



Taxing Atlas: Executive compensation, firm size, and their impact on optimal top income tax rates [☆]



Laurence Ales ^{a,*}, Antonio Andrés Bellofatto ^b, Jessie Jiayu Wang ^c

^a Tepper School of Business, Carnegie Mellon University, United States

^b School of Economics, University of Queensland, Australia

^c W.P. Carey School of Business, Arizona State University, United States

ARTICLE INFO

Article history:

Received 17 July 2016

Received in revised form 2 February 2017

Available online 16 February 2017

JEL classification:

D31

H21

H24

M12

M52

Keywords:

Optimal taxation

Managerial compensation

Span of control

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 show how to quantify the model using readily available firm-level data. Our benchmark calibration focuses on the US. Our results suggest that optimal top taxes are roughly in line with the current statutory rates and, thus, are significantly lower than what previous optimal taxation studies that ignore the scale-of-operations effect have shown. Similar quantitative findings hold when we extend the analysis to a panel of developed countries.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Heightened concerns over recent trends in income inequality necessarily bring the taxation of high earners to the forefront of the policy agenda.¹ 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 percent, others have challenged the validity of such elevated taxes on various grounds.² 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. Within

[☆] We thank Jeremy Bertomeu and Adriano Rampini for valuable discussions. We also thank Hank Bessembinder, Emilio Bisetti, Chris Edmond, Miguel Ferreira, Brent Glover, Burton Hollifield, Roozbeh Hosseini, Greg Kaplan, Finn Kydland, Christos Makridis, Pricila Maziero, Tomasz Piskorski, B. Ravikumar, Peter Rupert, Florian Scheuer, Ali Shourideh, Chris Sleet, Stephen Spear, Chris Telmer, Gustavo Ventura, Sunil Wahal, Randy Wright, Sevin Yeltekin, Ariel Zetlin-Jones and Marianne Holohan for editorial assistance. We also thank 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, 5th Tepper-LAEF conference in Pittsburgh, the University of Melbourne, 2015 WAMS/LAEF conference in Sydney, Arizona State University, University of Arizona, 2016 UNC Tax Symposium, and UNSW for feedback.

* Corresponding author.

E-mail addresses: ales@cmu.edu (L. Ales), a.bellofatto@uq.edu.au (A.A. Bellofatto), Jessie.Jiayu.Wang@asu.edu (J.J. Wang).

¹ 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).

² Mankiw et al. (2009), for instance, cast doubt on the identification of the ability distribution in Saez (2001), while Mankiw (2013) discusses various normative judgments which can invalidate the standard optimal tax formula used by Diamond and Saez (2011).

top income earners, we focus on 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. We find that top income taxes at the optimum are actually aligned with the current statutory rates in the US, unlike prevailing recommendations in the literature.

Our approach is to take the standard optimal taxation environment and augment it with a role for managers via a span of control technology. 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.³ 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, hire workers, and exert effort. The technology is such that managerial talent has a dual role. First, it affects effort productivity, which captures the quality of supervision or monitoring. Second, managerial ability affects the overall productivity of the firm. This channel captures the quality of indivisible managerial decisions and creates a so-called scale-of-operations effect.⁴ 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 scale-of-operations generates more right skewness in the distribution of firm sizes and pre-tax income. Such a mapping from talent to observables has profound implications on our quantitative analysis, as we discuss below.

We show that the constrained efficient allocation can be decentralized as a competitive equilibrium with taxes levied on income and on firm size (measured as the size 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.

We first look at optimal income taxation and provide a formula for the optimal top marginal tax rate. 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 implicit. 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 managers' 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.⁵

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 effect rises. This implies that any given dispersion of pre-tax income, for example, can now be rationalized with a smaller dispersion in skills. 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 a model which ignores the scale-of-operations effect. Hence, those individuals should not be subject to very high taxes, as implied by the textbook tax formula.

As anticipated above, our framework also provides normative grounds for the use of 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 managers. A tax formula for optimal marginal firm distortions is derived, through which we isolate sufficient conditions on the technology for marginal firm-size taxes to disappear. Such conditions are 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. Moreover, we show that the sign of the optimal marginal firm-size tax depends on the value of the elasticity of substitution in the production function.

We then take the model to the data to quantitatively evaluate the forces discussed above, and to provide specific normative recommendations on top optimal marginal tax rates. To determine tax rates, we require 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 \(1982\)](#). We start by deriving equilibrium restrictions relating the distribution of talent with firm size, sales, and profits. 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 well-established regularities of managerial compensation.

The second object that requires to be calibrated is the distribution of talent at the top. This is uncovered from the observed distribution of firm size and by exploiting the mapping between talent and firm size predicted by our model. Notably, here we depart from the established approach in public finance which instead recovers the ability distribution

³ 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%.

⁴ See [Mayer \(1960\)](#). [Rosen \(1982\)](#) shows that a positive scale-of-operations effect is necessary to reconcile stylized facts on managerial compensation.

⁵ 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 exhibits high levels of right skewness (or “thick” right tails).

directly from the distribution of income (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.⁶ 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 developed countries, including Australia, Canada, and nine European countries. We find that benchmark optimal top tax rates are mostly concentrated within the 34–50 percent range and are strikingly similar across the nations in our sample.

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 \(2008\)](#), and [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.⁷ The key in all of these models is that they introduce a nonlinear mapping between the compensation and talent of the manager. In particular, the distribution of compensation is more positively skewed than the distribution for talent, which is the key mechanism in our paper.⁸ 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 the optimal taxation of top income earners is vast.⁹ 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 classic taxation environment where wages are fixed exogenously. [Stiglitz \(1982\)](#) originally analyzes 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.¹⁰

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.¹¹ These authors consider 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.¹² Instead, the main mechanism which lowers top taxes in our paper is connected to the calibration of the skill distribution.¹³

[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. Differently from that work, our

⁶ [Saez et al. \(2012\)](#) report a top 1% marginal rate of approximately 42.5% for 2009. Using the Current Population Survey in the same period, we find top marginal income tax rates of 33.5% at the federal level and 5% at the state level. See Section 6 for details.

⁷ 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. \(2013\)](#).

⁸ 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.

⁹ For a review refer to [Mankiw et al. \(2009\)](#) and [Diamond and Saez \(2011\)](#).

¹⁰ 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 different forms of capital that interact differently with agents of diverse talent, while the latter study the problem of assigning workers with heterogeneous talents to tasks with heterogeneous complexity.

