

# Taxing Atlas:

## Executive Compensation, Firm Size and Their Impact on Optimal Top Income Tax Rates\*

Laurence Ales<sup>1</sup>, Antonio Andrés Bellofatto<sup>2</sup>, and Jessie Jiaxu Wang<sup>3</sup>

<sup>1</sup>*Tepper School of Business, Carnegie Mellon University*

<sup>2</sup>*School of Economics, University of Queensland - Australia*

<sup>3</sup>*W. P. Carey School of Business, Arizona State University*

This Version: December 27, 2015. First Version: June 10, 2014

### Abstract

We study the optimal taxation of top labor incomes. Top income earners are modeled as managers who operate a span of control technology as in [Rosen \(1982\)](#). Managers are heterogeneous across talent, which is both effort-augmenting and total-factor-productivity improving. The latter gives rise to a positive scale-of-operations effect. A tax formula for optimal taxes is derived linking optimal marginal tax rates to preferences and technology parameters. We calibrate the model to the US by using firm level data. Our results suggests that optimal top taxes are roughly in line to the current statutory rates and, thus, are significantly lower than what previous studies ignoring the scale-of-operations effect have shown. These findings are robust when extending the quantitative analysis to a panel of non-US countries. In addition, our model provides normative grounds for using firm size taxation. (*JEL D31, H21, H24, M12, M52*)

---

\*We thank Adriano Rampini for his valuable discussion and suggestions. We also thank Emilio Bisetti, Chris Edmond, Miguel Ferreira, Brent Glover, Burton Hollifield, Greg Kaplan, Finn Kydland, Christos Makridis, Pricila Maziero, Tomasz Piskorski, Peter Rupert, Florian Scheuer, Ali Shourideh, Chris Sleet, Stephen Spear, Chris Telmer, Randy Wright, Sevin Yeltekin, and Ariel Zetlin-Jones, as well as conference and seminar participants at Carnegie Mellon Ph.D. Macro-Finance lunch, Econometric Society summer meetings in Minneapolis, Society for Economic Dynamics annual meeting in Toronto, 5<sup>th</sup> Tepper-LAEF conference in Pittsburgh, the University of Melbourne, and 2015 WAMS/LAEF conference in Sydney for feedback. E-mail: [ales@cmu.edu](mailto:ales@cmu.edu) (Ales); [a.bellofatto@uq.edu.au](mailto:a.bellofatto@uq.edu.au) (Bellofatto); [Jessie.Jiaxu.Wang@asu.edu](mailto:Jessie.Jiaxu.Wang@asu.edu) (Wang).

# 1 INTRODUCTION

Heightened concerns over recent trends in income inequality necessarily bring the taxation of high earners to the forefront of the policy agenda.<sup>1</sup> However, the vast literature in public finance is far from reaching consensus on what the top income tax rate should be: While the dominant view sets the optimal top tax rate above 70 per cent,<sup>2</sup> others have challenged the validity of such elevated taxes on various grounds.<sup>3</sup> This paper introduces a novel approach to modeling and quantifying the behavior of highly talented individuals in the economy, thus contributing to the aforementioned debate. We identify high earners with managers, and show how to exploit well established facts on firm size and managerial compensation to pin down the key forces shaping optimal income taxes. Contrary to the prevailing recommendations in the literature, we find that top income taxes at the optimum are actually aligned with the current statutory rates in the US.

Our model economy combines elements from optimal taxation environments, on the one hand, and from span of control studies, on the other. The economy is static and it is populated by workers and managers. We focus on the optimal tax treatment of the latter class of individuals, motivated by the fact that a significant fraction of top income earners fall within managerial occupations in the data.<sup>4</sup> Managers are heterogeneous across talent, which is privately observed. A benevolent government values redistribution across types, but informational frictions hinder full insurance. Building on the seminal contribution of [Mirrlees \(1971\)](#), optimal tax functions are only restricted by such informational asymmetries.

The production process is modeled by extending the span of control technology in [Rosen \(1982\)](#) to a framework with elastic labor supply. Specifically, managers operate separate productive units by hiring workers and exerting effort. The technology is such that managerial talent has a dual role. First, it affects own effort productivity, which captures the quality of supervision or monitoring. Second, managerial ability affects the overall productivity of the firm. This channel encapsulates the quality of indivisible managerial decisions (e.g. the location of production), and creates a so-called scale-of-operations effect.<sup>5</sup> Crucially, this effect magnifies the impact of skill variations on firm size and compensation differentials. In other words, for a given distribution of skills, a larger the scale-of-operations generates more right

---

<sup>1</sup>In the US, the share of income going to the top 1% increased from 9% in 1970 to 23.5% in 2007 ([Diamond and Saez \(2011\)](#)). The top 1% accounted for 59.8% of average growth in income compared to just 9% of average growth accounted for by the bottom 90% over this period ([Piketty and Saez \(2003\)](#)).

<sup>2</sup>[Diamond and Saez \(2011\)](#) summarize these recommendations.

<sup>3</sup>[Mankiw et al. \(2009\)](#), for instance, cast doubt on the identification of the ability distribution in [Saez \(2001\)](#), while [Mankiw \(2013\)](#) discusses various normative judgements which can invalidate the standard optimal tax formula used by [Diamond and Saez](#).

<sup>4</sup>Using tax return data, [Bakija et al. \(2012\)](#) document that executives, managers and supervisors account for about 40% of the top 0.1% of income earners in recent years. When managers and professionals in the financial sector are included the number grows to 60%.

<sup>5</sup>See [Mayer \(1960\)](#). [Rosen \(1982\)](#) shows that a positive scale-of-operations effects is necessary to reconcile stylized facts on managerial compensation.

skewness in the distribution of firm sizes and pre-tax income. Such a mapping from talent to observables has profound implications on our quantitative exercise, as we discuss below.

We characterize the constrained efficient allocation and show that it can be decentralized as a competitive equilibrium with taxes levied on income and on firm size (measured as the value of the hired workforce). The former tax is standard in the public finance literature, while the latter is not. A positive marginal tax on firm size forces managers to operate below optimal scale.

Our first result provides normative grounds for using firm size taxation. We find that, in general, the government should forgo efficiency in the allocation of labor in order to relax the incentive constraints of the managers. We isolate a sufficient condition on the technology for optimal firm distortions to disappear, and show that such condition is satisfied when the scale-of-operations effect is shut down, or when the technology is Cobb-Douglas. In this sense, our environment nests the well known [Diamond and Mirrlees \(1971\)](#) efficiency result as a special case.

We then look at optimal income taxation. We start the analysis by providing a formula for optimal marginal tax rates. As it is standard, our tax formula links marginal rates to the assumed distribution of talent, the redistributive motives of the policy maker, and the elasticity of labor supply of the manager. In addition, a positive scale-of-operations affects the level of optimal taxes via two channels: the first one is explicit in the tax formula, while the second one is only implicit.<sup>6</sup> Explicitly, the scale-of-operations impacts optimal taxes by shaping the relative productivity of skills and effort (the two components of managerial input), and by modifying the sensitivity of such a relationship to changes in labor supply. We show that the combined effect is positive when the technology displays constant elasticity of substitution between managerial and workers' inputs. In a nutshell, given a distribution of talent, larger values for the scale-of-operations generate more skewness in pre-tax income. This creates a force for higher marginal taxes to level the playing field.<sup>7</sup>

The implicit effect arising from a positive scale-of-operations is connected to the shape of the underlying density of managerial skills. More precisely, given a distribution of earnings or other observables, the intrinsic distribution of skills becomes *less* skewed as the scale-of-operations rises. This implies that any given dispersion in pre-tax income, for example, can now be rationalized with a smaller dispersion in skills. So in contrast with the explicit effect discussed previously, this channel actually *reduces* optimal marginal taxes (all else equal). The logic is now reversed: High income managers operating large firms are not as high talent as implied by

---

<sup>6</sup>As will become clear in the body of the paper, to break down the conventional expression for optimal income taxes it is also required that the elasticity of substitution between managerial and workers' inputs be different from one (Cobb-Douglas case).

<sup>7</sup>A rule of thumb for income tax design is that high marginal taxes are attractive when few individuals are affected at the margin, but many individuals are taxed inframarginally (and, hence, without distortion). This occurs, for instance, whenever the distribution of pre-tax income, or the distribution of skills exhibit high levels of right skewness (or "thick" right tails).

a model which ignores the scale-of-operations. Hence, those individuals should not be levied very high taxes, as implied by the textbook tax formula.

We take the model to the data to quantitatively evaluate the forces discussed above, and to provide specific normative recommendations on optimal marginal tax rates. A successful calibration requires the identification of two key objects: the parameter governing the scale-of-operations effect, and the distribution of managerial talent at the top. We identify these parameters by following the insights from [Rosen \(1982\)](#) (adding elastic labor). In particular, we start by deriving equilibrium restrictions relating the distribution of talent with firm size, sales, and profits of the firm. Using such conditions, we first show that the scale-of-operations effect can be written as a function of the elasticity of firm size with respect to sales, and the elasticity of managerial compensation relative to sales, both of which can be backed out from the data. We pin down these elasticities by using COMPUSTAT data, and by applying stylized facts on managerial compensation.

The second calibration object is the distribution of top talent. This is uncovered from the observed distribution of firm sizes, by exploiting the mapping between talent and firm size predicted by our Rosen-type model. Notably, here we depart from the established approach in public finance which recovers the ability distribution from the distribution of income instead (see, e.g., [Saez \(2001\)](#)). Given that income data is usually confidential and top coded, we believe that the alternative route that we propose has certain advantages when it comes to availability and reliability. We find that, in the presence of a positive scale-of-operations effect, the talent distribution is substantially more compact than in previous studies: Assuming that the right tail is Pareto distributed, our estimate on the tail parameter is an order of magnitude larger than what was previously identified in the literature.

The optimal top tax rate in our benchmark calibration is 32.4 percent. This number is significantly lower than what is obtained in standard environments, such as [Diamond and Saez \(2011\)](#), where top tax rates can be as high as 80 percent. Moreover, the top tax rate in the US tax code falls within the range of estimates in our calibration.<sup>8</sup> Our span of control production function and the implied compactness in the upper tail of the skill distribution are key in generating these results. For comparison, the optimal rate is equal to 65.4 percent when using the same calibration but absent a scale-of-operations effect. We also extend the analysis to a panel of non-US countries, including Australia, Canada and nine European countries. We find that benchmark optimal top tax rates are mostly concentrated within the 34-50 per cent range, and are strikingly similar across the nations in our sample. We also quantify the optimal marginal tax on firm size. We find it to be positive, progressive, and creating a markup of about 2 percent over the real wage at the top.

---

<sup>8</sup>[Saez et al. \(2012\)](#) report a top 1 percent marginal rate of approximately 42.5 percent for 2009. Using the Current Population Survey in the same period we find top marginal income tax rates of 33.5% for federal and 5% for state.

## RELATED LITERATURE

This paper touches on two large literatures: the first one concerns managerial compensation and the second one deals with the taxation of top income earners.

A growing literature has modeled CEOs with heterogeneous talent that map into firm performance and managerial compensation schemes. [Lucas \(1978\)](#) and [Rosen \(1982\)](#) provide early frameworks where the compensation of the CEO (the owner of the span of control technology) can be analyzed jointly with firm size. [Terviö \(2008\)](#), [Gabaix and Landier \(2008\)](#), [Edmans et al. \(2009\)](#), on the other hand, consider models where firm size is fixed exogenously, and the most productive managers are assigned to the largest firms.<sup>9</sup> The key in all of these models is that they introduce a nonlinear mapping between compensation and talent of the manager. In particular, the distribution of compensation is more positively skewed than the one for talent, which is the key mechanism in our paper.<sup>10</sup> Contributing to this line of literature, we model the intensive margin of managerial effort. This is a necessary step to think about top income taxation.

The literature on optimal taxation of top income earners is vast.<sup>11</sup> Methodologically, our contribution with respect to this literature is twofold. First, to the best of our knowledge, this is the first paper that recovers the distribution of talent consistent with production functions which are nonlinear in skills, and that studies the corresponding tax implications. Second, our environment is one in which compensation of the agent (the manager in our case) is endogenous. This is a departure from the classical taxation environment where wages are fixed exogenously. [Stiglitz \(1982\)](#) originally analyzed taxation under endogenous wages in a model where workers of different types interact within an aggregate production function. As we clarify below, though, the nature of wage endogeneity in our model is quite different from the one that Stiglitz considers. More recent work on taxation under endogenous wages include [Slavík and Yazıcı \(2014\)](#) and [Ales et al. \(2015\)](#). The former focus on the endogenous accumulation of heterogeneous forms of capital that interact differently with agents of diverse talent, while the latter study an assignment problem of workers with heterogeneous talents to tasks with heterogeneous complexity.

Given that our approach is to map top income earners to managers, [Rothschild and Scheuer \(2013\)](#) and [Scheuer \(2014\)](#) are also related to our work.<sup>12</sup> These authors consider

---

<sup>9</sup>Other theoretical contributions in this area include [Baker and Hall \(2004\)](#), [Edmans and Gabaix \(2011\)](#), [Baranchuk et al. \(2011\)](#), and [Eisfeldt and Kuhnen \(2013\)](#). An incomplete list of empirical studies emphasizing the key role of CEO abilities include [Bertrand and Schoar \(2003\)](#), [Adams et al. \(2005\)](#), [Bennedsen et al. \(2007\)](#), [Kaplan et al. \(2012\)](#), [Custódio and Metzger \(2013\)](#), and [Custódio et al. \(2013\)](#).

<sup>10</sup>This important feature also differentiates our paper from [Baranchuk et al. \(2011\)](#), who embed moral hazard in a superstars model and derive implications on firm size and CEO compensation. Differently, we model explicitly the managerial scale-of-operations effect. The skewness of firm size distribution endogenously generated in our model is not only driven by the complementarity between talent and managerial effort (as in [Baranchuk et al. \(2011\)](#)), but also from the complementarity between talent and span-of-control.

<sup>11</sup>For a review refer to [Mankiw et al. \(2009\)](#) and [Diamond and Saez \(2011\)](#).

<sup>12</sup>The environment in this paper is static. For dynamic models that consider the modeling and taxation of entrepreneurial wealth, refer to [Quadrini \(2000\)](#), [Cagetti and Nardi \(2006\)](#), [Albanesi \(2011\)](#) or [Shourideh \(2012\)](#).

an environment where agents are characterized by a multidimensional skill/taste vector and decide whether to be a worker or a manager. These papers isolate a force for lowering top taxes, like we do, but our mechanism is quite different from theirs. Specifically, in the spirit of [Stiglitz \(1982\)](#), [Rothschild and Scheuer \(2013\)](#) and [Scheuer \(2014\)](#) obtain that reducing taxes on individuals with high ability increases the productivity of lower types, therefore relaxing incentive constraints. In our framework, on the other hand, managerial effort has no effect on other managers' productivities as they operate separate firms.<sup>13</sup> Instead the main mechanism which lowers top taxes in our paper is connected to the calibration of the skill distribution.<sup>14</sup>

[Piketty et al. \(2014\)](#) also considers a model of CEO taxation. In that paper, the CEO can extract surplus by imposing a negative externality on workers, thus raising her own compensation above her marginal product. This channel provides an upward pressure on marginal tax rates, which corrects for the negative CEO externality. Our approach to modeling managers, on the other hand, is more in line with the empirical evidence in [Kaplan and Rauh \(2013\)](#), according to which managers generate a positive externality on workers rather than a negative one.

Finally, this paper is related to the recent works of [Ales and Sleet \(2015\)](#) and [Scheuer and Werning \(2015\)](#). Both of these papers study an optimal managerial/superstar taxation problem in competitive assignment environments. In these papers, top income earners of heterogeneous abilities are matched to fixed factors (firm assets), as in [Terviö \(2008\)](#) and [Gabaix and Landier \(2008\)](#). In our paper, on the other hand, firm size is endogenous and it is affected by managerial ability as well as the tax code. One way to interpret this difference is to view [Ales and Sleet \(2015\)](#) and [Scheuer and Werning \(2015\)](#) as studying short run taxation implications where firm sizes do not adjust, while we provide a framework for the design of taxation in the long run where firm size is allowed to vary.

