &= \int_\Theta \mathbb{1}_{\{\tilde{\phi}_d(\theta)>\tilde{\phi}_{z^*}(\theta)\}}(\theta)G(\tilde{\phi}_d(\theta))\int_\Pi (R^d_\theta(\pi)-I)dH(\pi|\theta)dF(\theta)\\ &< \int_\Theta \mathbb{1}_{\{\tilde{\phi}_d(\theta)>\tilde{\phi}_{z^*}(\theta)\}}(\theta)G(\tilde{\phi}_d(\theta))\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta)\end{aligned} \tag{44}$$

since $\int_\Pi R^d_\theta(\pi)dH(\pi|\theta) < \int_\Pi R_{z^*}(\pi)dH(\pi|\theta)$ whenever $\tilde{\phi}_d(\theta) > \tilde{\phi}_{z^*}(\theta)$ by (42). Aggregate profits in the proposed equilibrium are

$$\Pi^* = \int_\Theta G(\tilde{\phi}_{z^*}(\theta))\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta) = 0 \tag{45}$$

by (19), and therefore, subtracting (45) from (44) yields

$$\begin{aligned}\Pi^d &< \int_\Theta \left(\mathbb{1}_{\{\tilde{\phi}_d(\theta)>\tilde{\phi}_{z^*}(\theta)\}}(\theta)G(\tilde{\phi}_d(\theta)) - G_{z^*}(\tilde{\phi}(\theta))\right)\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta)\\ &= \int_\Theta \left(\Delta G(\theta) - \mathbb{1}_{\{\tilde{\phi}_d(\theta)=\tilde{\phi}_{z^*}(\theta)\}}(\theta)G(\tilde{\phi}_{z^*}(\theta))\right)\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta)\end{aligned} \tag{46}$$

The following two lemmas establish that the RHS of (46) is non-positive.

Lemma 5. *In equation (46),*

$$\int_\Theta \Delta G(\theta)\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta) \le 0. \tag{47}$$

Proof. Find $\tilde{\theta}$ such that $\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\tilde{\theta}) = 0$, which exists and is unique by (19) and MLRP. Also, find the constant δ such that $\delta G(\tilde{\phi}_{z^*}(\tilde{\theta})) = \Delta G(\tilde{\theta})$. Then since $\Delta G(\theta)$ is decreasing and $\delta G(\tilde{\phi}_{z^*}(\theta))$ is increasing by Lemma 4, $\delta G(\tilde{\phi}_{z^*}(\theta)) \le \Delta G(\theta)$ for all $\theta \le \tilde{\theta}$, and $\delta G(\tilde{\phi}_{z^*}(\theta)) \ge \Delta G(\theta)$ otherwise. Thus,

$$\begin{aligned}&\int_{\underline{\theta}}^{\tilde{\theta}} \Delta G(\theta)\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta) + \int_{\tilde{\theta}}^{\overline{\theta}} \Delta G(\theta)\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta)\\ \le\ &\int_{\underline{\theta}}^{\tilde{\theta}} \delta G(\tilde{\phi}_{z^*}(\theta))\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta) + \int_{\tilde{\theta}}^{\overline{\theta}} \delta G(\tilde{\phi}_{z^*}(\theta))\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta)\\ =\ &0,\end{aligned} \tag{48}$$

where the inequality comes from $\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta) < 0$ for $\theta \le \tilde{\theta}$ and $\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta) \ge 0$ otherwise, and the equality from (19). □

Lemma 6. *In equation (46),*

$$\int_\Theta \mathbb{1}_{\{\tilde{\phi}_d(\theta)=\tilde{\phi}_{z^*}(\theta)\}}(\theta)G(\tilde{\phi}_{z^*}(\theta))\int_\Pi (R_{z^*}(\pi)-I)dH(\pi|\theta)dF(\theta) \ge 0. \tag{49}$$

Proof. There are 3 cases to be considered. If $\tilde{\phi}_d(\theta) = \tilde{\phi}_{z^*}(\theta)$ for all $\theta \in \Theta$, then (49) holds with equality due to (19). If there does not exist a $\theta \in \Theta$ such that $\tilde{\phi}_d(\theta) = \tilde{\phi}_{z^*}(\theta)$, then (49) also holds as an equality trivially. Finally, if $\tilde{\phi}_d(\theta) = \tilde{\phi}_{z^*}(\theta)$ holds for some but not all $\theta \in \Theta$, there must exist some threshold value $\hat{\theta} \in (\underline{\theta}, \overline{\theta})$ such that $\tilde{\phi}_d(\theta) > \tilde{\phi}_{z^*}(\theta)$ for all $\theta < \hat{\theta}$ and $\tilde{\phi}_d(\theta) = \tilde{\phi}_{z^*}(\theta)$ otherwise. This follows from Lemma 3, which has shown that $\Delta\tilde{\phi}(\theta)$ is decreasing in θ and, by the definition in (42), $\tilde{\phi}_d(\theta) \geq \tilde{\phi}_{z^*}(\theta)$. With this, (49) becomes

$$\int_{\hat{\theta}}^{\overline{\theta}} G(\tilde{\phi}_{z^*}(\theta)) \int_{\Pi} (R_{z^*}(\pi) - I) dH(\pi|\theta) dF(\theta) > 0$$

since $\hat{\theta} > \underline{\theta}$ and $\int_{\Pi}(R_{z^*}(\pi) - I)dH(\pi|\theta)$ is increasing in θ by MLRP. $\square$

Lemmas 5 and 6 together with equation (46) show that $\Pi^d < 0$, and hence there does not exist a profitable deviation.