¹¹ The environment in this paper is static. For dynamic models that consider the modeling and taxation of entrepreneurial wealth, refer to [Quadriini \(2000\)](#), [Cagetti and Nardi \(2006\)](#), [Albanesi \(2011\)](#), [Shourideh \(2012\)](#), [Brüggemann \(2016\)](#).

¹² The 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. See Section 6.1 for details.

¹³ [Golosov et al. \(2016\)](#) and [Huggett and Badel \(2015\)](#) uncover different forces lowering optimal top tax rates in dynamic models.

model features no externalities.¹⁴ Instead, our approach to modeling managers is more in line with the empirical evidence in Kaplan and Rauh (2013), who find support for superstar effects in the US using data on income and wealth at the top.

Finally, this paper is related to the contemporary works of Ales and Sleet (2016) and Scheuer and Werning (2017). 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 by the tax code. One way to interpret this difference is to view Ales and Sleet (2016) and Scheuer and Werning (2017) 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 efficiency. In Section 4, we look at the decentralization of the optimum and derive the optimal tax formulas. 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. Section 8 concludes. The Appendix contains all proofs, robustness checks, and a numerical evaluation of firm size distortions.

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.¹⁵

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 $\Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_{++}$. Managerial talent, θ , 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 θ produce final output y according to:

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

where L is hired labor and $\gamma \geq 0$ is the scale-of-operations parameter (see below). 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, θ is effort-augmenting, as it multiplies n within H . Second, θ improves total factor productivity since $\gamma \geq 0$. We refer to the latter effect γ as the *scale-of-operations effect*, following Mayer (1960).¹⁶ This formulation is in line with the one in Rosen (1982), where managers' actions naturally affect the productivity of all workers under their supervision. But unlike the technology in Rosen's paper, we also incorporate elastic managerial effort n as an intensive margin.¹⁷

In what follows, it will be convenient to define $n(y, L, \theta)$ as the effort required by a manager of talent θ 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 θ , $n(\theta)$ and $\theta n(\theta)$ are private information to each θ -agent.

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

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

and

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

¹⁴ For a general treatment of optimal taxation with positive or negative externalities, refer to Rothschild and Scheuer (2016).

¹⁵ Empirical analyses typically indicate that income effects are relatively small compared to substitution effects. See, e.g., Blundell and MaCurdy (1999). Moreover, as argued by Diamond (1998), assuming small income effects for very high income individuals appears suitable due to the presence of large estates.

¹⁶ See Bartelsman and Doms (2000) and references therein for additional details on the relationship between managerial talent, firm size, and firm productivity.

¹⁷ A model featuring capital (or any additional adjustable input) can be analyzed along the lines of Terviö (2008) and Ales and Sleet (2016). In this case, the additional assumptions required are that the choice for capital be undistorted neither by taxation nor by credit constraints.

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 θ .

3. Efficiency

In this section, we characterize efficient allocations. We use a direct mechanism in which managers report their talent θ 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 θ who mimics a manager of talent θ' . In this case, manager θ 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:

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

Efficient allocations solve the following planning problem:

$$\max_{c^w, \{c(\theta), y(\theta), L(\theta)\}_{\theta \in \Theta}} \Psi(c^w) + \int_{\Theta} \Psi(c(\theta) - v(n(y(\theta), L(\theta), \theta))) dF(\theta), \quad (PP)$$

s.t. (2), (3) and (4).

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_θ , 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.¹⁸ The condition holds, for example, when c , y , and L are increasing in θ , and $n_{\theta L}$ is small enough.^{19,20} 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:

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

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

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

$$U'(\theta) = -v'(n(y(\theta), L(\theta), \theta))n_\theta(y(\theta), L(\theta), \theta), \quad \text{for all } \theta \in \Theta. \quad (9)$$

¹⁸ 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. (2016), and Farhi and Werning (2013) apply similar techniques in dynamic environments.

¹⁹ It is straightforward to verify that $n_\theta, n_{\theta y} \leq 0$. The sign of $n_{\theta L}$ depends on the value of the elasticity of substitution between n and L . In particular, when H displays constant returns to scale, $n_{\theta L} \geq 0$ as long as the elasticity of substitution is greater than one.

²⁰ 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 .

Before presenting the decentralization and showing the properties of the optimal tax system, we discuss the incentive constraint (9) further. In the standard Mirrleesian environment, output, talent, and effort are related through $y(\theta) = \theta n(\theta)$, so that effort required to generate output y is $n(y, \theta) = y/\theta$. In that case, the term n_θ appearing on the right hand side of (9) 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 $y = \theta^\gamma Lh\left(\frac{\theta n(y, L, \theta)}{L}\right)$, where $h(x) \equiv H(x, 1)$ for all $x \geq 0$. Consequently, $n(y, L, \theta) = h^{-1}\left(\frac{y}{\theta L}\right)\frac{L}{\theta}$ and

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

which simplifies to

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

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_θ and, hence, the right hand side of the incentive constraint (9) 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.

4. Tax implementation

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

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

$$\begin{aligned} \max_{c, y, L} \quad & c - v(n(y, L, \theta)) & \text{(MP)} \\ \text{s.t.} \quad & c \leq y - wL - T_L(wL) - T(y - wL - T_L(wL)), & \text{(11)} \end{aligned}$$

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.²¹ Since workers in our environment supply labor inelastically, their problem is characterized by a simple budget constraint $c^w = w + \phi$, where ϕ 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 θ -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:

$$\int [T(y(\theta) - wL(\theta) - T_L(wL(\theta))) + T_L(wL(\theta))] dF(\theta) = G + \phi. \tag{12}$$

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 (PP). Then there exists 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. \square

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.

²¹ The split between income tax and taxes depending on firm size is indeterminate. However, the split considered in this paper is implementable with the information currently collected in the US from corporate and individual income tax returns.

4.1. Income taxation

We begin by analyzing income taxation, which is the main focus of the paper. To facilitate intuition, our optimal tax formulas will be written using derivatives of the production function rather than derivatives of required effort $n(y, L, \theta)$ (e.g., we exploit that, for a given level of output, $n_y = 1/y_n$). Assuming differentiability of T' , first order conditions from the manager's problem (MP) imply:

$$1 - T'(\pi(\theta)) = \frac{v'(n(\theta))}{y_n(n(\theta), L(\theta), \theta)}, \quad (13)$$

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

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 $\varepsilon > 0$.

The next proposition characterizes optimal marginal income tax rates. To simplify notation, we refer to $T'(\pi(\theta))$ as $T'(\theta)$, unless stated otherwise. Additionally, partial derivatives of the production function given $(n(\theta), L(\theta), \theta)$ are denoted by $y_i(\theta)$, for $i = n, L, \theta$.