The remainder of the paper is organized as follows. In Section 2 we describe the environment. In Section 3 we characterize Pareto optimality. In Section 4 we look at the decentralization of the optimum and derive the optimal tax rate formula. In Section 5 we show our identification strategy and calibrate the model. In Section 6 we discuss the quantitative results on income taxes for the US, and Section 7 computes income taxes across a panel of countries. In Section 8 we analyze optimal firm size distortions. Section 9 concludes.

---

<sup>13</sup>Effort of the manager does affect the productivity of the workers, but given that the latter are subject to a different tax schedule this has no impact on managerial incentives to work below full potential. See Section 6.1 for details.

<sup>14</sup>[Golosov et al. \(2015\)](#) and [Huggett and Badel \(2015\)](#) uncover different forces lowering optimal top tax rates in dynamic models.



## 2 ENVIRONMENT

The economy is static and it is populated by a unit measure of workers and a unit measure of managers. There is a single consumption good. Managers have quasi-linear preferences over consumption  $c$  and effort  $n$ , which are represented by the utility function

$$U(c, n) = c - v(n),$$

where  $v : \mathbb{R}_+ \rightarrow \mathbb{R}$  and is twice continuously differentiable with positive derivatives. Our specification for preferences abstracts from income effects.<sup>15</sup>

Workers have preferences over consumption and supply labor inelastically. Without loss of generality, we normalize the disutility from effort of the worker to zero and the amount of effective effort supplied to one. Consumption of the worker is denoted by  $c^w \in \mathbb{R}_+$ .

Managers are heterogeneous with respect to managerial talent, denoted by  $\theta \in \Theta$  with  $\Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+ \setminus 0$ . Managerial talent,  $\theta$ , is distributed according to the cumulative distribution function  $F : \Theta \rightarrow [0, 1]$  with density function  $f : \Theta \rightarrow \mathbb{R}_+$ . Following [Rosen \(1982\)](#) and [Lucas \(1978\)](#), managers operate a span of control technology. Specifically, managers with talent  $\theta$  produce final output  $y$  according to:

$$y(n, L, \theta) = \theta^\gamma H(\theta \cdot n, L), \quad (1)$$

where  $L$  is hired labor and  $\gamma$  is the scale-of-operation parameter. We focus on the case with  $\gamma > 0$ . The production function  $H : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is concave, strictly increasing in both arguments and features continuous derivatives. We further assume that  $y_{n\theta} \geq 0$ .

Managerial talent enters the production function (1) in two ways. First,  $\theta$  is effort-augmenting, as it multiplies  $n$  within  $H$ . Second,  $\theta$  improves total factor productivity since  $\gamma > 0$ . We refer to the latter effect  $\gamma$  as the *scale-of-operations effect*, following [Mayer \(1960\)](#).<sup>16</sup> This formulation is in line with the one in [Rosen \(1982\)](#), where managers' actions naturally affect the productivity of all workers under their supervision (irrespective of their number). But unlike the technology in Rosen's paper, we also incorporate elastic managerial effort  $n$  as an intensive margin.

In what follows, it is convenient to define  $n(y, L, \theta)$  as the effort required by a manager of talent  $\theta$  to generate output  $y$  when hired labor is  $L$ . An allocation in this economy is then defined as  $(c^w, c, y, L)$ , where  $c^w \in \mathbb{R}_+$ ,  $c : \Theta \rightarrow \mathbb{R}_+$ ,  $L : \Theta \rightarrow \mathbb{R}_+$ ,  $y : \Theta \rightarrow [0, \bar{y}]$ , and  $0 < \bar{y} < \infty$ . We assume that  $c(\theta)$ ,  $y(\theta)$  and  $L(\theta)$  are observable, while  $\theta$ ,  $n(\theta)$  and  $\theta \cdot n(\theta)$  are private information to each  $\theta$ -agent.

<sup>15</sup>Empirical analyses typically indicate that income effects are relatively small compared to substitution effects. See, e.g., [Blundell and MaCurdy \(1999\)](#).

<sup>16</sup>See [Bartelsman and Doms \(2000\)](#) and references therein for additional details on the relationship between managerial talent, firm size and firm productivity.

For a given level of (exogenous) government consumption  $G > 0$ , an allocation is feasible if

$$c^w + \int_{\Theta} c(\theta) dF(\theta) + G \leq \int_{\Theta} y(\theta) dF(\theta), \quad (2)$$

and

$$\int_{\Theta} L(\theta) dF(\theta) \leq 1. \quad (3)$$

Social welfare is evaluated according to the social welfare function

$$SWF = \Psi(c^w) + \int_{\Theta} \Psi(c(\theta) - v(n(y(\theta), L(\theta), \theta))) dF(\theta),$$

where  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  is a strictly increasing, differentiable and concave function which summarizes social preferences for redistribution. In particular, we refer to  $\Psi'(c(\theta) - v(n(\theta)))$  as the *social marginal welfare weight* on managers of talent  $\theta$ .

### 3 PARETO OPTIMALITY

In this section, we characterize Pareto optimal allocations using a direct mechanism where managers report their talent  $\theta$  to a social planner and are assigned an allocation for consumption  $c(\theta)$ , output  $y(\theta)$  and labor  $L(\theta)$  accordingly. Define  $n(y(\theta'), L(\theta'), \theta)$  as the level of effort exerted by a manager of talent  $\theta$  who mimics a manager of talent  $\theta'$ . In this case, manager  $\theta$  is assigned  $L(\theta')$  workers and is required to produce output  $y(\theta')$ . An allocation is *incentive-compatible* when truthful revelation is optimal for all managers, which requires:

$$c(\theta) - v(n(y(\theta), L(\theta), \theta)) \geq c(\theta') - v(n(y(\theta'), L(\theta'), \theta)), \quad \forall \theta, \theta' \in \Theta. \quad (4)$$

Pareto optimal allocations solve the following social planner's problem:

$$\begin{aligned} \max_{c^w, \{c(\theta), y(\theta), L(\theta)\}_{\theta \in \Theta}} \quad & \Psi(c^w) + \int_{\Theta} \Psi(c(\theta) - v(n(y(\theta), L(\theta), \theta))) dF(\theta), \\ \text{s.t.} \quad & (2), (3) \text{ and } (4). \end{aligned} \quad (\text{PO})$$

The next proposition provides a useful characterization of incentive compatibility with quasi-linear utility.

**Proposition 1.** *Let  $n(\theta) \equiv n(y(\theta), L(\theta), \theta)$ ,  $U(\theta) \equiv c(\theta) - v(n(\theta))$ , and denote by  $n_y$ ,  $n_L$  and  $n_\theta$ , the first derivatives of  $n(y, L, \theta)$  with respect to its first, second and third arguments, respectively, with similar notation for its second derivatives. Then incentive compatibility constraints (4) hold*



if and only if for all  $\theta \in \Theta$ :

$$U'(\theta) = -v'(n(y(\theta), L(\theta), \theta))n_\theta(y(\theta), L(\theta), \theta), \quad (5)$$

and

$$\frac{v''(n(\theta))}{v'(n(\theta))^2} c'(\theta) + \frac{n_{\theta y}(\theta)}{n_\theta(\theta)} y'(\theta) + \frac{n_{\theta L}(\theta)}{n_\theta(\theta)} L'(\theta) \geq 0. \quad (6)$$

*Proof.* See Appendix A.  $\square$

As it is standard in the optimal taxation literature, from here onwards we assume that the monotonicity condition (6) is satisfied at the optimum.<sup>17</sup> The condition holds, for example, when  $c$ ,  $y$  and  $L$  are increasing in  $\theta$  and  $n_{\theta L}$  is small enough.<sup>18,19</sup> In practice, we apply a *first order approach* to the planning problem, in which the original set of constraints (4) is replaced by local first order conditions. The relaxed version of the planner's problem can be written as:

$$\max_{c^w, \{U(\theta), y(\theta), L(\theta)\}_{\theta \in \Theta}} \Psi(c^w) + \int_{\Theta} \Psi(U(\theta)) dF(\theta), \quad (\text{PO-FOC})$$

$$\text{s.t.} \quad \int_{\Theta} [y(\theta) - c^w - U(\theta) - v(n(y(\theta), L(\theta), \theta))] dF(\theta) = G, \quad (7)$$

$$\int_{\Theta} L(\theta) dF(\theta) = 1, \quad (8)$$

$$U'(\theta) = -v'(n(y(\theta), L(\theta), \theta))n_\theta(y(\theta), L(\theta), \theta). \quad (9)$$

Before presenting the decentralization and showing the properties of the optimal tax system, we discuss the incentive constraint (5) further. In the standard Mirrleesian environment, output, talent, and effort are related through  $y(\theta) = \theta \cdot n(\theta)$ , so that effort required to generate output  $y$  is  $n(y, \theta) = y/\theta$ . In that case, the term  $n_\theta$  appearing on the right hand side of (5) is given by  $n_\theta(y, \theta) = -n(\theta)/\theta$  and, hence, the incentive constraint can be written *only* in terms of the level of effort.

In our environment, on the other hand, the incentive constraints will generally incorporate labor as an additional margin for incentive provision. To be precise, suppose (as in Rosen (1982)) that  $H$  satisfies constant returns to scale. Then by definition  $y = \theta^\gamma \cdot L \cdot h\left(\frac{\theta n(y, L, \theta)}{L}\right)$ , where  $h(x) = H(x, 1)$  for all  $x \geq 0$ . Consequently,  $n(y, L, \theta) = h^{-1}\left(\frac{y}{\theta^\gamma L}\right) \frac{L}{\theta}$  and

$$n_\theta(y, L, \theta) = -h^{-1}\left(\frac{y}{\theta^\gamma L}\right) \frac{L}{\theta^2} - \frac{L}{\theta} \frac{1}{h'\left(h^{-1}\left(\frac{y}{\theta^\gamma L}\right)\right)} \frac{y}{L} \frac{\gamma}{\theta^{\gamma+1}},$$

<sup>17</sup>In our numerical simulations we verify the validity of this assumption. See Scheuer (2014) for a recent example of this approach. Kapička (2013), Golosov et al. (2015), and Farhi and Werning (2013) apply similar techniques in dynamic environments.

<sup>18</sup>It is straightforward to verify that  $n_\theta, n_{\theta y} \leq 0$  and  $n_{\theta L} \geq 0$ .

<sup>19</sup>Relative to the standard Mirrleesian environment, where it is required that  $c$  and  $y$  be increasing in talent, (6) also imposes a condition on the allocation of  $L$ .

which simplifies to

$$n_\theta(y, L, \theta) = -\frac{n(\theta)}{\theta} - \frac{\gamma}{\theta^2} \frac{H(\theta n, L)}{h'(\frac{\theta n}{L})}. \quad (10)$$

The first term in (10) is the same expression which emerges in the standard Mirrleesian environment discussed previously. Indeed, when  $\gamma = 0$  the incentive constraint boils down to the traditional formulation of Mirrlees. In contrast, the second term in (10) is novel to our environment. Through this term,  $n_\theta$  and, hence, the right hand side of the incentive constraint (5), will explicitly depend on  $L$ . This implies that by properly choosing labor, the social planner may relax the incentive constraint. In the sections below we show that the distorted choice of  $L$  is implemented with a nonlinear tax on firm size, which is at the cost of production efficiency.

Before closing this section, it is worth noting that the ability of the planner to affect incentives by changing  $L$  depends on the functional form of  $H$ . In particular, suppose that  $H$  is Cobb-Douglas so that:  $y = \theta^\gamma (\theta \cdot n)^\alpha L^{1-\alpha}$  for some  $0 < \alpha < 1$ . We have:

$$n_\theta(y, L, \theta) = -\frac{n(\theta)}{\theta} - \frac{\gamma}{\theta^2} \frac{(\theta \cdot n)^\alpha L^{1-\alpha}}{\alpha(\theta \cdot n)^{\alpha-1} L^{1-\alpha}} = -\frac{n(\theta)}{\theta} (1 + \gamma/\alpha), \quad (11)$$

so that in this case the incentive constraint does not explicitly depend on labor. In our discussion below, however, we focus on a more general formulation for  $H$  featuring constant elasticity of substitution (CES). In fact, our calibration exercise demonstrates the Cobb-Douglas case is not empirically relevant when it comes to reconciling a positive scale-of-operations effect using firm level data.

## 4 OPTIMAL TAXATION

Next we construct a decentralization of the optimum in (PO) that relies on nonlinear taxes on firm size ( $T_L$ ) and nonlinear taxes on income ( $T$ ). We then discuss certain properties of these tax functions.

In the decentralized environment, managers of talent  $\theta$  solve the following problem, taking wages and tax rates as given:

$$\begin{aligned} \max_{c, y, L} \quad & c - v(n(y, L, \theta)) \\ \text{s.t.} \quad & c \leq y - wL - T_L(wL) - T(y - wL - T_L(wL)), \end{aligned} \quad \begin{aligned} & \text{(MP)} \\ & (12) \end{aligned}$$

where  $w \in \mathbb{R}_+$  is the real wage,  $T : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a nonlinear income tax, and  $T_L : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a nonlinear tax on firm size. Since workers in our environment supply labor inelastically, their problem is characterized by a simple budget constraint  $c^w = w + \phi$ , where  $\phi$  is a government transfer to the worker. We can now define a competitive equilibrium for our environment:

**Definition 1.** For a given level of government consumption  $G$ , a tax distorted competitive equilibrium is an allocation  $\{c, y, L\}$ , a tax system  $\{T, T_L, \phi\}$ , and a wage  $w$  such that:

1. Taking as given  $\{w, T, T_L\}$  each  $\theta$ -manager solves (MP);
2. The worker's budget constraint holds:  $c^w = w + \phi$ ;
3. Goods and labor markets clear: equations (7) and (8) hold;
4. The government's budget constraint is balanced:

$$\int [T(y(\theta) - wL(\theta) - T_L(wL(\theta))) + T_L(wL(\theta))] dF(\theta) = G + \phi. \quad (13)$$

By applying a version of the *taxation principle* (see, e.g., [Guesnerie \(1981\)](#)) we derive the following proposition:

**Proposition 2.** *Let  $X \equiv \{c^w, (c(\theta), y(\theta), L(\theta))_{\theta \in \Theta}\}$  be an optimal allocation solving (PO). Then there exist a tax system  $\{T, T_L, \phi\}$  and a wage  $w$  such that  $X$  can be decentralized as a tax distorted competitive equilibrium.*

*Proof.* See Appendix [B.1](#). □

We refer to the tax system that implements the allocation  $X$  above as the optimal one. We next proceed to characterize the optimal tax system.

#### 4.1 FIRM SIZE TAXATION

We begin by looking at the distortions on firm size implied by the constrained efficient allocation. Such distortions provide a normative rationale for firm size taxation in the decentralization. Assuming differentiability of  $T_L$ , first order conditions from the manager's problem (MP) give:

$$w(1 + T'_L(wL(\theta))) = y_L(\theta n(\theta), L(\theta), \theta), \quad (14)$$

where  $y_L$  is the marginal product of the worker.<sup>20</sup> Equation (14) shows that if  $T'_L(wL(\theta)) \neq 0$  for some  $\theta$ , the worker's marginal product is not be equalized across firms. This implies a break down of the well known [Diamond and Mirrlees \(1971\)](#) productive efficiency result.

The reason behind the willingness of the planner to distort labor allocations lies in the ability to relax incentive constraints by affecting  $L$  as discussed at the end of Section [3](#). The next proposition provides a formula for optimal firm size distortions and gives sufficient conditions under which it is optimal not to distort firm level employment.