Proposition 3. Suppose Assumption 1 holds. Then at any solution to (PP-FOC), $T'(\theta)$ satisfies:

$$\frac{T'(\theta)}{1 - T'(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \cdot \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \cdot \frac{y_\theta(\theta)}{y_n(\theta)} \frac{\theta}{n(\theta)} \cdot \left[\frac{1}{\varepsilon} + \frac{d \ln \left(\frac{y_\theta(\theta)}{y_n(\theta)}\right)}{d \ln n(\theta)}\right], \quad (14)$$

where

$$D(\theta) \equiv \frac{1}{1 - F(\theta)} \int_{\theta}^{\bar{\theta}} \Psi'(U(\theta)) dF(\theta). \quad (15)$$

Proof. See Appendix B.2. \square

Equation (14) highlights the main forces that shape optimal marginal income tax rates in our framework. The first two terms are well known in the literature. 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 θ are attractive when the mass of managers above θ , given by $(1 - F(\theta))$, is large. At the same time, the resulting distortion is proportional to the mass of individuals at θ 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 decrease as the corresponding social welfare weight rises.

The third term 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, $(y_\theta/y_n)(\theta/n)$ 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 channel is encapsulated by the Frisch elasticity ε , which measures how changes in after-tax compensation affect the supply of effort (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.

In addition to the conventional channel just described, effort variations also modify the relative contribution of skills and effort into output and, hence, informational rents. Such an effect is captured by the elasticity $d \ln (y_\theta/y_n) / d \ln n$. Given that this coefficient is positive,²² reductions in effort lower the relative productivity of talent. This latter lowers informational rents, which increases the social surplus. In turn, a high value of the elasticity $d \ln (y_\theta/y_n) / d \ln n$ provides a rationale for high tax rates since, all else equal, the resulting decrease in effort can increase social welfare. To be more precise about this channel, note that, for a given θ , the envelope condition (9) can be written as

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} v'(n(\tilde{\theta})) \frac{y_\theta(\tilde{\theta})}{y_n(\tilde{\theta})} d\tilde{\theta}, \quad (16)$$

²² Recall that $y_{\theta n} \geq 0$ and $y_{nn} \leq 0$.

where we used $n_\theta = -y_\theta/y_n$. This expression makes it clear that informational rents of the managers, encapsulated in $U(\theta)$, decrease as y_θ/y_n falls (*ceteris paribus*).

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:

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

Substituting into (14), we obtain the classic tax formula from Diamond (1998) or Saez (2001) in terms of underlying structural parameters:

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

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

4.2. Top income taxation under CES technology

Proposition 3 applies to quite general production functions H . To lay the groundwork for our quantitative analysis, we make the following parametric assumption on the production function²³:

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

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

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

The following corollary characterizes optimal income taxes under Assumption 2.^{24,25}

Corollary 1. Suppose Assumptions 1 and 2 hold. Then at any solution to (PP-FOC), $T'(\theta)$ satisfies:

$$\frac{T'(\theta)}{1 - T'(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \left(1 - \frac{D(\theta)}{D(\theta)} \right) \underbrace{\left(1 + \frac{\gamma}{1 - \kappa(\theta)} \right)}_{= \frac{y_\theta(\theta)}{y_n(\theta)} \frac{\theta}{n(\theta)}} \left[\frac{1}{\varepsilon} + \underbrace{\frac{1 + \frac{\gamma}{1 - \kappa(\theta)} (1 - \rho \kappa(\theta))}{1 + \frac{\gamma}{1 - \kappa(\theta)}}}_{= d \ln \left(\frac{y_\theta(\theta)}{y_n(\theta)} \right) / d \ln n(\theta)} \right], \tag{17}$$

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

Proof. See Appendix B.3. □

The scale-of-operations effect The parameter γ impacts optimal tax rates via three channels. The first two are explicit in formula (17), but the last one is not. We discuss each effect in turn.

First, a larger γ leads to a larger output-talent elasticity relative to effort through $(y_\theta/y_n)(\theta/n)$, which translates into higher marginal taxes. Second, under the CES specification, $d \ln (y_\theta/y_n) / d \ln n$ decreases with γ . In other words, the relative weight of skills and effort into output becomes less sensitive to changes in n as γ 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 (17).

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 γ grows. This effect reduces optimal tax rates through the first term on the right-hand side of (17). While this channel will be analyzed in detail in Section 5 below, here we provide a heuristic illustration.

²³ 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.

²⁴ The case of a Cobb–Douglas production function (i.e., $\rho = 0$) reverts to equation (DS) after renormalizing skills properly. We can also use the income distribution to recover a standard expression. This analysis is contained in Section 6.1.

²⁵ Note that, under Assumption 2, we specialize the discussion to a constant return to scale production function. Scheuer and Werning (2017) consider a related environment which does not impose constant returns to scale, but which does not incorporate a scale-of-operations effect as considered in this paper.

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

$$\frac{1 - F(\theta)}{\theta f(\theta)} = \frac{1 - F_z(z)}{zf_z(z)} \left(\frac{d \ln z}{d \ln \theta} \right)^{-1}, \quad (18)$$

so that the left hand side, and hence optimal tax rates, are inversely related to the elasticity of z with respect to θ . Crucially, in a decentralized environment, such an elasticity increases with γ , 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 more 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: $\lim_{\theta \rightarrow \bar{\theta}} D(\theta) = 0$.²⁶

Let $\lim_{\theta \rightarrow \bar{\theta}} \kappa(\theta) = \bar{\kappa}$. Taking limits on (17) 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 the optimal marginal tax at the top satisfies:

$$\frac{T'(\bar{\theta})}{1 - T'(\bar{\theta})} = \frac{1}{a} \left(\frac{1}{\varepsilon} + 1 \right) + \frac{1}{a} \frac{\gamma}{1 - \bar{\kappa}} \left(\frac{1}{\varepsilon} + 1 - \rho \bar{\kappa} \right). \quad (19)$$

Corollary 2 contains the main tax formula in our analysis. In the following sections, we quantitatively evaluate the different effects shaping top tax rates in equation (19).

4.3. Firm size taxation

While our main focus is on the shape of optimal income taxes, firm distortions are generally necessary to implement the efficient allocations, as implied by Proposition 2. We thus close this section by characterizing optimal taxes on firm size T_L .

Assuming differentiability of T_L , first order conditions from the manager's problem (MP) give:

$$1 + T'_L(wL(\theta)) = \frac{y_L(n(\theta), L(\theta), \theta)}{w}, \quad (20)$$

where y_L is the marginal product of the worker.²⁷

Equation (20) shows that if $T'_L(wL(\theta)) \neq 0$ for some θ , the worker's marginal product is not equalized across firms. This implies a breakdown of the well-known Diamond and Mirrlees (1971) productive efficiency result. The reason behind the presence of labor misallocation at the optimum was already discussed in Section 3: by affecting L , the social planner can relax incentive constraints. The next proposition provides a formula for optimal firm size distortions. We use $T'_L(\theta)$ instead of $T'_L(wL(\theta))$ to avoid clutter.

Proposition 4. At any solution of (PP-FOC), there exists a w such that $T'_L(\theta)$ satisfies:

$$\frac{T'_L(\theta)}{(1 + T'_L(\theta))(1 - T'(\theta))} = \frac{1 - F(\theta)}{\theta f(\theta)} \cdot \left(1 - \frac{D(\theta)}{D(\underline{\theta})} \right) \cdot \frac{y_\theta(\theta)}{y_L(\theta)} \frac{\theta}{L(\theta)} \cdot \frac{d \ln \left(\frac{y_\theta(\theta)}{y_n(\theta)} \right)}{d \ln L(\theta)}, \quad (21)$$

where $D(\theta)$ is defined by (15).

²⁶ By assuming zero social weight at the top, we implicitly focus on an upper bound for the optimal marginal income tax at the top. This is clearly an extreme assumption, but still provides a benchmark that allows for easy comparison with the bulk of the optimal taxation literature. More recently, Weinzierl (2014) looks at survey evidence on preferred societal welfare criteria. Saez and Stantcheva (2016) derive tax formulas under arbitrary marginal social weights.

²⁷ Here we use that, for a given level of output, $y_L = -n_L/n_y$.

Proof. See Appendix B.4. \square

In the expression of Proposition 4, marginal firm size taxes T'_L are being normalized by $(1 + T'_L)(1 - T')$, which is the after-tax cost of labor in wage units.²⁸ The first two terms on the right hand side of (21) embed the effects of the talent distribution and of redistributive tastes on marginal taxes. Those coefficients were also present in our income tax formula (14), and the same intuition applies here. In contrast, the last two terms in (21) do deserve some further explanation, to which we now turn.

The coefficient $(y_\theta/y_L)(\theta/L)$ is the ratio between the output elasticity of talent and the output elasticity of labor input. A high value of this coefficient implies that the relative contribution of skills to output is high compared to the contribution of employed labor. In that case, firm distortions affecting labor input decisions do not entail large efficiency costs, which permits the planner using a high T'_L to provide incentives. Clearly, the ratio $(y_\theta/y_L)(\theta/L)$ is the analogue of $(y_\theta/y_n)(\theta/n)$ showing up in the income tax formula (14).

The last term, $d \ln(y_\theta/y_n)/d \ln L$, captures how variations in labor demand affect the informational rent of the managers. As discussed in Section 4.1, informational rents vary with the size of y_θ/y_n across the skill distribution (see equation (16)). In the previous section, this fact created a force for increasing marginal income tax rates, as captured by the elasticity $d \ln(y_\theta/y_n)/d \ln n \geq 0$ in (14). The term $d \ln(y_\theta/y_n)/d \ln L$ has an analogous effect on T'_L . Specifically, if that elasticity is positive (see below), introducing a positive marginal tax on firm size is optimal, as the corresponding decrease in L lowers the informational rent of the managers (all else equal). Essentially, in our model, informational rents change not only by affecting y_θ/y_n via n (the standard channel in the literature), but also by changing L .

The elasticity $d \ln(y_\theta/y_n)/d \ln L$ is particularly important as it pins down the sign of T'_L . Unfortunately, it is not possible to sign this coefficient unless parametric assumptions are made. The next corollary derives the expression of T'_L under the CES technology of Assumption 2 and shows that the sign of $d \ln(y_\theta/y_n)/d \ln L$ actually coincides with ρ .

Corollary 3. Suppose Assumption 2 holds. Then, at any solution of (PP-FOC), there exists a w such that $T'_L(\theta)$ satisfies:

$$\frac{T'_L(\theta)}{(1 + T'_L(\theta))(1 - T'(\theta))} = \frac{1 - F(\theta)}{\theta f(\theta)} \left(1 - \frac{D(\theta)}{D(\underline{\theta})} \right) \gamma \rho, \quad (22)$$

where $D(\theta)$ is defined by (15).

Proof. See Appendix B.5. \square

As evident from equation (22), firm size distortions vanish in the absence of a scale-of-operations effect ($\gamma = 0$), or when working with a Cobb–Douglas production function ($\rho = 0$). Moreover, equation (22) allows us to sign T'_L based on the value of the elasticity of substitution between θn and L embedded in ρ . In particular, T'_L is strictly positive (negative) if $\rho > 0$ (< 0), i.e., if there is high (low) substitutability between managerial effort and labor. Interestingly, this pattern goes against the conventional inverse rule between taxes and elasticities.²⁹

5. Identification and estimation

The tax formula in (19), 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 γ 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 γ and a , respectively. The ongoing assumption of this section is that firms at the top are not subject to any size-dependent form of taxation.

5.1. Firm level elasticities

Our approach is along the lines of Rosen (1982). We make the following assumption that will simplify the identification of the key parameters of the paper:

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

²⁸ More precisely, each additional unit of labor employed carries a marginal tax of wT'_L . So after taxes, a manager pays $w(1 + T'_L)(1 - T')$ per unit of labor. The latter after-tax rate incorporates the factor $(1 - T')$ since labor costs are being deducted from taxable income in our decentralization (see (11)).

²⁹ In Appendix C, we quantify the optimal marginal tax on firm size T'_L within a numerical example.

Consider a competitive equilibrium where the manager faces a linear tax τ 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.³⁰ Due to the constant returns to scale assumption, we can write the θ -manager's problem as

$$\begin{aligned} \max_{\{L, n, \pi\}} & (1 - \tau)\chi\pi - v(n) \\ \text{s.t.} & \pi = \left[\theta^\gamma L h \left(\frac{\theta n}{L} \right) - wL \right], \end{aligned} \quad (23)$$

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 (23) are given by:

$$\theta^\gamma \left[h \left(\frac{\theta n}{L} \right) - \frac{\theta n}{L} h' \left(\frac{\theta n}{L} \right) \right] = w, \quad (24)$$

and

$$(1 - \tau)\chi\theta^{\gamma+1} h' \left(\frac{\theta n}{L} \right) = v'(n). \quad (25)$$

Equations (24) and (25) together imply the following lemma, where we show how firm size, output, and profits move together with respect to managerial talent.