**Proposition 3.** *Let  $\{y(\theta), n(\theta), L(\theta)\}_{\theta \in \Theta}$  be a solution of (PO-FOC) and let  $T_L$  be part of an optimal tax system. Then:*

---

<sup>20</sup>Here we use that  $y_L = -n_L/n_y$ , which simply follows from the implicit function theorem.

1. Optimal marginal firm size distortions satisfy:

$$T'_L(wL(\theta)) = \frac{f(\theta) - \mu(\theta)v'(n(\theta))\frac{n_{\theta L}(\theta)}{\lambda^l}}{f(\theta) + \mu(\theta)v'(n(\theta))\frac{n_{\theta y}(\theta)}{\lambda^r}} - 1, \quad (15)$$

where  $\mu(\theta) \geq 0$  is the multiplier on the incentive constraint (5), and  $\lambda^l > 0$  and  $\lambda^r > 0$  are the multipliers on (8) and (7), respectively.

2. If there is a differentiable  $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $n_\theta = g(n(\theta))$  for all  $\theta \in \Theta$ , then  $T'_L(wL(\theta)) = 0$  for all  $\theta \in \Theta$ .

*Proof.* See Appendix B.2. □

Part 1 of Proposition 3 illustrates how, in general, a binding incentive constraint generates nonzero firm level distortions. Part 2 provides a sufficient condition for distortions to disappear. Following the discussion at the end of Section 3, it is immediate that this condition is satisfied in the case in which  $\gamma = 0$  and  $H$  is constant return to scale, or when  $H$  is Cobb-Douglas.<sup>21</sup>

## 4.2 INCOME TAXATION

We now move to income taxation. Assuming differentiability of  $T'$ , first order conditions from the manager's problem (MP) imply:

$$1 - T'(\pi(\theta)) = v'(n(y(\theta), L(\theta), \theta))n_y(y(\theta), L(\theta), \theta), \quad (16)$$

where  $\pi(\theta) \equiv y(\theta) - wL(\theta) - T_L(L(\theta))$  corresponds to income of a manager with talent  $\theta$ .

For the rest of the analysis, we make a standard assumption on preferences:

**Assumption 1.** The disutility for effort is isoelastic:  $v(n) = n^{1+\frac{1}{\varepsilon}}/(1 + \frac{1}{\varepsilon})$  with  $\varepsilon > 0$ .

Under Assumption 1, the Frisch elasticity of labor supply is constant and equals  $\varepsilon$ . This only simplifies the expressions that follow without loss of generality in the interpretation. The next proposition characterizes optimal marginal income tax rates. To simplify notation, we refer to  $T'(\pi(\theta))$  as  $T'(\theta)$ , unless stated otherwise.

**Proposition 4.** Suppose Assumption 1 holds. Let  $\{y(\theta), n(\theta), L(\theta)\}_{\theta \in \Theta}$  be a solution of (PO-FOC), and let  $\{T, T_L, w\}$  be an optimal tax system. We have that for all  $\theta$ :

$$\frac{T'(\theta)}{1 - T'(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \cdot \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \cdot \frac{y_\theta}{y_n} \frac{\theta}{n} \cdot \left[\frac{1}{\varepsilon} + \frac{d \ln \left(\frac{y_\theta}{y_n}\right)}{dn}\right], \quad (17)$$

<sup>21</sup>Scheuer (2014) considers an environment similar to ours with two important differences. First, the firm level production function features  $\gamma = 0$ . Second, workers and managers are heterogeneous in two dimensions. He considers two cases: one in which the government can tax workers and managers differently, and one in which it cannot. This second case features firm level distortions. However, this is not due to the presence of  $L$  in the incentive constraint, but because of an additional “no-discrimination” constraint which is absent in our environment.

where

$$D(\theta) \equiv \frac{1}{1 - F(\theta)} \int_{\theta}^{\bar{\theta}} \Psi'(U(\theta)) dF(\theta).$$

*Proof.* See Appendix B.3. □

Equation (17) reveals the main forces that shape optimal marginal income tax rates in our framework. The first two terms are well known and are present in previous contributions. The first term looks at the effect of the shape of the talent distribution on marginal tax rates. In particular, high marginal taxes at talent level  $\theta$  are attractive when the mass of managers above  $\theta$ , given by  $(1 - F(\theta))$ , is large. At the same time, the resulting distortion is proportional to the mass of individuals at  $\theta$  and to their productivity level, explaining the negative dependence on  $\theta f(\theta)$ . The second term summarizes the redistributive tastes of the government. Unsurprisingly, optimal marginal taxes on a given manager decreases as the corresponding social welfare weight rises.

The third expression in the formula is novel and captures the relative contribution of the two components of managerial input, i.e. skills and effort, into output. More precisely,  $(y_{\theta}/y_n) \cdot (\theta/n)$  is the ratio between the output elasticity of talent, and the output elasticity of effort. When this term is large, high marginal taxes are warranted. The reason is that, in such a scenario, the relative contribution of innate skills to income inequality is more pronounced than the contribution of elastic effort. Consequently, redistributive taxes can level the playing field without triggering large efficiency costs.

The last term in square brackets represents the impact of effort responses on optimal tax rates. The standard labor supply channel is encapsulated by the Frisch elasticity  $\varepsilon$ , which measures how changes in after-tax wages affect the labor supply (holding the marginal utility of wealth constant). A low Frisch elasticity translates into a small effort response to tax hikes, which raises optimal marginal tax rates. But labor supply variations also modify the relative contribution of skills and effort into output and, hence, compensation. This effect is captured by the elasticity  $d \ln(y_{\theta}/y_n) / dn$ . Given that this coefficient is positive,<sup>22</sup> reductions in effort lower the relative productivity of talent, which is the source of managerial rents. In turn, a high value of this elasticity deters the manager from reducing effort, thus providing a rationale for high tax rates.

Our tax formula nests traditional expressions for optimal income taxes in the literature. In particular, suppose that  $\gamma = 0$  so that  $y = H(\theta n, L)$ . In that case, it is straightforward to show that:

$$\frac{y_{\theta}}{y_n} \frac{\theta}{n} = 1, \quad \text{and} \quad \frac{d \ln \left( \frac{y_{\theta}}{y_n} \right)}{d \ln n} = \frac{y_{\theta n}}{y_{\theta}} n - \frac{y_{nn}}{y_n} n = 1.$$

Substituting into (17) we obtain the classical tax formula from [Diamond \(1998\)](#) or [Saez \(2001\)](#)

---

<sup>22</sup>Recall that it is assumed that  $y_{\theta n} \geq 0$  and  $y_{nn} \leq 0$ .

in terms of underlying structural parameters:

$$\frac{T'(\theta)}{1 - T'(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \left( 1 - \frac{D(\theta)}{D(\underline{\theta})} \right) \left( \frac{1}{\varepsilon} + 1 \right). \quad (\text{DS})$$

The next section untangles how a positive scale-of-operations parameter generates a departure from the standard tax formula in (DS).

#### 4.3 TOP INCOME TAXATION UNDER CES TECHNOLOGY

Proposition 4 applies to quite general production functions  $H$ . To lay the groundwork for our quantitative analysis we make the following parametric assumption on the production function:<sup>23</sup>

**Assumption 2.** *The production function has constant elasticity of substitution:*

$$y(n(\theta), L(\theta), \theta) = \theta^\gamma [\beta(\theta n(\theta))^\rho + (1 - \beta)L(\theta)^\rho]^{\frac{1}{\rho}},$$

where  $\rho \in [-\infty, 1]$  and the elasticity of substitution between  $\theta n(\theta)$  and  $L(\theta)$  is given by  $\frac{1}{1-\rho} \in [0, \infty]$ .

The following corollary characterizes optimal income taxes under Assumption 2.<sup>24</sup>

**Corollary 1.** *Suppose Assumptions 1 and 2 hold, and that (6) is satisfied. Then at any Pareto optimum,  $T'(\theta)$  satisfies*

$$\frac{T'(\theta)}{1 - T'(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \left( 1 - \frac{D(\theta)}{D(\underline{\theta})} \right) \underbrace{\left( 1 + \frac{\gamma}{1 - \kappa(\theta)} \right)}_{=\frac{y_\theta}{y_n} \frac{\theta}{n}} \left[ \frac{1}{\varepsilon} + \underbrace{\frac{1 + \frac{\gamma}{1 - \kappa(\theta)} (1 - \rho \kappa(\theta))}{1 + \frac{\gamma}{1 - \kappa(\theta)}}}_{=d \ln \left( \frac{y_\theta}{y_n} \right) / d \ln n} \right], \quad (18)$$

where  $\kappa(\theta) \equiv y_L(\theta)L(\theta)/y(\theta)$  is the share of labor costs to total sales for managers of talent  $\theta$ .

*Proof.* See Appendix B.4. □

**The Scale-of-Operations Effect** The parameter  $\gamma$  impacts optimal tax rates via three channels. The first two are explicit in formula (18), but the last one is not. We discuss each effect in turn.

First, a larger  $\gamma$  leads to a larger output-talent elasticity relative to effort through  $(y_\theta/y_n) \cdot (\theta/n)$ , which translates into higher marginal taxes. Second, under the CES specification  $d \ln (y_\theta/y_n) / d \ln n$  decreases with  $\gamma$ . In words, the relative weight of skills and effort into output

<sup>23</sup>The estimation procedure described in Section 5 can be extended to other production functions as long as the production function satisfies constant returns to scale.

<sup>24</sup>Note that under Assumption 2 we specialize the discussion to a constant return to scale production function. Refer to Scheuer and Werning (2015) for additional details on taxation with production functions that do not satisfy constant returns to scale.



becomes less sensitive to changes in  $n$  as  $\gamma$  grows. The latter leads to lower marginal taxes, as discussed in the previous section. Under CES technology, though, the first effect dominates, which can be verified by inspecting (18).

The third effect is implicit in the tax formula. In a nutshell, *given a distribution of earnings* or other observables, the underlying distribution of skills becomes less “spread out” when  $\gamma$  grows. This effect reduces optimal tax rates through the first term on the right hand side of (18). While this channel will be analyzed in detail in Section 5 below, here we provide a heuristic illustration.

Consider an observable variable  $z$  (such as income or firm size) which is monotonic in  $\theta$ , with distribution  $F_z$  and density  $F'_z = f_z$ . By construction,  $f(\theta) = f_z(z)dz/d\theta$  which implies

$$\frac{1 - F(\theta)}{\theta f(\theta)} = \frac{1 - F_z(z)}{z f_z(z)} \left( \frac{d \ln z}{d \ln \theta} \right)^{-1}, \quad (19)$$

so that the left hand side, and hence optimal tax rates, are inversely related to the elasticity of  $z$  with respect to  $\theta$ . Crucially, in a decentralized environment, such an elasticity increases with  $\gamma$ , which is the nature of superstar effects considered by Rosen. In that sense, a positive scale-of-operations effect is consistent with a distribution for managerial ability which is much less compressed than what previous studies have considered.

**Top Income Taxation** To provide a benchmark for the optimal top income tax rate we make the following assumption:

**Assumption 3.**

(a) *The talent distribution has a right Pareto tail with parameter  $a > 0$ :*

$$\lim_{\theta \rightarrow \bar{\theta}} \frac{1 - F(\theta)}{\theta f(\theta)} = \frac{1}{a}.$$

(b) *There is zero social marginal welfare weight at the top:  $\lim_{\theta \rightarrow \bar{\theta}} D(\theta) = 0$ .*<sup>25</sup>

Let  $\lim_{\theta \rightarrow \bar{\theta}} \kappa(\theta) = \bar{\kappa}$ . Taking limits on (18) and using Assumption 3 we get an expression for the optimal marginal tax rate at the top:

**Corollary 2.** *Suppose Assumptions 1, 2, and 3 hold. Then at any Pareto optimum the optimal marginal tax at the top satisfies:*

$$\frac{T'(\bar{\theta})}{1 - T'(\bar{\theta})} = \frac{1}{a} \left( \frac{1}{\varepsilon} + 1 \right) + \frac{1}{a} \frac{\gamma}{1 - \bar{\kappa}} \left( \frac{1}{\varepsilon} + 1 - \rho \bar{\kappa} \right). \quad (20)$$

---

<sup>25</sup>Assuming zero social weight at the top provides a benchmark that allows easy comparison with the bulk of the optimal taxation literature. It is clearly an extreme assumption. Weinzierl (2014) looks at survey evidence on preferred societal welfare criterions. Saez and Stantcheva (2015) derive tax formulas under arbitrary marginal social weights.

In the next section, we take equation (20) to the data and quantitatively evaluate the overall effect of the scale-of-operations effect on optimal tax rates.

## 5 IDENTIFICATION AND ESTIMATION

The tax formula in (20) provides insights on the forces that shape top marginal tax rates. In this section we quantify these forces. Two parameters which are crucial for this quantitative evaluation are the scale-of-operations parameter  $\gamma$  and the Pareto tail parameter of the talent distribution  $a$ . Our main contribution in this section is to show how to estimate such parameters using firm level data.

As a first step, in Section 5.1 we derive equilibrium restrictions which relate the distribution of talent with firm size, sales, and profits. Those relationships are then used in Sections 5.2 and 5.3 to estimate  $\gamma$  and  $a$ , respectively.

### 5.1 FIRM LEVEL ELASTICITIES

The approach is along the lines of Rosen (1982). We make the following assumption:

**Assumption 4.** *The production function  $H$  satisfies constant returns to scale.*

Consider a competitive equilibrium where the manager faces a linear tax  $\tau$  on her income, pays wage  $w$  (taken as given) to each unit of labor input  $L$ , and gets a fraction  $\chi \in (0, 1]$  of total profits.<sup>26</sup> Due to the constant returns to scale assumption, we can write the  $\theta$ -manager's problem as

$$\begin{aligned} \max_{\{L, n, \pi\}} \quad & (1 - \tau)\chi\pi - v(n) \\ \text{s.t.} \quad & \pi = \left[ \theta^\gamma L h\left(\frac{\theta n}{L}\right) - wL \right], \end{aligned} \tag{21}$$

where  $h\left(\frac{\theta n}{L}\right) \equiv H\left(\frac{\theta n}{L}, 1\right)$ ,  $h' > 0$ , and  $h'' < 0$ .

The first order conditions with respect to  $L$  and  $n$  in (21) are given by:

$$\theta^\gamma \left[ h\left(\frac{\theta n}{L}\right) - \frac{\theta n}{L} h'\left(\frac{\theta n}{L}\right) \right] = w, \tag{22}$$

and

$$(1 - \tau)\chi\theta^{\gamma+1} h'\left(\frac{\theta n}{L}\right) = v'(n). \tag{23}$$

Equations (22) and (23) together imply the following lemma where we show how firm size, output, and profits grow with respect to managerial talent.

---

<sup>26</sup>The assumption that managers are subject to a constant marginal tax rate is motivated by the progressivity of the US income tax system together with the fact that managers, in general, are located at the top of the income distribution.

**Lemma 1.** Suppose Assumptions 1 and 4 hold. Let  $\{L(\theta), n(\theta), \pi(\theta)\}$  solve the  $\theta$ -manager's problem in (21), and let  $\sigma$  be the elasticity of substitution between  $\theta n$  and  $L$ . Then the following relationships hold:

$$\frac{d \ln L(\theta)}{d \ln \theta} = 1 + \frac{\gamma \sigma}{1 - \kappa(\theta)} + \varepsilon \left( 1 + \frac{\gamma}{1 - \kappa(\theta)} \right), \quad (24)$$

$$\frac{d \ln y(\theta)}{d \ln \theta} = 1 + \gamma + \varepsilon \left( 1 + \frac{\gamma}{1 - \kappa(\theta)} \right) + \frac{\kappa(\theta)}{1 - \kappa(\theta)} \gamma \sigma, \quad (25)$$

$$\frac{d \ln \pi(\theta)}{d \ln \theta} = \left( 1 + \frac{\gamma}{1 - \kappa(\theta)} \right) (1 + \varepsilon), \quad (26)$$

where  $\kappa(\theta) \equiv wL(\theta)/y(\theta)$ .

*Proof.* See Appendix C.1. □

Lemma 1 reveals a key property of the scale-of-operations effect: for a given distribution of talent, the distributions of firm size, sales, and profits become more *skewed* as  $\gamma$  grows. In particular once we shut down the scale-of-operations effect ( $\gamma = 0$ ) then the growth rates in (24)-(26) are all equal to  $(1 + \varepsilon)$ . However, in general,  $L$ ,  $y$ , and  $\pi$  respond *more* to differences in managerial talent when  $\gamma$  is positive.<sup>27</sup> Finally, it is worthwhile observing that as long as the ownership share  $\chi$  and the top marginal income tax rate  $\tau$  are constant across talents, they do not impact the growth rates of firm size, output, and profits in Lemma 1 and, hence, it does not bias our estimation strategy.<sup>28</sup>

## 5.2 ESTIMATING THE SCALE-OF-OPERATIONS EFFECT $\gamma$

Using the expressions in Lemma 1 we obtain the next proposition:

**Proposition 5.** Suppose Assumptions 1 and 4 hold. Then for any solution of the  $\theta$ -manager's problem in (21) the following relationship holds:

$$\gamma = \frac{1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}}{\frac{d \ln L(\theta)}{d \ln y(\theta)} - \frac{d \ln \pi(\theta)}{d \ln y(\theta)} \frac{\sigma + \varepsilon}{1 + \varepsilon}}. \quad (27)$$

*Proof.* See Appendix C.2. □

Equation (27) forms the basis for the estimation of  $\gamma$ . Specifically, to evaluate  $\gamma$  we require the elasticity of firm size with respect to sales ( $d \ln L(\theta)/d \ln y(\theta)$ ), the elasticity of managerial

<sup>27</sup>This property is at the heart of Rosen's analysis. In his own words: "This is what maintains the observed long-tailed distribution of income at the top ranks: the distribution of rewards is more skewed than the distribution of talent (...)" (Rosen (1982), page 317). In fact, it is straightforward to verify that if  $\varepsilon = 0$  the relationships in Lemma 1 map to equations (14)-(16) in his paper.

<sup>28</sup>If managers were subject to a nonlinear income tax schedule, the formulas in Lemma 1 would incorporate terms with second derivatives of the tax function. However, the impact of those terms is negligible at the top, given that the tax schedule is progressive but concave in US data. The widely used parameterization in Gouveia and Strauss (1994), for example, has this implication. In fact, under that specification the marginal tax rate converges to a constant.

compensation with respect to sales ( $d \ln \pi(\theta)/d \ln y(\theta)$ ), the elasticity of substitution ( $\sigma$ ) and the Frisch elasticity ( $\varepsilon$ ). Below we discuss how each of these are estimated.

**Elasticity of firm size with respect to sales** ( $d \ln L/d \ln y$ ) To estimate this elasticity we consider the following linear relationship:

$$\ln y_t(\theta_i) = \alpha_0 + \alpha_1 \ln L_t(\theta_i) + \sum_{j=1}^{10} \alpha_{3,j} \text{Div}_j + \varepsilon_{i,t}, \quad (28)$$

where  $\ln y_t(\theta_i)$  is the log of firm sales,  $\ln L_t(\theta_i)$  is the log of firm size, and  $\text{Div}_j$  are industry division dummies.

We look at data from publicly traded US firms in COMPUSTAT. The sample is constructed at an annual frequency from 2000 to 2012.<sup>29</sup> Data on firm sales is taken from *Gross Sales* in the Income Statement, and data on the total number of employees is taken from the *Employees* item. Nominal variables are deflated using the CPI for all urban consumers, all goods. Division dummies are based on Standard Industrial Classification (SIC) as defined by the Occupational Safety & Health Administration.<sup>30</sup>

Our employment data does not distinguish between managerial and non managerial employees. To overcome this limitation we assume that the number of top executives is fixed across firms. Also, we only consider firms above (and including) the median firm size for each year, which minimizes the impact of assuming a fixed number of top executives. To see this, notice that for large firms the relationship between sales and non managerial employees can be approximated as follows:

$$\begin{aligned} \ln(\text{Sales}_{i,t}) &= \alpha_0 + \alpha_1 \ln \left( \text{Employees}_{i,t} - \text{Number of top executives}_{i,t} \right) \\ &= \alpha_0 + \alpha_1 \ln \left( \text{Employees}_{i,t} \right) + \alpha_1 \ln \left( 1 - \frac{\text{Number of top executives}_{i,t}}{\text{Employees}_{i,t}} \right) \\ &\approx \alpha_0 + \alpha_1 \ln \left( \text{Employees}_{i,t} \right). \end{aligned}$$

As a benchmark, we assume that the number of top executives is 20. From (28) we estimate a value of  $\hat{\alpha}_1 = \frac{d \ln y}{d \ln L} = .951$  (0.002), where  $\frac{d \ln y}{d \ln L}$  denotes the average value of  $\frac{d \ln y(\theta)}{d \ln L(\theta)}$  in our sample. The estimated elasticity is consistent with the *making-do-with-less* effect which implies a coefficient smaller than one, as in Lazear et al. (2014).

In Table 1 we report details about our benchmark estimation (Column (1)) along with additional robustness checks. Columns (2) – (4) look at the impact of extending the time period and the effect of either industry or year dummies. Columns (5) – (6) look at the effect of

<sup>29</sup>From our sample we drop firms that report negative or zero sales and firms with duplicate CUSIP.

<sup>30</sup>Division refers to industry groupings. The 10 divisions considered are: Agriculture, Forestry and Fishing; Mining; Construction; Manufacturing; Transportation, Communications, Electric, Gas and Sanitary Services; Wholesale Trade; Retail Trade; Finance, Insurance and Real Estate; Services; Public administration.

changing the decile of firm size included. We observe that our estimate, with either slightly larger or smaller values, is robust to changes in specification. Finally, in Figure 1 we also report our estimates when changing the number of top executives from 1 to 50. The Figure also includes the comparison between our benchmark estimation and the case in which firms below the median size are included.

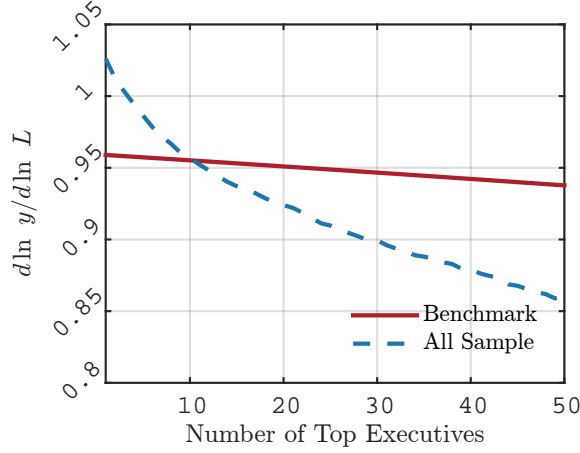


Figure 1: Estimates of  $\frac{d \ln y}{d \ln L}$  by number of top executives. “Benchmark” refers to estimates of (28) using firms above the median size, “All Sample” refers to estimates of (28) also including firms below the median firm size.

**Elasticity of managerial compensation with respect to sales ( $d \ln \pi / d \ln y$ )** Starting from [Roberts \(1956\)](#), there is a vast literature estimating the elasticity of managerial compensation with respect to firm size in the cross-section.<sup>31</sup> This literature has highlighted an empirical regularity, usually denoted as “*Roberts’s Law*,” which states that, on average, managerial compensation is proportional to a power of 1/3 on own firm size. Accordingly, in our benchmark calculation we set  $\frac{d \ln \pi}{d \ln y} = 0.34$ .

**Frisch elasticity ( $\varepsilon$ ), and elasticity of substitution ( $\sigma$ )** We set the value of  $\varepsilon$  based on previous studies. Specifically, following the guidelines of [Chetty et al. \(2011\)](#) we set  $\varepsilon$  equal to 0.5 for our benchmark calibration.<sup>32</sup> To pin down the elasticity of substitution  $\sigma$ , we assume it is constant (as in the case with CES technology) and rely on equilibrium modeling restrictions.<sup>33</sup>

<sup>31</sup>See also [Lewellen and Huntsman \(1970\)](#), [Baker et al. \(1988\)](#), [Gabaix and Landier \(2008\)](#), [Frydman and Saks \(2010\)](#), or [Alder \(2012\)](#).

<sup>32</sup>There exists an extensive literature devoted to the estimation of the elasticity of labor supply. For prime age males [MaCurdy \(1981\)](#) and [Altonji \(1986\)](#) estimate an elasticity between 0 to 0.54. [Saez \(2003\)](#) using the NBER tax panel from 1979 to 1981, estimates a labor elasticity of 0.25. Similar ranges are estimated by [Blundell et al. \(2014\)](#) and [French \(2005\)](#). [Chetty et al. \(2011\)](#) find values equal to 0.5 on the intensive margin and 0.25 on the extensive margin.

<sup>33</sup>There is little direct empirical evidence on how managerial effort substitutes with hired factors of production, such as the number of workers. A possible strategy to identify  $\sigma$  is to use the aggregate behavior of wages over time. For example, [Katz and Murphy \(1992\)](#) use an aggregate production function and estimate an elasticity

Table 1: Estimating the Elasticity of Firm Size With Respect to Sales

	ln(Sales)					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Workers)	0.951 [0.002]	0.933 [0.003]	0.956 [0.001]	0.968 [0.001]	0.912 [.0008]	0.972 [0.004]
Year dummy				Yes	Yes	
Division dummy	Yes		Yes	Yes	Yes	Yes
All time period			Yes	Yes	Yes	
Deciles Included	$\geq 5$	$\geq 5$	$\geq 5$	$\geq 5$	All	$\geq 8$
Observations	50,267	50,267	171,044	171,044	265,764	25,131
$R^2$	0.77	0.71	0.79	0.80	0.84	0.69

Notes: Estimates of  $\alpha_1$  in (28). Column (1) displays benchmark calculation using COMPUSTAT data (2000-2012). "Year dummy" denotes the inclusion or not of yearly dummies. "Division dummy" highlights the inclusion or not in (28) of dummies based on Standard Industrial Classification (SIC) from Occupational Safety & Health Administration. "All time period" denotes the usage of the entire dataset up to 1950. "Decile Included" denote the sample of firms by size included in the estimation of  $\alpha_1$ . We report standard errors in square parentheses.

Given the values for the elasticities  $d \ln \pi / d \ln y$  and  $d \ln y / d \ln L$  estimated previously, equation (27) does not return a positive value of  $\gamma$  for every possible combination of  $\varepsilon$  and  $\sigma$ . Indeed, using  $\varepsilon = 0.5$  we have that  $\gamma > 0$  if and only if  $\sigma > 4.25$ . Based on this threshold, we set  $\sigma = 5$  in the benchmark calibration and perform a robustness check on this parameter in Section 6.<sup>34</sup> Other values of  $\varepsilon$  would yield a different feasible range for  $\sigma$ , as it is shown in Figure 2. Overall, from Figure 2 we see that the feasible range for  $\sigma$  precludes a Cobb-Douglas production function ( $\sigma = 1$ ) for empirically relevant values of  $\varepsilon$ .

We can now determine  $\gamma$  using equation (27) in Proposition 5. We get:

$$\gamma = \frac{1 - \frac{1}{0.951}}{\frac{1}{0.951} - 0.34 \times \frac{0.5+5}{0.5+1}} = 0.30.$$

Next we discuss the estimation of the Pareto parameter  $a$  for the distribution of managerial talent.

### 5.3 ESTIMATING THE TAIL OF THE TALENT DISTRIBUTION $a$

In this subsection we show how to recover the shape of the tail of the talent distribution using the distribution of firm sizes. This approach differs from the standard approach of

of substitution between skilled and unskilled workers of 1.4. Replicating their approach and partitioning the population into workers in managerial and non managerial positions, we arrive at an estimate of the elasticity of substitution between workers and managers of around 4, which is close to the benchmark value in our calibration.

<sup>34</sup>This choice for  $\sigma$  corresponds to a value of  $\rho = \frac{4}{5}$  in the production function of Assumption 2.



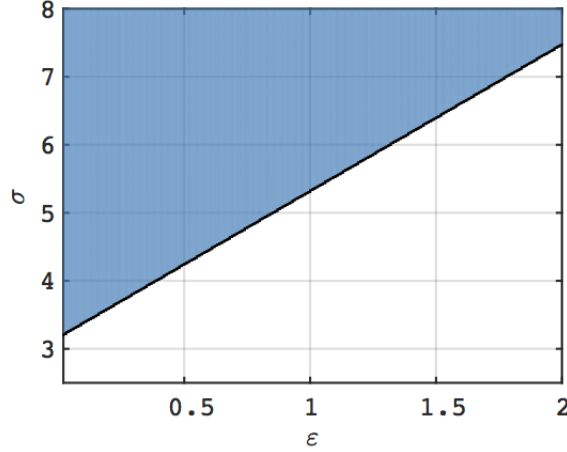


Figure 2: Region for  $(\varepsilon, \sigma)$  consistent with  $\gamma > 0$ .

estimating such a parameter based on the observed distribution of incomes (which is discussed in Appendix E for robustness purposes). The main advantage of our approach is that firm level data is readily available and comprehensive. This is a striking difference with respect to income data which in many instances is survey based and top-coded.

We start by deriving a relationship between the tail of the talent distribution and the tail of the firm size distribution. Given Assumption 3, the maximum likelihood estimate of  $a$  satisfies:<sup>35</sup>

$$\frac{1}{\hat{a}} = \frac{1}{N} \sum_{i=1}^N (\ln(\theta_i) - \ln(\underline{\theta})), \quad (29)$$

where  $\{\theta_1, \dots, \theta_N\}$  is a given a realization of managerial talent, and  $\underline{\theta}$  is the minimum possible value of  $\theta$ . If we let  $a_L$  denote the tail parameter of the Pareto distribution of firm size, the analogue to (29) yields:

$$\frac{1}{\hat{a}_L} = \frac{1}{N} \sum_{i=1}^N (\ln L(\theta_i) - \ln(L(\underline{\theta}))). \quad (30)$$

From equation (24) in Lemma 1 we have that:

$$\ln L(\theta) - \ln L(\underline{\theta}) = \left(1 + \frac{\gamma\sigma}{1 - \kappa(\theta)} + \varepsilon \left(1 + \frac{\gamma}{1 - \kappa(\theta)}\right)\right) (\ln(\theta) - \ln(\underline{\theta})). \quad (31)$$

Finally, combining (29)-(31) and taking limits we can link the tail parameters on the talent and firm size distributions as follows:<sup>36</sup>

**Proposition 6.** *Suppose Assumptions 1 and 4 hold. Then for any solution of the  $\theta$ -manager's*

<sup>35</sup>See, e.g., Malik (1970).

<sup>36</sup>We assume that regularity conditions necessary for consistency of maximum likelihood estimates hold.

problem in (21) we have:

$$a = \left( 1 + \frac{\gamma\sigma}{1 - \bar{\kappa}} + \varepsilon \left( 1 + \frac{\gamma}{1 - \bar{\kappa}} \right) \right) \times a_L. \quad (32)$$

Using equation (32) and the parameters of Section 5.2, the Pareto tail parameter  $a$  can be inferred from the observed value for  $a_L$ , and the share of labor costs at the top  $\bar{\kappa}$ . We pin down these parameters as follows. First, it is well documented that the distribution of firm size exhibits a Pareto distribution with tail parameter close to one. Taking the estimate from Axtell (2001), we set  $a_L = 1.06$ . Second, combining equations (24)-(26) from Lemma 1 we obtain:<sup>37</sup>

$$1 - \kappa(\theta) = \frac{1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}}{\frac{d \ln \pi(\theta)}{d \ln y(\theta)} - \frac{d \ln L(\theta)}{d \ln y(\theta)}}, \quad (33)$$

using the estimated elasticities  $d \ln L / d \ln y$  and  $d \ln \pi / d \ln y$  into (33) we get  $\bar{\kappa} \approx 0.93$ .<sup>38</sup> Plugging in our estimates into equation (32), we obtain

$$a = \left( 1 + \frac{0.30 \times 5}{1 - 0.93} + 0.5 \times \left( 1 + \frac{0.30}{1 - 0.93} \right) \right) \times 1.06 \approx 26.13.$$

The above analysis shows that the distribution of talent is significantly less skewed than the distribution of firm size:  $a$  is an order of magnitude larger than  $a_L$ .<sup>39</sup> Fundamentally, this big difference relies on a positive scale-of-operations effect  $\gamma$ .