Lemma 1. *Suppose Assumptions 1 and 4 hold. Let $\{L(\theta), n(\theta), \pi(\theta)\}$ solve the θ -manager's problem in (23), and let σ be the elasticity of substitution between θn and L . Then the following relationships hold:*

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

$$\frac{d \ln y(\theta)}{d \ln \theta} = 1 + \gamma + \varepsilon \left(1 + \frac{\gamma}{1 - \kappa(\theta)} \right) + \frac{\kappa(\theta)}{1 - \kappa(\theta)} \gamma\sigma, \quad (27)$$

$$\frac{d \ln \pi(\theta)}{d \ln \theta} = \left(1 + \frac{\gamma}{1 - \kappa(\theta)} \right) (1 + \varepsilon), \quad (28)$$

where $\kappa(\theta) \equiv y_L(\theta)L(\theta)/y(\theta)$.

Proof. See Appendix C.1. \square

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 γ grows. In particular, once we shut down the scale-of-operations effect ($\gamma = 0$), then the growth rates in (26)–(28) are all equal to $(1 + \varepsilon)$. However, L , y , and π respond *more* to differences in managerial talent when γ is positive.³¹

It should be noted that the ownership share χ and the top marginal income tax rate τ do not impact the elasticities in Lemma 1. Consequently, our estimation strategy in the following section does not require information on χ and τ . A few assumptions underlying this feature are key. First, χ and τ are constant across talent.³² Second, preferences are quasi-linear. Third, the disutility for effort is isoelastic.

5.2. Estimating the scale-of-operations effect γ

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 θ -manager's problem in (23) the following relationship holds:*

³⁰ 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.

³¹ 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.

³² 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.

$$\gamma = \frac{1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}}{\frac{d \ln L(\theta)}{d \ln y(\theta)} - \frac{d \ln \pi(\theta)}{d \ln y(\theta)} \frac{\sigma + \varepsilon}{1 + \varepsilon}}. \quad (29)$$

Proof. See Appendix C.2. \square

Equation (29) forms the basis for the estimation of γ . Specifically, to evaluate γ , we require the elasticity of firm size with respect to sales ($d \ln L(\theta)/d \ln y(\theta)$), the elasticity of managerial compensation with respect to sales ($d \ln \pi(\theta)/d \ln y(\theta)$), the elasticity of substitution (σ), and the Frisch elasticity (ε). 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:

$$\ln y_t(\theta_i) = \alpha_0 + \alpha_1 \ln L_t(\theta_i) + \sum_{j=1}^{10} \alpha_{2,j} \text{Div}_j + \varepsilon_{i,t}, \quad (30)$$

where $\ln y_t(\theta_i)$ is the log of firm sales, $\ln L_t(\theta_i)$ is the log of firm size, and 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 2012.³³ 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.³⁴

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 (30) 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. (2016).

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 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 Fig. 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.

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.³⁵ This literature has highlighted an empirical regularity, usually denoted as “*Roberts’s Law*,” which states that, on average, managerial compensation is a power law of firm size. When proxying firm size using sales, the corresponding exponent is estimated at around 1/3, so in our benchmark calculation we set $\frac{d \ln \pi}{d \ln y} = 1/3$.³⁶

³³ From our sample we drop firms that report negative or zero sales and firms with duplicate CUSIP.

³⁴ 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.

³⁵ See also Lewellen and Huntsman (1970), Baker et al. (1988), Gabaix and Landier (2008), Frydman and Saks (2010), or Alder (2016).

³⁶ Baker, Jensen and Murphy state that “*The best documented empirical regularity regarding levels of executive compensation is an elasticity of compensation with respect to firm sales of about 0.3 (...)*” (Baker et al., 1988, page 609).

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	≥ 5	≥ 5	≥ 5	≥ 5	All	≥ 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 elasticity of sales with respect to workers (α_1) in (30). 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 (30) 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” denotes the sample of firms by size included in the estimation of α_1 . We report standard errors in square parentheses.

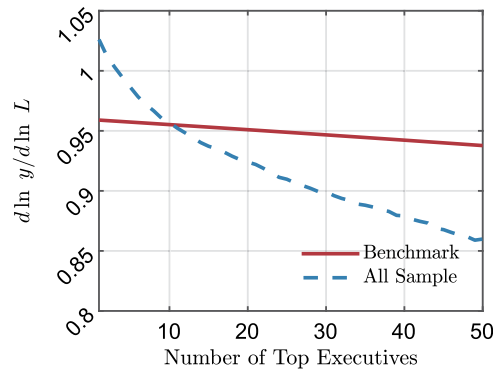


Fig. 1. Estimates of elasticity of sales with respect to workers ($\frac{d \ln y}{d \ln L}$) by number of top executives. “Benchmark” refers to estimates of (30) using firms above the median size. “All Sample” refers to estimates of (30) also including firms below the median firm size.

Frisch elasticity (ε) and elasticity of substitution (σ) We set the value of ε based on previous studies. Specifically, following the guidelines of Chetty et al. (2011), we set ε equal to 0.5 for our benchmark calibration.³⁷ As for the elasticity of substitution σ , we assume it is constant (as in the case with CES technology) and rely on equilibrium modeling restrictions to pin down this parameter. Given the values for the elasticities $d \ln \pi / d \ln y$ and $d \ln y / d \ln L$ estimated previously, equation (29) does not return a positive value of γ for every possible combination of ε and σ . Indeed, using $\varepsilon = 0.5$, we have that $\gamma \geq 0$ if and only if $\sigma \geq 4.25$. Based on this threshold, we set $\sigma = 5$ ($\rho = 4/5$) in the benchmark calibration and perform a robustness check on this parameter in Section 6. Other values of ε would yield a different feasible range for σ , as it is shown in Fig. 2. Overall, from Fig. 2, we see that the feasible range for σ precludes a Cobb–Douglas production function ($\sigma = 1$) for empirically relevant values of ε .

We can now determine γ using equation (29) 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 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.

³⁷ 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. (2016) and French (2005). Chetty et al. (2011) find values equal to 0.5 on the intensive margin and 0.25 on the extensive margin.

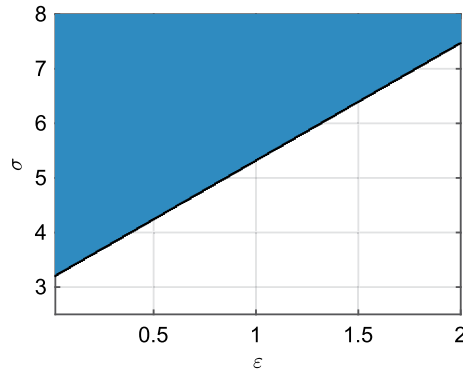


Fig. 2. Region for the elasticity of substitution (σ) and Frisch elasticity (ε) consistent with a positive scale-of-operations parameter (γ).

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 satisfies³⁸:

$$\frac{1}{\hat{a}} = \frac{1}{N} \sum_{i=1}^N (\ln(\theta_i) - \ln(\underline{\theta})), \tag{31}$$

where $\{\theta_1, \dots, \theta_N\}$ is a given a realization of managerial talent, and $\underline{\theta}$ is the minimum possible value of θ . If we let a_L denote the tail parameter of the Pareto distribution of firm size, the analogue to (31) yields:

$$\frac{1}{\hat{a}_L} = \frac{1}{N} \sum_{i=1}^N (\ln(L(\theta_i)) - \ln(L(\underline{\theta}))). \tag{32}$$

From equation (26) in Lemma 1, we have that:

$$\ln(L(\theta)) - \ln(L(\underline{\theta})) = \left(1 + \frac{\gamma\sigma}{1 - \kappa(\theta)} + \varepsilon \left(1 + \frac{\gamma}{1 - \kappa(\theta)}\right)\right) (\ln(\theta) - \ln(\underline{\theta})). \tag{33}$$

Finally, combining (31)–(33) and taking limits, we can link the tail parameters on the talent and firm size distributions as follows³⁹:

Proposition 6. Suppose Assumptions 1 and 4 hold. Then for any solution of the θ -manager’s problem in (23) we have:

$$a = \left(1 + \frac{\gamma\sigma}{1 - \bar{\kappa}} + \varepsilon \left(1 + \frac{\gamma}{1 - \bar{\kappa}}\right)\right) \times a_L. \tag{34}$$

Using equation (34) 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 (26)–(28) from Lemma 1, we obtain⁴⁰:

$$1 - \kappa(\theta) = \frac{1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}}{\frac{d \ln \pi(\theta)}{d \ln y(\theta)} - \frac{d \ln L(\theta)}{d \ln y(\theta)}}. \tag{35}$$

Our approach is to use the estimated elasticities $d \ln L/d \ln y$ and $d \ln \pi/d \ln y$ into (35) to approximate $\bar{\kappa}$. We get $\bar{\kappa} \approx 0.93$.⁴¹

³⁸ See, e.g., Malik (1970).

³⁹ We assume that regularity conditions necessary for consistency of maximum likelihood estimates hold.

⁴⁰ See equation (C.10) in Appendix C.2.

⁴¹ 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 $\frac{0.08}{20} / \frac{0.93}{20,000} \approx 86.2$. 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 PAYSACLE (see www.payscale.com/data-packages/ceo-income/full-list). The average across these top firms is 87.4, which closely matches what the estimated κ implies.

Table 2
Benchmark parameter calibration.

Parameter	Symbol	Value
Scale-of-operations parameter	γ	0.30
Pareto tail parameter of the talent distribution	a	26.13
Frisch elasticity	ε	0.5
Elasticity of substitution	σ	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: The elasticity of substitution (σ) and Frisch elasticity (ε) are imposed exogenously. Firm elasticities are estimated from the data as described in Section 5.2. γ , a , and κ are computed using (29), (34), and (35), respectively.

Table 3
Top marginal tax rates.

σ	$\varepsilon = 0.25$		$\varepsilon = 0.50$	
	4.5	10	4.5	10
$T'(\bar{\theta})$	51.5%	53.2%	31.7%	34.9%
γ	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 (19). The values for the scale-of-operations parameter (γ) and the Pareto tail parameter of the talent distribution (a) are computed using (29) and (34), respectively. $T'_{\gamma=0}(\bar{\theta})$ denotes the optimal tax rates with the exogenous constraint of $\gamma = 0$.

Plugging in our estimates into equation (34), 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 .⁴² Fundamentally, this big difference relies on a positive scale-of-operations effect γ .

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 (19), 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 γ and its influence on a . To see this, it is instructive to compare the result with the case in which the scale-of-operations effect 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 (ε) and the degree of substitutability across inputs (σ) on the optimal top marginal tax rates. These are reported in Table 3. As expected, we observe that tax rates are decreasing in ε . To evaluate the impact of σ , we use two values. The first one ($\sigma = 4.5$) is close to the lower bound consistent with $\gamma \geq 0$ in the benchmark calibration. The second value ($\sigma = 10$) is an arbitrarily large number. We see that marginal tax rates are increasing in σ , but optimal tax rates are relatively insensitive to variations in this parameter. It is worth noting that whenever we change ε or γ , we recompute γ and a using (29) and (34), 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. Fig. 3 displays the previous observations for wider ranges of ε and σ .

⁴² The same conclusion holds when calibrating a using the income distribution, as discussed in Appendix E.

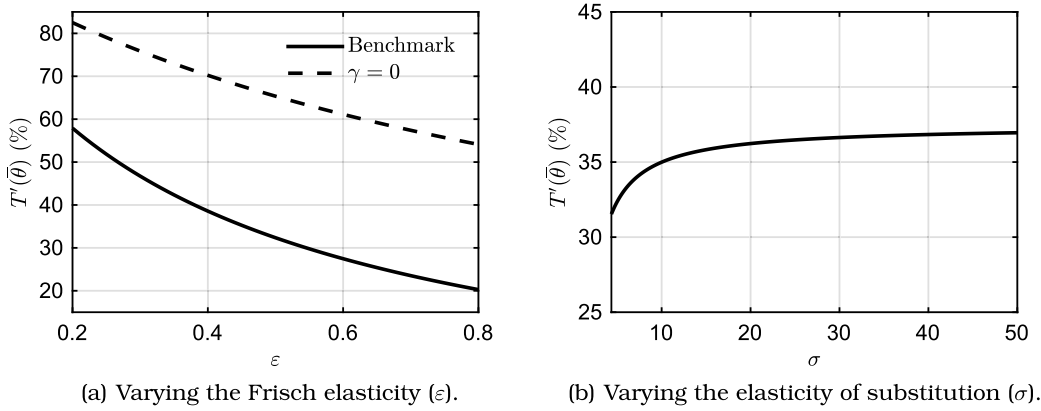


Fig. 3. Comparative statics of optimal top tax rates.