To conclude this section, Table 2 summarizes the benchmark parameters and moments used in the calibration. We next compute the value for optimal taxes at the top.

## 6 OPTIMAL TOP INCOME TAXATION IN THE US

Substituting the parameters of the benchmark calibration from Table 2 into our top tax formula (20), we obtain the optimal tax rate implied by our environment:

$$T'(\bar{\theta}) = \frac{1}{1 + a \left[ \frac{1}{\varepsilon} + 1 + \frac{\gamma}{1 - \bar{\kappa}} \left( \frac{1}{\varepsilon} + 1 - \rho \bar{\kappa} \right) \right]^{-1}} = 32.4\%.$$

The prescribed value for top marginal rates relies crucially on the estimated value of  $\gamma$  and its influence on  $a$ . To see this, it is instructive to compare the result with the case in which

<sup>37</sup>See equation (C.10) in Appendix C.2.

<sup>38</sup>This estimate provides an opportunity for a testable implication. In COMPUSTAT we consider the top quintile firms in terms of employment size. For these firms we observe an average work force of 20,000 individuals. This implies that the pay ratio of top executives over an average worker (i.e. the ratio of the compensation of a top executive over that of an average worker in the same firm) is roughly  $\frac{0.08}{20} / \frac{0.93}{20000} \approx 86.2$ . To determine an empirical counterpart we take the CEO-to-worker pay ratio among the 100 highest-grossing publicly traded companies in the United States from Payscale (see <http://www.payscale.com/data-packages/ceo-income/full-list>). The average across these top firms is 87.4, which closely matches what the estimated  $\kappa$  implies.

<sup>39</sup>The same conclusion holds when calibrating  $a$  using the income distribution, as discussed in Appendix E.

Table 2: Benchmark Parameter Calibration.

Parameter	Symbol	Value
Scale-of-operations effect	$\gamma$	0.30
Pareto tail parameter of the talent distribution	$a$	26.13
Frisch elasticity	$\varepsilon$	0.5
Elasticity of substitution	$\sigma$	5
Elasticity of firm size w.r.t. sales	$\frac{d \ln L}{d \ln y}$	1.051
Elasticity of executive compensation w.r.t. sales	$\frac{d \ln \pi}{d \ln y}$	0.34
Share of labor costs to total sales at the top	$\bar{\kappa}$	0.93

Notes:  $\varepsilon$  and  $\sigma$  are imposed exogenously. Firm elasticities are estimated from the data as described in Section 5.2.  $\gamma$ ,  $a$ , and  $\kappa$  are computed using (27), (32), and (33), respectively.

the scale-of-operations effects is shut down: By imposing  $\gamma = 0$ , we obtain  $a_{\gamma=0} = 1.59$  and the corresponding tax rate is almost double our benchmark at 65.4 percent. (For comparison, [Diamond and Saez \(2011\)](#) use a value of  $a = 1.5$ .)

We next proceed to study the effect of the elasticity of labor supply ( $\varepsilon$ ) and the degree of substitutability across inputs ( $\sigma$ ) on the optimal top marginal tax rates. These are reported in Table 3. As expected, we observe that tax rates are decreasing in  $\varepsilon$ . To evaluate the impact of  $\sigma$ , we use two values. The first one ( $\sigma = 4.5$ ) is close to the lower bound consistent with  $\gamma > 0$  in the benchmark calibration. The second value ( $\sigma = 10$ ) is an arbitrarily large number. We see that marginal tax rates are increasing in  $\sigma$ , but optimal tax rates are relatively insensitive to variations in this parameter. It is worth noting that whenever we change  $\varepsilon$  or  $\gamma$ , we recompute  $\gamma$  and  $a$  using (27) and (32), respectively.

The table also shows marginal taxes when imposing  $\gamma = 0$ . Compared to our benchmark results, marginal taxes are higher, but the difference between the two (in relative terms) is decreasing as labor supply becomes more inelastic. Figure 3 displays the previous observations for wider ranges of  $\varepsilon$  and  $\sigma$ .

To conclude, it is useful to compare the baseline optimal tax rate of 32.4 percent with what we see in US data. We look at the March edition of the CPS from 2000 to 2010. For every individual in our sample we compute federal and state taxes of labor income using the NBER TAXSIM calculator.<sup>40</sup> For the top 99<sup>th</sup> percentile we find an effective marginal federal tax rate of 33.5 percent and a marginal state tax rate of 5 percent. [Saez et al. \(2012\)](#) report a marginal rate of approximately 50 percent, for the top 1 percent of workers post 2000. With these values in mind, our benchmark prescribes a tax rate in line to what we currently observe in US data.

<sup>40</sup>We drop individuals with negative income and labor income below \$100. Also dropped are individuals for which labor income is less than 60% of total income or more than 120% of total income. Tax rates are computed using the NBER TAXSIM calculator version 9.2. Rates reported are applied to the head of household inclusive of transfer received. Refer to [Ales et al. \(2015\)](#) for further details.

Table 3: Top Marginal Tax Rates.

	$\varepsilon = 0.25$		$\varepsilon = 0.50$	
$\sigma$	4.5	10	4.5	10
$T'(\bar{\theta})$	51.5%	53.2%	31.7%	34.9%
$\gamma$	0.24	0.03	0.87	0.04
$a$	18.1	5.9	65.5	7.8
$T'_{\gamma=0}(\bar{\theta})$	79.1%		65.3%	
$a_{\gamma=0}$	1.3		1.6	

Notes:  $T'(\bar{\theta})$  denotes optimal tax rate as imputed by (20). The values for  $\gamma$  and  $a$  are computed using (27) and (32), respectively.  $T'_{\gamma=0}(\bar{\theta})$  denotes the optimal tax rates with the exogenous constraint of  $\gamma = 0$ .

## 6.1 CONNECTIONS TO PREVIOUS STUDIES

**Relationship with Saez (2001)** In Section 4 we showed how the classical (DS) tax formula is recovered in the case with  $\gamma = 0$ . That formula was expressed in terms of primitives of the environment, namely the distribution of managerial talent and the structural parameters of the production function. We now relate our formula to an alternative expression in Saez (2001) which utilizes the income distribution, instead of the distribution of skills.

In our environment, for a given value of  $a$  we can recover the tail parameter for the distribution of income,  $a_\pi$ , by applying the following identity (see Appendix E for a derivation):

$$a = \left(1 + \frac{\gamma}{1 - \hat{\kappa}}\right) (1 + \varepsilon) a_\pi, \quad (34)$$

substituting in (20) we get:

$$T'(\bar{\theta}) = \frac{1}{1 + a_\pi \left(1 + \frac{\gamma}{1 - \hat{\kappa}}\right) \varepsilon \left[1 + \frac{\gamma}{1 - \hat{\kappa}} \left(1 - \frac{\rho \kappa \varepsilon}{1 + \varepsilon}\right)\right]^{-1}}. \quad (35)$$

It is then immediate that in the case with no scale-of-operations effect ( $\gamma = 0$ ) or in the Cobb-Douglas case ( $\rho = 0$ ) then we get  $T' = 1/(1 + a_\pi \varepsilon)$  as in Saez (2001) and Diamond and Saez (2011).

Equation (35) allows us to understand further the novel forces present in this environment. As discussed in Diamond and Saez (2011) marginal taxes at the top can be understood by looking at the tail parameter of the distribution of talent ( $a$ ) and the income elasticity of the after tax rate  $e \equiv \frac{\partial \log \pi(\theta)}{\partial \log(1 - \tau)}$ . In equation (34) we see how the presence of the scale-of-operations effect creates a wedge between the distribution of income and the distribution of managerial

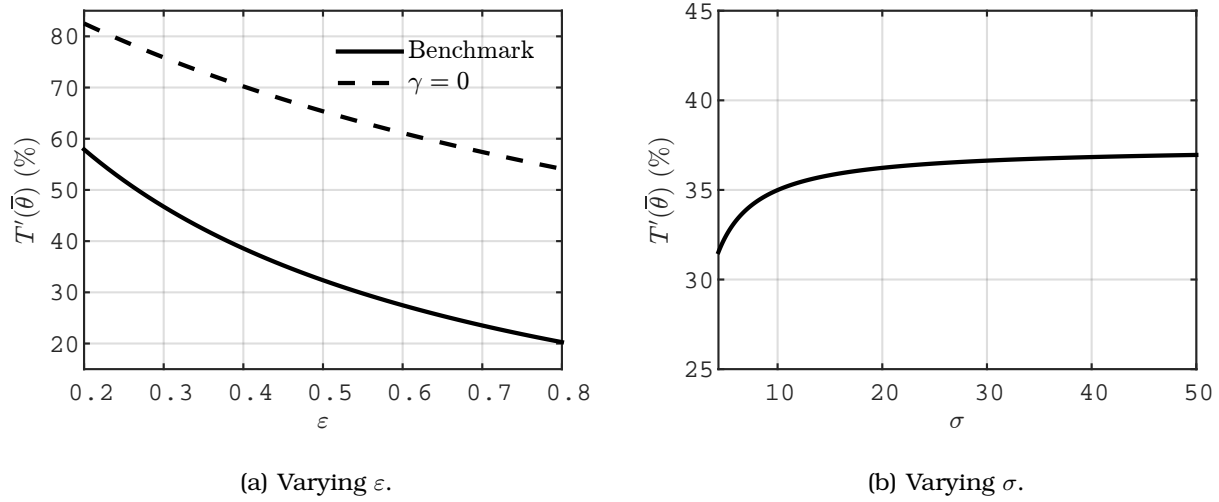


Figure 3: Comparative Statics of Optimal Top Tax Rates.

talent. This wedge is a force for lower taxes (since it points towards a higher value of  $a$ ). At the same time, firm level distortions that emerge with the scale-of-operations effect (as long as  $\rho \neq 0$ ) generate a lower response of income to taxes, which is a force for higher marginal tax rates. In the two cases discussed above, these two forces cancel each other perfectly. However, this is not true in general as our benchmark environment has demonstrated.

**Remark 1.** *The standard approach in public finance to the issue of optimal top income taxation relies heavily on estimating the elasticity of taxable income in reduced form (see, for example Saez et al. (2012) and Piketty et al. (2014)). Our approach, on the other hand, recognises that this elasticity is endogenous to policy, and that firm level distortions may have a particularly strong effect. Indeed, as we show in Appendix D, absent firm size distortions the value of the elasticity of taxable income is given by the Frisch elasticity of labor supply. This would be a case in which the value of “ $e$ ” generated by the model would be consistent with the one estimated in the data.*

**Relationship with Stiglitz (1982)** Stiglitz (1982) shows that the presence of endogenous wages provides a rationale for lowering top income tax rates. The logic is the following. Suppose that labor supplies of individuals with different skills are complements into the production function. Under this scenario, reducing marginal taxes on individuals with high ability induces them to increase their labor supply and, thus, increases the productivity of individuals with lower ability. The latter reduces wage disparities across skills, which relaxes incentive constraints. This insight has been recently extended to richer assignment models (see, e.g. Rothschild and Scheuer (2013)).

While our quantitative results also imply lower top tax rates than in environments with exogenous wages, it is important to note that we abstract from Stiglitz-type effects both *within*

and across occupations. First of all, effort decisions of any given manager have no impact on other managers' wages. The reason is that managers of different abilities are perfect substitutes into the aggregate production function (which is the sum of firms' outputs). Second, while an increment in managerial effort does increase the productivity of the workers, the resulting reduction in wage dispersion across occupations does not relax incentive constraints of the manager. This is because occupations in our model are perfectly observable. A crucial assumption in [Stiglitz \(1982\)](#) is that the government uses a single tax schedule on different types of labor. Our environment, on the other hand, features differential tax treatment of managers and workers.<sup>41,42</sup>

## 7 CROSS-COUNTRY RESULTS

An important contribution of this paper is to make operational the study of optimal top income tax rates with firm-level data. This is appealing since, in most countries, this data is publicly available given the regulatory requirements on publicly traded firms. For example, COMPUSTAT GLOBAL contains detailed fundamental data for major companies trading on international exchanges dating back to 1987. ORBIS, the multi-country database published by Bureau van Dijk, covers information including sales on more than 50 million companies world-wide. Compensation data for individual top executives in fourteen countries with mandated disclosure rules and many other countries can be obtained from BoardEx, which is compiled by the UK-based firm Management Diagnostics Limited.

Next we calculate optimal top income tax for a panel of eleven non-US countries with mandated disclosure of executive compensation, including nine European countries (Belgium, France, Germany, Ireland, Italy, Netherlands, Sweden, Switzerland, UK), Australia, and Canada.<sup>43</sup> For this purpose, we take the estimates of firm sales Pareto tail parameters from [di Giovanni and Levchenko \(2013\)](#) who estimate the country-by-country power laws using firm-level sales from ORBIS. The firm size Pareto tail parameter is then obtained by  $a_L = \frac{d \ln y}{d \ln L} a_y$ . For the CEO pay data, we restrict our sample to those that have mandated disclosure of executive compensation. In particular, the estimates of CEO pay-sales elasticities  $\frac{d \ln \pi}{d \ln y}$  are taken from [Fernandes et al. \(2013\)](#). The sales-size elasticity  $\frac{d \ln y}{d \ln L}$  is estimated using the same methodology in Section 5.2 based on COMPUSTAT NORTH AMERICA for Canada and COMPUSTAT GLOBAL for the other non-US countries.

In Table 4, we present the estimates for  $a_L$ ,  $\frac{d \ln \pi}{d \ln y}$ , and  $\frac{d \ln y}{d \ln L}$ , along with the implied values for  $\gamma$ ,  $a$ , and optimal top tax rates with and without the scale-of-operations effect. The results for

<sup>41</sup>[Scheuer \(2014\)](#) analyzes both cases in an entrepreneurial model with endogenous entry.

<sup>42</sup>Similarly as before, there are no Stiglitz-type forces affecting the taxation on firm sizes.

<sup>43</sup>We thank Miguel Ferreira for sharing the estimates of CEO pay-sales elasticities in the thirteen non-US countries with mandated disclosure of executive compensation. Norway and South Africa are excluded because Norway has low BoardEx coverage resulting a negative Roberts' law estimate, and that South Africa has low ORBIS coverage.