To conclude, it is useful to compare the baseline optimal tax rate of 32.4 percent with what we see in US data. The empirical top tax rate is recovered as follows. 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.⁴³ For the top 99th 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 with what we currently observe in US data.

6.1. Connections to previous studies

Relationship with Saez (2001) In Section 4, we showed how the classic (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_π , by applying the following identity (see Appendix E for a derivation):

$$a = \left(1 + \frac{\gamma}{1 - \bar{k}}\right) (1 + \varepsilon) a_\pi. \tag{36}$$

We can next substitute this value in our previous tax formula (19) to get:

$$T'(\bar{\theta}) = \frac{1}{1 + a_\pi \varepsilon \left(1 + \frac{\gamma}{1 - \bar{k}}\right) \left[1 + \frac{\gamma}{1 - \bar{k}} \left(1 - \frac{\rho \bar{k} \varepsilon}{1 + \varepsilon}\right)\right]^{-1}}. \tag{37}$$

As discussed in Diamond and Saez (2011), marginal taxes at the top can be written in terms of the tail parameter of the distribution of income a_π , and the income elasticity of the after-tax rate $e \equiv \frac{\partial \log \pi(\theta)}{\partial \log(1 - \tau)}$. In our environment we are unable to provide such a characterization. The reason is that our measure of e depends on the curvature of firm size taxes (i.e., T_L''), which cannot be pinned down analytically.⁴⁴ Nonetheless, equation (37) shows that we can still write our top tax formula in terms of ε , which is the elasticity of taxable income in the absence of firm size taxes (see Appendix D). This equation also allows us to provide an immediate connection with the literature. In particular, with no scale-of-operations effect ($\gamma = 0$), or in the Cobb–Douglas case ($\rho = 0$), equation (37) simplifies to $T' = 1/(1 + a_\pi \varepsilon)$. This is the same formulation presented in Saez (2001), and it is consistent with Corollary 3 and the fact either of these cases lead to zero firm size taxes. For all other cases with $\gamma > 0$ and $\rho > 0$, we can think of the term $\left(1 + \frac{\gamma}{1 - \bar{k}}\right) \left[1 + \frac{\gamma}{1 - \bar{k}} \left(1 - \frac{\rho \bar{k} \varepsilon}{1 + \varepsilon}\right)\right]^{-1}$ in (37) as providing an (upward) adjustment to the income elasticity of the after-tax rate when firm distortions are zero. We refer the reader to Scheuer and Werning (2017) for an analysis on how various elasticity measures need to be adjusted in a related environment with superstar effects.

⁴³ 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.

⁴⁴ Appendix G numerically characterizes the progressivity of firm size taxes within some calibrated examples.

Remark 1. The standard approach in public finance to the issue of optimal top income taxation relies 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, recognizes that this elasticity is *endogenous* to policy, and that firm level distortions may have a particularly strong effect. 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 from 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 and Ales et al., 2015).

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. Indeed, a crucial assumption in Stiglitz (1982) is that the government uses a single tax schedule on different types of labor (this is relaxed in Scheuer (2014), who analyzes both cases in an entrepreneurial model with endogenous entry). Our environment, on the other hand, features differential tax treatment of managers and workers.

7. Cross-country results

A contribution of this paper is to make operational the study of optimal top income tax rates by using 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. In addition, ORBIS, the multi-country database published by Bureau van Dijk, covers information (including sales) on more than 50 million companies world-wide. Moreover, compensation data across countries for individual top executives 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.⁴⁵ 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 $a_L = \frac{d \ln y}{d \ln L} a_y$. The estimates of CEO pay-sales elasticities $\frac{d \ln \pi}{d \ln y}$ are taken from Fernandes et al. (2013), whose sample is restricted to countries having mandated disclosure of executive compensation. 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 γ , a , and optimal top tax rates with and without the scale-of-operations effect. The results for the US are displayed at the bottom of the table for comparison. Fig. 4 displays the scatter plots by country of the optimal top income tax rates versus $\log(a+1)$ and $\log(\gamma+1)$.

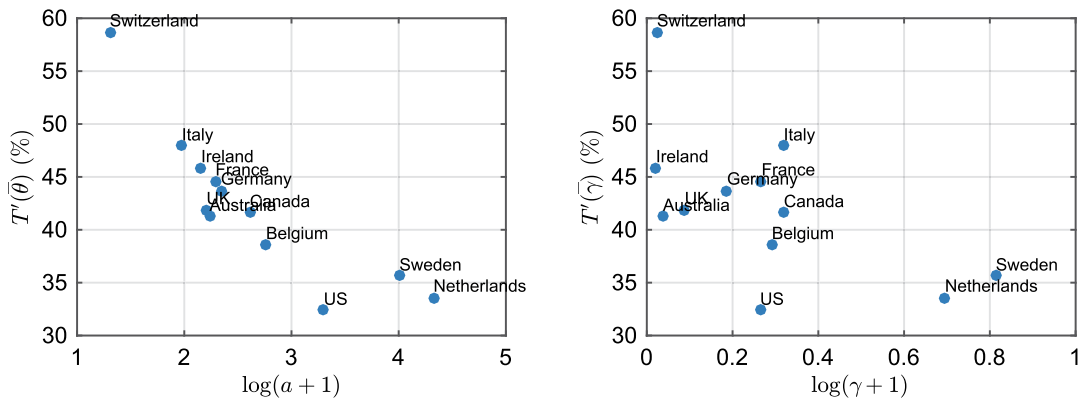
Our findings on the optimal top income tax rates are strikingly robust across countries. For all the non-US countries, the estimated scale-of-operations effect γ is positive, and the estimated distribution of talent is significantly less skewed than firm size. Consequently, the optimal tax rates are 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 are explained by variations in the firm size tail index (Column 1) and in the pay-size elasticity (Column 2). The US, the Netherlands, and Sweden have high values of γ 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 vary more significantly with a than γ . This confirms that the effect from a more compressed managerial talent distribution dominates in lowering top tax rates. Finally, in the last column, Table 4 contains actual top statutory personal income tax rates taken from OECD data. When comparing optimal taxes to their statutory counterparts, those rates are generally much closer to each other in our framework than in the case with $\gamma = 0$ (Sweden is the exception due to the combination of two factors: a high value of a and a high statutory rate.)