Table 4: Top Tax Rates Across Countries

Country	$a_L$	$\frac{d \ln \pi}{d \ln y}$	$\frac{d \ln y}{d \ln L}$	$\gamma$	$a$	$T'(\bar{\theta})$	$T'_{\gamma=0}(\bar{\theta})$
Australia	0.97	0.41	0.98	0.04	8.34	0.41	0.67
Belgium	0.89	0.37	0.92	0.34	14.84	0.39	0.69
Canada	0.80	0.38	0.90	0.38	12.61	0.42	0.71
France	0.79	0.42	0.89	0.31	8.88	0.45	0.72
Germany	0.79	0.39	0.93	0.21	9.39	0.44	0.72
Ireland	0.75	0.38	0.99	0.02	7.64	0.46	0.73
Italy	0.82	0.52	0.80	0.38	6.22	0.48	0.71
Netherland	0.86	0.30	0.95	1.00	74.50	0.34	0.70
Sweden	0.82	0.31	0.92	1.26	53.94	0.36	0.71
Switzerland	0.77	0.60	0.97	0.02	2.69	0.59	0.72
UK	0.97	0.42	0.96	0.09	8.08	0.42	0.67
US	1.06	0.33	0.95	0.30	25.75	0.32	0.65

Notes: For all the non-US countries, we obtain  $a_L$ ,  $\frac{d \ln \pi}{d \ln y}$ , and  $\frac{d \ln y}{d \ln L}$  according to the following procedure. The firm size Pareto tail parameter  $a_L$  is estimated from  $a_L = \frac{d \ln y}{d \ln L} a_y$ . The estimates of firm sales Pareto tail parameter  $a_y$  are taken from [di Giovanni and Levchenko \(2013\)](#) based on ORBIS database. The estimates of CEO pay-sales elasticities  $\frac{d \ln \pi}{d \ln y}$  are taken from [Fernandes et al. \(2013\)](#). The sales-size elasticity  $\frac{d \ln y}{d \ln L}$  is estimated using the same methodology described in Section 5.2 for 2000-2012. The data for Canada is from COMPUSTAT NORTH AMERICA, whereas those for all the other non-US countries are from COMPUSTAT GLOBAL.

the US are displayed at the bottom of the table for comparison. Figure 4 displays the scatter plots by country of the log optimal top income tax rates versus  $\log(a + 1)$  and  $\log(\gamma + 1)$ .

Our findings on the optimal top income tax rates are strikingly robust across countries. For all the non-US countries, the estimated scale-of-operations effect  $\gamma$  is positive, and the estimated distribution of talent is significantly less skewed than firm size. Consequently, the optimal tax rates are lower once we consider the scale-of-operations effect relative to the case without it (compare  $T'(\bar{\theta})$ , vs.  $T'_{\gamma=0}(\bar{\theta})$ ). Most of the differences in top tax rates across countries is explained by variations in the firm size tail index (Column 1) and in the pay-size elasticity (Column 2). The U.S., the Netherlands, and Sweden, have high values of  $\gamma$  and  $a$  and hence are at the bottom of the marginal tax rate. In comparison, Switzerland and Italy have low values of  $a$ , thereby standing at the top of the spectrum. Comparing the patterns in scatter plots 4(a) and 4(b), model implied optimal top tax rates varies more significantly with estimated  $a$  than  $\gamma$ . This confirms that the effect from a flatter managerial talent distribution dominates in lowering top tax rates.

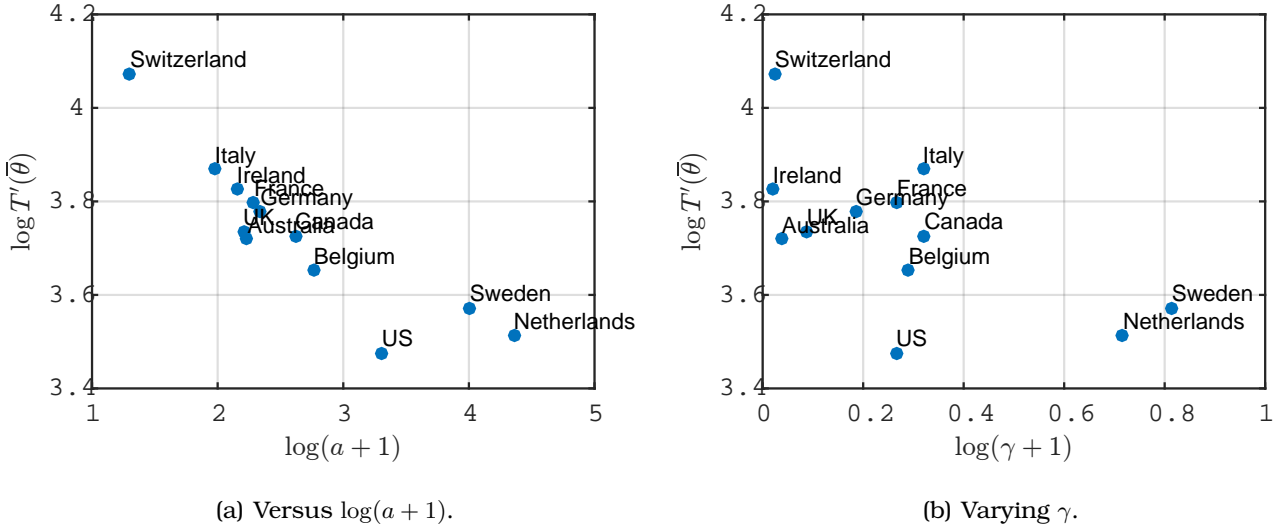


Figure 4: Log Optimal Top Tax Rates by Country.

## 8 OPTIMAL FIRM SIZE TAXATION

As emphasized in Section 4.1, the marginal product of labor at the optimum is typically not equalized across firms. This feature is necessary for incentive provision and, in our decentralization, translates into nonzero marginal taxes on firm size (see equation (14)). In this section we characterize optimal taxes on firm size in a calibrated example. Unlike in previous sections where we focused on taxation at the top, here we compute firm size taxes over the entire talent distribution in order to study progressivity.

We focus on US data and take the values for the parameters  $a$ ,  $\varepsilon$ ,  $\sigma$ , and  $\gamma$  from Table 2. In addition, here we assume that the talent distribution is Pareto-Lognormal<sup>44</sup> with  $\theta \sim \text{PlN}(\zeta, \iota^2, a)$ , and following Mankiw et al. (2009) we set  $\zeta = 2.76$  and  $\iota = 0.56$ . To calibrate  $\beta$  (the share parameter on the production function), we use the definition of  $\kappa$  which implies:

$$\kappa(\theta) = \frac{MP_{L(\theta)}L(\theta)}{y(\theta)} = \frac{1}{1 + \frac{\beta}{1-\beta} \left(\frac{n\theta}{L}\right)^\rho}.$$

From the NBER-CES Manufacturing Industry Database, the average team size defined by the number of production workers per non-production worker is estimated to be 3.58. Using this number to proxy  $\left(\frac{n\theta}{L}\right)^{-1}$ , the equation above and our benchmark value of  $\kappa \approx 0.93$  together imply  $\beta \approx 0.18$ . Finally, the social welfare function (refer to (PO)) is  $\Psi(U) = U^{1/2}$ .

Table 5 reports optimal marginal firm size taxes across firm size percentiles. Two facts stand out. First, computed firm distortions are positive and economically significant. For the median firm, for example, these distortions raise the marginal labor cost by 1.75 percent above

<sup>44</sup>That is,  $\theta \stackrel{d}{=} \theta_1 \theta_2$ , where  $\theta_1 \sim \text{LN}(\zeta, \iota^2)$  and  $\theta_2 \sim P(a)$ .

Table 5: Optimal Marginal Tax on Firm Size.

	Firm Size Percentile				
	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>	99 <sup>th</sup>
$T'_L(wL)$	1.54%	1.75%	1.91%	2.01%	2.02%

the wage rate. Second, the marginal tax on labor use is progressive in the range of firm sizes reported. In particular,  $T'_L$  increases by 30 percent between the 25<sup>th</sup> percentile and the top, where it asymptotes at around 2 percent.

## 9 CONCLUSION

The title of this paper is a reference to the thought provoking novel of [Rand \(1957\)](#).<sup>45</sup> In this dystopian novel, top income earners have a vital role in the workings of the economy and threaten to stop the “world’s motor” in response to increasing government regulation. In this paper we aim to quantify the scope of such types of actions within an optimal taxation framework. We model top income earners as managers whose effort, talent, and hired labor jointly contribute to generate output. We quantify the forces shaping optimal tax rates by estimating the distribution of talent at the top, and the impact of managers on the firm’s overall productivity (here referred to as the scale-of-operation effect).

This paper makes three main contributions. First, we find that top tax rates should be substantially lower than what previous recommendations ignoring the scale-of-operation effect have found. The second contribution is methodological, as we show how to exploit firm level data (as opposed to surveys or censuses eliciting workers’ income) to pin down key parameters relevant for income taxation. Lastly, we provide normative grounds for implementing firm size taxes.

The logical next step involves taking a closer look at the role of managers within the production process. Two extensions come to mind. The first one is to analyze optimal taxes in richer hierarchical organizations rather than in two-rank firms. A second extension is to study how has the contribution of managerial talent evolved over time, as well as its consequences for tax design.

## REFERENCES

- Adams, R. B., H. Almeida, and D. Ferreira (2005). Powerful CEOs and Their Impact on Corporate Performance. *Review of Financial Studies* 18(4), 1403–1432.
- Albanesi, S. (2011). Optimal Taxation of Entrepreneurial Capital with Private Information. Working paper.

<sup>45</sup>The excellent survey of [Slemrod \(2000\)](#) also features a similar title.

- Alder, S. (2012). In the Wrong Hands: Complementarities, Resource Allocation, and Aggregate TFP. Working paper.
- Ales, L., M. Kurnaz, and C. Sleet (2015). Technical change, wage inequality, and taxes. *American Economic Review* 105(10), 3061–3101.
- Ales, L. and C. Sleet (2015). Taxing Top Incomes. Working paper.
- Altonji, J. G. (1986). Intertemporal Substitution in Labor Supply: Evidence from Micro Data. *Journal of Political Economy* 94(3), 176–215.
- Axtell, R. L. (2001). Zipf Distribution of US Firm Sizes. *Science* 293(5536), 1818–1820.
- Baker, G. P. and B. J. Hall (2004). Ceo Incentives and Firm Size. *Journal of Labor Economics* 22(4), 767–798.
- Baker, G. P., M. C. Jensen, and K. J. Murphy (1988). Compensation and Incentives: Practice vs. Theory. *The Journal of Finance* 43(3), 593–616.
- Bakija, J., A. Cole, and B. Heim (2012). Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data. Working paper.
- Baranchuk, N., G. MacDonald, and J. Yang (2011). The Economics of Super Managers. *Review of Financial Studies*.
- Bartelsman, E. J. and M. Doms (2000). Understanding Productivity: Lessons from Longitudinal Micro-data. *Journal of Economic Literature* 38(3), 569–594.
- Bennedsen, M., K. M. Nielsen, F. Perez-Gonzalez, and D. Wolfenzon (2007). Inside the Family Firm: The Role of Families in Succession Decisions and Performance. *The Quarterly Journal of Economics* 122(2), 647–691.
- Bertrand, M. and A. Schoar (2003). Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics* 118(4), 1169–1208.
- Blundell, R. and T. MaCurdy (1999). Labor Supply: A Review and Alternative Approaches. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 27, pp. 1559–1695. Elsevier.
- Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2014). Consumption Inequality and Family Labor Supply. Working paper.
- Cagetti, M. and M. D. Nardi (2006). Entrepreneurship, Frictions, and Wealth. *Journal of Political Economy* 114(5), 835–870.
- Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins. *The American Economic Review* 101(3), 471–475.
- Custódio, C., M. A. Ferreira, and P. Matos (2013). Generalists versus Specialists: Lifetime Work Experience and Chief Executive Officer Pay. *Journal of Financial Economics* 108(2), 471 – 492.
- Custódio, C. and D. Metzger (2013). How Do CEOs Matter? The Effect of Industry Expertise on Acquisition Returns. *Review of Financial Studies*.
- di Giovanni, J. and A. Levchenko (2013). Firm Entry, Trade, and Welfare in Zipf’s World. *Journal of International Economics* 89(2), 283–296.
- Diamond, P. (1998). Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review* 88(11), 83–95.

- Diamond, P. A. and J. A. Mirrlees (1971). Optimal Taxation and Public Production I: Production Efficiency. *American Economic Review* 61(1), 8–27.
- Diamond, P. A. and E. Saez (2011). The Case for a Progressive Tax: From Basic Research to Policy Recommendations. *Journal of Economic Perspectives* 25(4), 165–90.
- Edmans, A. and X. Gabaix (2011). The Effect of Risk on the CEO Market. *Review of Financial Studies* 24(8), 2822–2863.
- Edmans, A., X. Gabaix, and A. Landier (2009). A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium. *Review of Financial Studies* 22(12), 4881–4917.
- Eisfeldt, A. L. and C. M. Kuhnen (2013). CEO Turnover in a Competitive Assignment Framework. *Journal of Financial Economics* 109(2), 351 – 372.
- Farhi, E. and I. Werning (2013). Insurance and Taxation over the Life Cycle. *The Review of Economic Studies* 80(2), 596–635.
- Fernandes, N., M. A. Ferreira, P. Matos, and K. J. Murphy (2013). Are U.S. CEOs Paid More? New International Evidence. *Review of Financial Studies* 26(2), 323–367.
- French, E. (2005). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour. *Review of Economic Studies* 72(2), 395–427.
- Frydman, C. and R. E. Saks (2010). Executive Compensation: A New View from a Long-Term Perspective, 1936–2005. *Review of Financial Studies* 23(5), 2099–2138.
- Gabaix, X. and A. Landier (2008). Why Has CEO Pay Increased So Much? *The Quarterly Journal of Economics* 123(1), 49–100.
- Golosov, M., M. Troshkin, and A. Tsyvinski (2015). Redistribution and Social Insurance. *American Economic Review*, forthcoming.
- Gouveia, M. and R. Strauss (1994). Effective Individual Income Tax Functions: An Exploratory Empirical Analysis. *National Tax Journal* 47(2), 317–39.
- Guesnerie, R. (1981). On Taxation and Incentives: Further Reflections on the Limits to Redistribution. Working paper.
- Huggett, M. and A. Badel (2015). Taxing Top Earners: A Human Capital Perspective. Working paper.
- Kapička, M. (2013). Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach. *The Review of Economic Studies* 80(3), 1027–1054.
- Kaplan, S. N., M. M. Klebanov, and M. Sorensen (2012). Which CEO Characteristics and Abilities Matter? *The Journal of Finance* 67(3), 973–1007.
- Kaplan, S. N. and J. Rauh (2013). It's the Market: The Broad-Based Rise in the Return to Top Talent. *Journal of Economic Perspectives* 27(3), 35–56.
- Katz, L. F. and K. M. Murphy (1992). Changes in Relative Wages, 1963–1987: Supply and Demand Factors. *The Quarterly Journal of Economics* 107(1), 35–78.
- Lazear, E. P., K. L. Shaw, and C. Stanton (2014). Making Do With Less: Working Harder During Recessions. Working paper.
- Lewellen, W. G. and B. Huntsman (1970). Managerial Pay and Corporate Performance. *The American Economic Review* 60(4), 710–720.
- Lucas, R. E. (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics* 9(2), 508–523.

- MaCurdy, T. E. (1981). An Empirical Model of Labor Supply in a Life-Cycle Setting. *The Journal of Political Economy* 89(6), 1059–1085.
- Malik, H. J. (1970). Estimation of the Parameters of the Pareto Distribution. *Metrika* 15(1), 126–132.
- Mankiw, G. (2013). Defending the One Percent. *Journal of Economic Perspectives* 27(3), 21–34.
- Mankiw, G., M. Weinzierl, and D. Yagan (2009). Optimal Taxation in Theory and Practice. *Journal of Economic Perspectives* 23(4), 147–174.
- Mayer, T. (1960). The Distributions of Ability and Earnings. *The Review of Economics and Statistics* 42(2), 189–195.
- Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies* 38(2), 175–208.
- Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics* 118(1), 1–39.
- Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy* 6(1), 230–71.
- Quadrini, V. (2000). Entrepreneurship, Saving and Social Mobility. *Review of Economic Dynamics* 3(1), 1–40.
- Rand, A. (1957). *Atlas Shrugged*. New York: Signet.
- Roberts, D. R. (1956). A General Theory of Executive Compensation Based on Statistically Tested Propositions. *The Quarterly Journal of Economics* 70(2), 270–294.
- Rosen, S. (1982). Authority, Control, and the Distribution of Earnings. *The Bell Journal of Economics* 13(2), 311–323.
- Rothschild, C. and F. Scheuer (2013). Redistributive Taxation in the Roy Model. *The Quarterly Journal of Economics* 128(2), 623–668.
- Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies* 68(1), 205–229.
- Saez, E. (2003). The Effect of Marginal Tax Rates on Income: A Panel Study of ‘Bracket Creep’. *Journal of Public Economics* 87(5–6), 1231–1258.
- Saez, E., J. Slemrod, and S. H. Giertz (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature* 50(1), 3–50.
- Saez, E. and S. Stantcheva (2015). Generalized social welfare weights for optimal tax theory. *NBER Working Paper [forthcoming at the American Economic Review]* 18835.
- Scheuer, F. (2014). Entrepreneurial Taxation with Endogenous Entry. *American Economic Journal: Economic Policy* 6(2), 126–163.
- Scheuer, F. and I. Werning (2015). The Taxation of Superstars. Working paper.
- Shourideh, A. (2012). Optimal Taxation of Entrepreneurial Income: A Mirrleesian Approach to Capital Accumulation. Working paper.
- Slavík, C. and H. Yazıcı (2014). Machines, Buildings and Optimal Dynamic Taxes. *Journal of Monetary Economics* 66(0), 47–61.
- Slemrod, J. (2000). *Does Atlas Shrug?: The Economic Consequences of Taxing the Rich*. Harvard University Press.



- Stiglitz, J. E. (1982). Self-Selection and Pareto Efficient Taxation. *Journal of Public Economics* 17(2), 213–240.
- Terviö, M. (2008). The Difference that CEOs Make: An Assignment Model Approach. *American Economic Review* 98(3), 642–668.
- Weinzierl, M. (2014). The promise of positive optimal taxation: normative diversity and a role for equal sacrifice. *Journal of Public Economics* 118, 128 – 142.

## APPENDIX

### A PROOF OF PROPOSITION 1

Define  $M(\theta', \theta) \equiv c(\theta') - v(n(y(\theta'), L(\theta'), \theta))$ . Incentive compatibility (4) requires that for all  $\theta \in \Theta$ ,  $M(\theta', \theta)$  attains a global maximum at  $\theta' = \theta$ . We start by characterizing local maxima of  $M(\theta', \theta)$  at  $\theta' = \theta$  using the following lemma.

**Lemma 2.** *Let  $M(\theta', \theta) \equiv c(\theta') - v(n(\theta', \theta))$  where  $n(\theta', \theta) \equiv n(y(\theta'), L(\theta'), \theta)$ . A local maximum of  $M(\theta', \theta)$  at  $\theta' = \theta$  is attained if and only if for all  $\theta \in \Theta$ :*

$$c'(\theta) - v'(n(\theta)) [n_y(\theta)y'(\theta) + n_L(\theta)L'(\theta)] = 0, \quad (\text{A.1})$$

and

$$\begin{aligned} & y'(\theta) [v''(n(\theta))n_y(\theta)n_\theta(\theta) + v'(n(\theta))n_{\theta y}(\theta)] + \\ & + L'(\theta) [v''(n(\theta))n_L(\theta)n_\theta(\theta) + v'(n(\theta))n_{\theta L}(\theta)] \leq 0, \end{aligned} \quad (\text{A.2})$$

where  $n(\theta, \theta) = n(\theta)$  and  $n_i(\theta, \theta) = n_i(\theta)$  for  $i = y, L, \theta, y\theta, L\theta$ .

*Proof.* The first order condition for  $\theta' = \theta$  to be a local maximum of  $M(\theta', \theta)$  is  $M_1(\theta, \theta) = 0$ .<sup>46</sup> This is equivalent to (A.1). Differentiating the first order condition  $M_1(\theta, \theta) = 0$  with respect to  $\theta$  gives  $M_{11}(\theta) + M_{12}(\theta) = 0$ . Hence, the second order condition  $M_{11}(\theta) \leq 0$  can be written as  $-M_{12}(\theta) \leq 0$ , which gives (A.2).  $\square$

We now go back to the proof of Proposition 1, which shows that  $M(\theta', \theta)$  attains a global maximum at  $\theta' = \theta$  when (A.1) and (A.2) hold. The proof follows standard arguments.

*Proof.* We want to show that  $M_1(\theta', \theta)$  has the sign of  $(\theta - \theta')$ . First note that

$$M_1(\theta', \theta) = c'(\theta') - v'(n(\theta', \theta)) [n_y(\theta', \theta)y'(\theta') + n_L(\theta', \theta)L'(\theta')]. \quad (\text{A.3})$$

We also have that (A.1) evaluated at  $\theta'$  gives

$$c'(\theta') = v'(n(\theta')) [n_y(\theta')y'(\theta') + n_L(\theta')L'(\theta')]. \quad (\text{A.4})$$

Using (A.4) into (A.3) gives

$$M_1(\theta', \theta) = J(\theta', \theta') - J(\theta', \theta), \quad (\text{A.5})$$

where  $J(\theta', \theta) \equiv v'(n(\theta', \theta)) [n_y(\theta', \theta)y'(\theta') + n_L(\theta', \theta)L'(\theta')]$ . Differentiating with respect to the second argument:

$$\begin{aligned} J_2(\theta', \theta) &= y'(\theta') [v''(n(\theta'))n_y(\theta')n_\theta(\theta') + v'(n(\theta'))n_{\theta y}(\theta')] + \\ & L'(\theta') [v''(n(\theta'))n_L(\theta')n_\theta(\theta') + v'(n(\theta'))n_{\theta L}(\theta')]; \end{aligned}$$

From (A.2) we have that  $J_2(\theta', \theta') \leq 0$ . Then (A.5) implies that  $M_1(\theta', \theta) \geq 0$  if and only if  $\theta' \leq \theta$ . Finally (6) is obtained by combining (A.1) and (A.2). This completes the proof.  $\square$

<sup>46</sup>The subscript  $i = 1, 2$  denotes derivative with respect to the first or second argument.

## B PROOFS OF SECTION 4

### B.1 PROOF OF PROPOSITION 2

As a first step, we show that there exist a wage  $w$  and a tax system  $\{T, T_L, \phi\}$  such that for a given  $\theta \in \Theta$ , the optimal allocation  $\{c(\theta), y(\theta), L(\theta)\}$  solves the manager's problem in (MP). To that end, define the retention function

$$R(y, L) \equiv \max_c \left\{ c : c(\theta) - v(n(y(\theta), L(\theta), \theta)) \geq c - v(n(y, L, \theta)), \forall \theta \in \Theta \right\}, \quad (\text{B.1})$$

and the budget set

$$B \equiv \left\{ (c, y, L) : c \leq R(y, L) \right\}.$$

Now consider the following claim:

**Claim 1.** *Take a  $\theta \in \Theta$  and let  $\{c(\theta), y(\theta), L(\theta)\}$  be the optimal allocation assigned to the  $\theta$ -manager. Then  $\{c(\theta), y(\theta), L(\theta)\}$  solves the problem:*

$$\{c(\theta), y(\theta), L(\theta)\} \in \arg \max_{\{c, y, L\} \in B} c - v(n(y, L, \theta)). \quad (\text{B.2})$$

*Proof.* To prove this claim, we follow two steps: (1) show that  $\{c(\theta), y(\theta), L(\theta)\} \in B$ , for all  $\theta$ , and (2) show that for each  $\theta$ ,  $\{c(\theta), y(\theta), L(\theta)\}$  solves (B.2). Step (1) follows by contradiction. Specifically, suppose that there exists a  $\hat{\theta}$  such that  $\{c(\hat{\theta}), y(\hat{\theta}), L(\hat{\theta})\} \notin B$ . Then, by construction, it must be that

$$\begin{aligned} c(\hat{\theta}) &> R(y(\hat{\theta}), L(\hat{\theta})) \\ &\geq \max_c \left\{ c : c(\theta') - v(n(y(\theta'), L(\theta'), \theta')) \geq c - v(n(y(\hat{\theta}), L(\hat{\theta}), \theta')) \right\} \end{aligned}$$

for some  $\theta' \in \Theta$ , implying that

$$c(\hat{\theta}) - v(n(y(\hat{\theta}), L(\hat{\theta}), \theta')) \geq c(\theta') - v(n(y(\theta'), L(\theta'), \theta')),$$

which violates incentive compatibility. Given that  $\{c(\theta), y(\theta), L(\theta)\} \in B$ , for all  $\theta$ , Step (2) is immediate by incentive compatibility.  $\square$

Now define taxes  $T$ ,  $T_L$ , and a wage  $w$  such that:

$$y - wL - T_L(wL) - T(y - wL - T_L(wL)) = R(y, L), \quad (\text{B.3})$$

where  $R(y, L)$  is the retention function defined in (B.1).

Clearly, many different tax-wage combinations satisfy the relationship in (B.3).<sup>47</sup> Claim 1 then implies that each of those combinations,  $\{c(\theta), y(\theta), L(\theta)\}$  solves the  $\theta$ -manager's problem in (MP).

To complete the proof of the decentralization, define the transfer  $\phi \equiv c^w - w$ , where  $c^w$  is the consumption of the worker at the optimum. Given this level  $\phi$ , the worker's budget constraint holds with equality at the optimal allocation. The optimum also satisfies market clearing conditions by construction, while the government's budget is balanced by Walras' law.

<sup>47</sup>For this reason, the levels of  $T$ ,  $T_L$  and  $w$  are not determined in the decentralization.

## B.2 PROOF OF PROPOSITION 3

We compute the Pareto optimal allocation by solving the optimal control problem (PO-FOC) where  $y(\theta)$  and  $L(\theta)$  are the controls and  $U(\theta)$  is the state variable. After integrating by parts, the Lagrangian to the planner's problem is (suppressing dependencies with respect to  $\theta$ ,  $y$  and  $L$ ):

$$\mathcal{L} = \Psi(c^w) + \int \Psi(U) dF - \int [\mu' U - \mu v'(n) n_\theta] d\theta + \lambda^r \int [y - c^w - U - v(n)] dF - \lambda^l \int [L - 1] dF,$$

where  $\lambda^r$  is the multiplier on (7),  $\lambda^l$  is the multiplier on (8) and  $\mu(\theta)$  is the costate on (5) that also satisfies the boundary conditions  $\mu(\underline{\theta}) = \lim_{\theta \rightarrow \bar{\theta}} \mu(\theta) = 0$ . It is straightforward to show that all of these multipliers are positive.

Optimality conditions with respect to the controls  $y$  and  $L$  are, respectively,

$$\lambda^r (1 - v'(n) n_y) f + \mu (v''(n) n_y n_\theta + v' n_{\theta y}) = 0, \quad (\text{B.4})$$

$$-\lambda^l f - \lambda^r v'(n) n_L f + \mu (v''(n) n_L n_\theta + v' n_{\theta L}) = 0 \quad (\text{B.5})$$

and the costate equation is

$$\mu' = (\Psi'(U) - \lambda^r) f. \quad (\text{B.6})$$

Now let  $(y, L, \theta)$  be such that  $n(y, L, \theta) = \bar{n}$ , where  $\bar{n}$  is a given level of effort. By applying the implicit function theorem we have that

$$y_L = -n_L / n_y. \quad (\text{B.7})$$

Rearranging the first order conditions (B.4) and (B.5) we get

$$n_y [\lambda^r v' f - \mu v'' n_\theta] = \lambda^r f + \mu v' n_{\theta y},$$

and

$$n_L [\lambda^r v' f - \mu v'' n_\theta] = -\lambda^l f + \mu v' n_{\theta L},$$

so that

$$y_L = -\frac{n_L}{n_y} = \frac{\lambda^l f - \mu v' n_{\theta L}}{\lambda^r f + \mu v' n_{\theta y}}. \quad (\text{B.8})$$

Using (14), (B.8) and substituting the expression for  $w = \lambda^l / \lambda^r$  we derive equation (15).

To prove Part 2, suppose that  $n_\theta = g(n)$  for all  $\theta$ . It follows that for all  $\theta$

$$\frac{n_{\theta L}}{n_{\theta y}} = \frac{n_L}{n_y}. \quad (\text{B.9})$$

Rearranging (B.8) and substituting (B.9) we get  $-\frac{n_{\theta L}}{n_{\theta y}} \lambda^r = \lambda^l$ . Simplifying and substituting (B.9) we get  $-n_L / n_y = \lambda^l / \lambda^r$ , so that  $T'_L = 0$ .

## B.3 PROOF OF PROPOSITION 4

Integrating (B.6) between  $\theta$  and  $\bar{\theta}$  and using the transversality condition we get

$$\mu(\theta) = \int_{\theta}^{\bar{\theta}} (\lambda^r - \Psi'(U(\theta))) f(\theta) d\theta. \quad (\text{B.10})$$

Evaluating (B.10) at  $\underline{\theta}$  gives

$$\lambda^r = \int_{\Theta} \Psi'(U(\theta)) f(\theta) d\theta. \quad (\text{B.11})$$

Let  $D(\theta) \equiv \frac{1}{1-F(\theta)} \int_{\theta}^{\bar{\theta}} \Psi'(U(\theta)) dF(\theta)$  so that  $D(\underline{\theta}) = \lambda^r$ . Substituting into (B.10) we obtain

$$\mu(\theta) = (1 - F(\theta)) (D(\underline{\theta}) - D(\theta)), \quad (\text{B.12})$$

from (B.4)

$$-\mu n_y v' \left[ \frac{v''}{v'} n_{\theta} + \frac{n_{\theta y}}{n_y} \right] = \lambda^r [1 - v'(n) n_y] f.$$

Substitute (B.12) into the above yields

$$(1 - F(\theta)) (D(\underline{\theta}) - D(\theta)) n_y v' \left[ -\frac{v''}{v'} n_{\theta} - \frac{n_{\theta y}}{n_y} \right] = \lambda^r [1 - v'(n) n_y] f.$$

By Assumption 1 and (B.11) we get

$$\frac{(1 - F(\theta))}{\theta f(\theta)} \left( 1 - \frac{D(\theta)}{D(\underline{\theta})} \right) \left[ -\frac{1}{\varepsilon} \frac{n_{\theta}}{n} \theta - \frac{n_{\theta y}}{n_y} \theta \right] = \frac{[1 - v'(n) n_y]}{v'(n) n_y},$$

so using (16) yields

$$\frac{T'}{1 - T'} = \frac{1 - F(\theta)}{\theta f(\theta)} \left( 1 - \frac{D(\theta)}{D(\underline{\theta})} \right) \left[ -\frac{1}{\varepsilon} \frac{n_{\theta}}{n} \theta - \frac{n_{\theta y}}{n_y} \theta \right]. \quad (\text{B.13})$$

Next we write the partial derivatives of  $n$  in (B.13) in terms of partial derivatives of  $y$ . Let  $(n, L, \theta)$  be such that  $y(n, L, \theta) = \bar{y}$ , where  $\bar{y}$  is a given level of output. The implicit function theorem gives

$$n_L = -\frac{y_L}{y_n}, \quad (\text{B.14})$$

and

$$n_{\theta} = -\frac{y_{\theta}}{y_n}. \quad (\text{B.15})$$

Combining (B.7) and (B.14) gives  $n_y(y, L, \theta) = 1/y_n(n(y, L, \theta), L, \theta)$ . By differentiating both sides with respect to  $\theta$  we have  $n_{\theta y} = -\frac{y_{nn} n_{\theta} + y_{n\theta}}{y_n^2}$ , which implies

$$\begin{aligned} -\frac{n_{\theta y}}{n_y} \theta &= \left( \frac{y_{nn}}{y_n} n \right) \left( \frac{n_{\theta}}{n} \theta \right) + \frac{y_{n\theta}}{y_n} \theta \\ &= \left[ \frac{y_{n\theta}}{y_{\theta}} n - \frac{y_{nn}}{y_n} n \right] \frac{y_{\theta}}{y_n} \frac{\theta}{n} \\ &= \frac{d \ln \left( \frac{y_{\theta}}{y_n} \right)}{d \ln n} \frac{y_{\theta}}{y_n} \frac{\theta}{n}. \end{aligned} \quad (\text{B.16})$$

Substituting (B.15) and (B.16) into (B.13) gives the result.

#### B.4 PROOF OF COROLLARY 1

Denote with  $g(\theta) = \theta^\gamma$ . Given Assumption 2 we have

$$n(y, L, \theta) = \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho - \frac{1-\beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right]^{\frac{1}{\rho}}. \quad (\text{B.17})$$

Taking derivatives from (B.17) we obtain

$$n_\theta(y, L, \theta) = \left( -\frac{1}{\theta} \right) \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho - \frac{1-\beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right]^{\frac{1}{\rho}-1} \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho (1+\gamma) - \frac{1-\beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right],$$

using the above and (B.17) we get

$$\frac{n_\theta(y, L, \theta)}{n(y, L, \theta)} = -\frac{1}{\theta} \frac{\left( \frac{y}{g(\theta)} \right)^\rho (1+\gamma) - (1-\beta)L^\rho}{\left( \frac{y}{g(\theta)} \right)^\rho - (1-\beta)L^\rho}. \quad (\text{B.18})$$

Also

$$n_y(y, L, \theta) = \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho - \frac{1-\beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right]^{\frac{1}{\rho}-1} \frac{1}{\beta y} \left( \frac{y}{\theta g(\theta)} \right)^\rho, \quad (\text{B.19})$$

$$\begin{aligned} n_{\theta y}(y, L, \theta) &= \left( -\frac{1-\rho}{\theta} \right) \frac{1}{\beta y} \left( \frac{y}{\theta g(\theta)} \right)^\rho \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho - \frac{1-\beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right]^{\frac{1}{\rho}-2} \times \\ &\quad \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho (1+\gamma) - \frac{1-\beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right] - \frac{\rho}{\theta} \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho - \frac{1-\beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right]^{\frac{1}{\rho}-1} \frac{1}{\beta y} \left( \frac{y}{\theta g(\theta)} \right)^\rho (1+\gamma). \end{aligned}$$

The two above imply:

$$\frac{n_{\theta y}(y, L, \theta)}{n_y(y, L, \theta)} = -\frac{1}{\theta} \left[ (1-\rho) \frac{\left( \frac{y}{g(\theta)} \right)^\rho (1+\gamma) - (1-\beta)L^\rho}{\left( \frac{y}{g(\theta)} \right)^\rho - (1-\beta)L^\rho} + \rho(1+\gamma) \right]. \quad (\text{B.20})$$

By Assumption 2 we have

$$\left( \frac{y(\theta)}{g(\theta)} \right)^\rho = \beta(\theta n)^\rho + (1-\beta)L^\rho.$$

Then

$$\frac{\left( \frac{y}{g(\theta)} \right)^\rho (1+\gamma) - (1-\beta)L^\rho}{\left( \frac{y}{g(\theta)} \right)^\rho - (1-\beta)L^\rho} = 1 + \gamma \left( 1 + \frac{1-\beta}{\beta} \left( \frac{L}{\theta n} \right)^\rho \right). \quad (\text{B.21})$$

Also, Assumption 2 implies:

$$\frac{\kappa(\theta)}{1-\kappa(\theta)} = \frac{1-\beta}{\beta} \left( \frac{L}{\theta n} \right)^\rho, \quad (\text{B.22})$$

where  $\kappa(\theta) \equiv y_L(\theta)L(\theta)/y(\theta)$  denotes the *share of labor costs to total sales* for manager  $\theta$ .

Using (B.22) in (B.21) gives

$$\frac{\left(\frac{y}{g(\theta)}\right)^\rho (1 + \gamma) - (1 - \beta)L^\rho}{\left(\frac{y}{g(\theta)}\right)^\rho - (1 - \beta)L^\rho} = 1 + \gamma \left(1 + \frac{\kappa(\theta)}{1 - \kappa(\theta)}\right). \quad (\text{B.23})$$

Combining (B.15) with (B.18) and (B.23) gives the expression for  $\frac{y_\theta}{y_n} \frac{\theta}{n}$ :

$$\frac{y_\theta}{y_n} \frac{\theta}{n} = -\frac{n_\theta}{n} \theta = 1 + \frac{\gamma}{1 - \kappa(\theta)}. \quad (\text{B.24})$$

As for the expression for  $\frac{d \ln \left(\frac{y_\theta}{y_n}\right)}{d \ln n}$ , note that (B.16) implies

$$\begin{aligned} \frac{d \ln \left(\frac{y_\theta}{y_n}\right)}{d \ln n} &= -\frac{n_{\theta y} \theta}{n_y} \left(-\frac{n_\theta}{n} \theta\right)^{-1} \\ &= \left[(1 - \rho) \left(1 + \frac{\gamma}{1 - \kappa(\theta)}\right) + \rho(1 + \gamma)\right] \left(1 + \frac{\gamma}{1 - \kappa(\theta)}\right)^{-1} \\ &= \left[1 + \frac{\gamma}{1 - \kappa(\theta)}(1 - \rho\kappa(\theta))\right] \left(1 + \frac{\gamma}{1 - \kappa(\theta)}\right)^{-1}, \end{aligned} \quad (\text{B.25})$$

where the next to last line follows from (B.20), (B.23), and (B.24). This completes the proof.

## C PROOFS OF SECTION 5

### C.1 PROOF OF LEMMA 1

We start by establishing the following result (henceforth we suppress the arguments of all functions):

**Lemma 3.**

$$\frac{1}{\sigma} = -\left(\frac{\theta n}{L}\right) \frac{h''}{h'} \frac{1}{\kappa}. \quad (\text{C.1})$$

*Proof.* From the definition of  $\kappa$  we can write

$$1 - \kappa = \frac{\theta n}{L} \frac{h'}{h}. \quad (\text{C.2})$$

Let  $f(\theta n, L) = Lh(\theta n/L)$ . Define the elasticity of substitution between  $\theta n$  and  $L$  as  $\sigma \equiv -\frac{d \ln(L/\theta n)}{d \ln(f_2/f_1)}$ .

By definition of  $f$  we have  $f_1 = h'$  and  $f_2 = h - \frac{\theta n}{L} h'$  which implies  $\frac{f_2}{f_1} = \frac{h}{h'} - \frac{\theta n}{L}$ . Therefore

$$\frac{1}{\sigma} = -\frac{d \ln(f_2/f_1)}{d \ln(L/\theta n)} = -\left(\frac{h}{h'} - \frac{\theta n}{L}\right)^{-1} \frac{L}{\theta n} \frac{d\left(\frac{h}{h'} - \frac{\theta n}{L}\right)}{d(L/\theta n)}.$$

Differentiating and re-arranging, we get:

$$\frac{1}{\sigma} = -\left(\frac{\theta n}{L}\right) \frac{h''}{h'} \left(1 - \frac{\theta n}{L} \frac{h'}{h}\right)^{-1}.$$

Substituting (C.2) in the above we obtain the result.  $\square$



We now move to the proof of Lemma 1. As notation, let  $g = \theta^\gamma$  and  $g' = \gamma\theta^{\gamma-1}$ .

*Proof.* Differentiating (22) and (23) we get

$$\frac{d \ln L}{d \ln \theta} - \frac{d \ln n}{d \ln \theta} = 1 - \theta \frac{g'}{g} \left( \frac{L}{\theta n} \right)^2 \frac{1}{h''} \left( h - \frac{\theta n}{L} h' \right), \quad (\text{C.3})$$

$$\frac{n\theta}{L} g h'' \frac{d \ln L}{d \ln \theta} + \left[ \frac{n}{\theta} \frac{v''}{(1-\tau)\chi} - \frac{\theta n}{L} g h'' \right] \frac{d \ln n}{d \ln \theta} = g' \theta h' + g h' + g h'' \frac{n\theta}{L}. \quad (\text{C.4})$$

Combining (C.3) and (C.4),

$$\frac{n}{\theta} \frac{v''}{(1-\tau)\chi} \frac{d \ln n}{d \ln \theta} + \frac{n\theta}{L} g h'' \left( 1 + \frac{\gamma\sigma}{1-\kappa} \right) = g' \theta h' + g h' + g h'' \frac{n\theta}{L},$$

where we applied (C.1), (C.2) and the definition of  $\gamma$  on the last term.

Further rearranging (C.1) and (C.2) gives

$$\frac{v''}{(1-\tau)\chi} \frac{n}{\theta} \frac{d \ln n}{d \ln \theta} = g h' \left( 1 + \frac{\gamma}{1-\kappa} \right). \quad (\text{C.5})$$

From the first order condition (23) we have

$$\frac{d \ln n}{d \ln \theta} = \varepsilon \left( 1 + \frac{\gamma}{1-\kappa} \right), \quad (\text{C.6})$$

where we used that  $\frac{v'}{v''n} = \varepsilon$ .

Plugging in (C.6) into (C.3),

$$\frac{d \ln L}{d \ln \theta} = \varepsilon \left( 1 + \frac{\gamma}{1-\kappa} \right) + 1 - \gamma \left( \frac{L}{\theta n} \right)^2 \frac{1}{h''} \left( h - \frac{\theta n}{L} h' \right).$$

Then applying (C.1) and (C.2) and rearranging gives (24).

Now we obtain equation (25). By constant returns to scale we can write  $y = gLh$ , so that

$$\begin{aligned} \frac{d \ln y}{d \ln \theta} &= \frac{d \ln g(\theta)}{d \ln \theta} + \frac{d \ln L}{d \ln \theta} + \frac{d \ln h}{d \ln \theta} \\ &= \gamma + \frac{d \ln L}{d \ln \theta} + \frac{\theta n}{L} \frac{h'}{h} \left( 1 + \frac{d \ln n}{d \ln \theta} - \frac{d \ln L}{d \ln \theta} \right). \end{aligned}$$

Substituting (C.2) in the above gives

$$\frac{d \ln y}{d \ln \theta} = \gamma + \kappa \frac{d \ln L}{d \ln \theta} + (1-\kappa) \left( 1 + \frac{d \ln n}{d \ln \theta} \right).$$

So substituting (24) and (C.6) into the above expression gives (25).

Finally, we derive equation (26). Profits are given by  $\pi = y - wL$ . Then

$$\frac{d \ln \pi}{d \ln \theta} = \frac{d \ln y}{d \ln \theta} \frac{y}{\pi} - w \frac{L}{\pi} \frac{L}{\pi},$$

or

$$\frac{d \ln \pi}{d \ln \theta} = \frac{d \ln y}{d \ln \theta} \frac{1}{1-\kappa} - w \frac{L}{\pi} \frac{\kappa}{1-\kappa}, \quad (\text{C.7})$$

where  $\kappa = wL/y$ . Substituting (24) and (25) into (C.7) and rearranging gives (26). □

## C.2 PROOF OF PROPOSITION 5

*Proof.* The following relationships are derived by combining (24)-(26):

$$\frac{d \ln L(\theta)}{d \ln y(\theta)} = \frac{(1 - \kappa(\theta))(1 + \varepsilon) + \gamma(\sigma + \varepsilon)}{(1 - \kappa(\theta))(1 + \gamma + \varepsilon) + \gamma(\kappa(\theta)\sigma + \varepsilon)}, \quad (\text{C.8})$$

$$\frac{d \ln \pi(\theta)}{d \ln y(\theta)} = \frac{(1 - \kappa(\theta) + \gamma)(1 + \varepsilon)}{(1 - \kappa(\theta))(1 + \gamma + \varepsilon) + \gamma(\kappa(\theta)\sigma + \varepsilon)}. \quad (\text{C.9})$$

From (C.8) and (C.9) we obtain

$$1 - \kappa(\theta) = \frac{1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}}{\frac{d \ln \pi(\theta)}{d \ln y(\theta)} - \frac{d \ln L(\theta)}{d \ln y(\theta)}}, \quad (\text{C.10})$$

rearranging equation (C.8), we have:

$$\gamma = \frac{\left(1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}\right) (1 - \kappa(\theta))(1 + \varepsilon)}{\frac{d \ln L(\theta)}{d \ln y(\theta)} (1 - \kappa(\theta) + \kappa(\theta)\sigma + \varepsilon) - (\sigma + \varepsilon)}. \quad (\text{C.11})$$

Substituting (C.10) into equation (C.11), we obtain equation (27). □

## D FIRM DISTORTIONS AND TAX ELASTICITIES

In this section we show that if there are no firm level distortions, a wage rate exists for the effective effort of the manager such that the income elasticity of the after tax rate equals the Frisch elasticity of labor supply.

We first show that at the optimum it is possible to write the income of managers of talent  $\theta$  as  $\pi(\theta) = \omega(\theta, w)n$  where  $\omega(\theta, w)$  is the wage of managers of talent  $\theta$  exercising effort  $n$ . The first order condition with respect to  $L$  is:  $\theta^\gamma H_2(\theta n, L) = w$ . Since  $H_2$  is a homogeneous of degree zero function we have  $\theta^\gamma H_2(\theta n/L, 1) = w$  so that  $\frac{\theta n}{L} = H_2^{-1}\left(\frac{w}{\theta^\gamma}, 1\right)$ . This relationship implies that for a given  $\theta$  and  $w$  the relationship between  $\theta n$  and  $L$  is linear. Define  $m(\theta, w) = 1/H_2^{-1}\left(\frac{w}{\theta^\gamma}, 1\right)$ . So that  $L = m(\theta, w)\theta n$ . Substituting in the expression for profits we have:  $\pi(\theta, n) = \theta^\gamma H(\theta n, m(\theta, w)\theta n) - m(\theta, w)w\theta n$ . Since  $H$  is homogeneous of degree one we have:

$$\pi(\theta, n) = \left[ \theta^{\gamma+1} H(1, m(\theta, w)) - w m(\theta, w) \theta \right] n = \omega(\theta, w)n.$$

We can now write the problem of the manager as:

$$\max c(\theta) - v(n(\theta)) \quad s.t. \quad c(\theta) = (1 - \tau)\omega(\theta, w)n.$$

First order conditions of the above problem can be written as  $n(\theta) = (v')^{-1}[(1 - \tau)\omega(\theta, w)]$ , so that:

$$\frac{\partial n}{\partial (1 - \tau)} = \frac{1}{v''(n(\theta))} \cdot \omega(\theta, w) = \frac{v'(n(\theta))}{v''(n(\theta))} \cdot \frac{1}{(1 - \tau)}, \quad (\text{D.1})$$

where the second equality follows from the first order condition. Substituting (D.1):

$$e \equiv \frac{\partial \log \omega(\theta, w)n}{\partial \log(1 - \tau)} = \frac{\partial n(\theta)}{\partial(1 - \tau)} \cdot \frac{1 - \tau}{n(\theta)} = \varepsilon.$$

This analysis would not apply in the case of a firm being subject to distortionary taxes, or if the size of the firm size is fixed.

## E ESTIMATING $a$ FROM THE INCOME DISTRIBUTION

In Subsection 5.3 we estimated  $a$  using the distribution of firm sizes. In this section we proceed similarly but focusing on the distribution of income, instead. From equation (26) in Lemma 1 we have that

$$\ln \pi(\theta) = \left(1 + \frac{\gamma}{1 - \kappa(\theta)}\right) (1 + \varepsilon) \ln \theta.$$

Approximating  $\kappa(\theta) = \bar{\kappa}$  and substituting in (29) we get

$$\frac{1}{a} = \frac{1}{\left(1 + \frac{\gamma}{1 - \bar{\kappa}}\right) (1 + \varepsilon)} \frac{1}{N} \sum_{i=1}^N \ln \pi(\theta^i).$$

Assume that in the data income is distributed according to a Pareto distribution with tail parameter  $a_\pi$ . We then have  $\frac{1}{a_\pi} = \frac{1}{N} \sum_{i=1}^N \ln \pi(\theta^i)$ . Substituting in the above we get:

$$a = \left(1 + \frac{\gamma}{1 - \hat{\kappa}}\right) (1 + \varepsilon) a_\pi.$$