⁴⁵ We thank Miguel Ferreira for sharing the estimates of CEO pay-sales elasticities in thirteen non-US countries with mandated disclosure of executive compensation. Norway and South Africa are excluded from the sample: Norway has low BoardEx coverage resulting in a negative Roberts’ law estimate, while South Africa has low ORBIS coverage.

Table 4
Top tax rates across countries.

Country	a_L	$\frac{d \ln \pi}{d \ln y}$	$\frac{d \ln y}{d \ln L}$	γ	a	$T'(\bar{\theta})$	$T'_{\gamma=0}(\bar{\theta})$	T'_{Data}
Australia	0.97	0.41	0.98	0.04	8.34	0.41	0.67	0.49
Belgium	0.89	0.37	0.92	0.34	14.84	0.39	0.69	0.45
Canada	0.80	0.38	0.90	0.38	12.61	0.42	0.71	0.50
France	0.79	0.42	0.89	0.31	8.88	0.45	0.72	0.54
Germany	0.79	0.39	0.93	0.21	9.39	0.44	0.72	0.48
Ireland	0.75	0.38	0.99	0.02	7.64	0.46	0.73	0.47
Italy	0.82	0.52	0.80	0.38	6.22	0.48	0.71	0.49
Netherlands	0.86	0.30	0.95	1.00	74.50	0.34	0.70	0.49
Sweden	0.82	0.31	0.92	1.26	53.94	0.36	0.71	0.57
Switzerland	0.77	0.60	0.97	0.02	2.69	0.59	0.72	0.36
UK	0.97	0.42	0.96	0.09	8.08	0.42	0.67	0.45
US	1.06	0.33	0.95	0.30	26.13	0.32	0.65	0.46

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. T'_{Data} denotes the top statutory personal income tax rate. It is taken from the OECD *Taxing Wages Database* Table 1.7.



(a) Varying the Pareto tail parameter of the talent distribution (a). (b) Varying the scale-of-operations parameter (γ).

Fig. 4. Optimal top tax rates by country.

8. Conclusion

The title of this paper is a reference to the thought-provoking novel of [Rand \(1957\)](#).⁴⁶ 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 in an optimal taxation framework. Within top income earners, we focus on 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-operations 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-operations 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.

Our theoretical results extend to all income earners who have a control over hired inputs that enter in the production (e.g., top lawyers, medical doctors, etc.). Nevertheless, an open question is to quantitatively evaluate the way in which top earners other than managers generate income and how they respond to changes in marginal tax rates. The logical next step involves taking a closer look at the role of managers in 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. An important second extension is to consider a dynamic model of CEOs. Such a model would allow us to analyze the optimal taxation of different forms

⁴⁶ The excellent survey of [Slemrod \(2000\)](#) also features a similar title.

of CEO compensation (long and short term), as well as the interaction between the tax code, managerial compensation, and turnover.

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

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

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

Proof. The first order condition for $\theta' = \theta$ to be a local maximum of $M(\theta', \theta)$ is $M_1(\theta, \theta) = 0$.⁴⁷ This is equivalent to (A.1). Differentiating the first order condition $M_1(\theta, \theta) = 0$ with respect to θ 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 θ' 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:

$$J_2(\theta', \theta) = y'(\theta') [v''(n(\theta'))n_y(\theta')n_\theta(\theta') + v'(n(\theta'))n_{\theta y}(\theta')] + L'(\theta') [v''(n(\theta'))n_L(\theta')n_\theta(\theta') + v'(n(\theta'))n_{\theta L}(\theta')].$$

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 $\theta' \leq \theta$. Equation (5) follows from (A.1), and (6) is obtained by combining (A.1) and (A.2). This completes the proof. \square

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

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

and the budget set

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

Now consider the following claim:

⁴⁷ The subscript $i = 1, 2$ denotes derivative with respect to the first or second argument.

Claim 1. Take a $\theta \in \Theta$ and let $\{c(\theta), y(\theta), L(\theta)\}$ be the optimal allocation assigned to the θ -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 θ , and (2) show that for each θ , $\{c(\theta), y(\theta), L(\theta)\}$ 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

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

which violates incentive compatibility. Given that $\{c(\theta), y(\theta), L(\theta)\} \in B$, for all θ , 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).⁴⁸ Claim 1 then implies that for each of those combinations, $\{c(\theta), y(\theta), L(\theta)\}$ solves the θ -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 ϕ , 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.

B.2. Proof of Proposition 3

We compute efficient allocations by solving the optimal control problem (PP-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 θ , 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 λ^r is the multiplier on (7), λ^l is the multiplier on (8) and $\mu(\theta)$ is the costate on (9) 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

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

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

$$\mu(\theta) = \int_{\theta}^{\bar{\theta}} (\lambda^r - \Psi'(U(\tilde{\theta}))) f(\tilde{\theta}) d\tilde{\theta}. \quad (\text{B.7})$$

Evaluating (B.7) at $\underline{\theta}$ gives

$$\lambda^r = \int_{\Theta} \Psi'(U(\tilde{\theta})) f(\tilde{\theta}) d\tilde{\theta}. \quad (\text{B.8})$$

⁴⁸ For this reason, the levels of T , T_L and w are not determined in the decentralization.

Let $D(\theta) \equiv \frac{1}{1-F(\theta)} \int_{\theta}^{\bar{\theta}} \Psi'(U(\tilde{\theta})) dF(\tilde{\theta})$ so that $D(\underline{\theta}) = \lambda^r$. Substituting into (B.7), we obtain

$$\mu(\theta) = (1 - F(\theta)) (D(\underline{\theta}) - D(\theta)). \quad (\text{B.9})$$

From (B.4)

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

Substituting (B.9) 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.8), 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 (13) yields

$$\frac{T'}{1 - T'} = \frac{1 - F(\theta)}{\theta f(\theta)} \left(1 - \frac{D(\theta)}{D(\underline{\theta})} \right) \left[-\frac{1}{\varepsilon} \frac{n_{\theta}}{n} \theta - \frac{n_{\theta y}}{n_y} \theta \right]. \quad (\text{B.10})$$

Next, we write the partial derivatives of n in (B.10) in terms of partial derivatives of y . Consider the identity $y(n(\bar{y}, L, \theta), L, \theta) = \bar{y}$. Differentiation yields:

$$n_y = \frac{1}{y_n}, \quad (\text{B.11})$$

$$n_L = -\frac{y_L}{y_n}, \quad (\text{B.12})$$

and

$$n_{\theta} = -\frac{y_{\theta}}{y_n}. \quad (\text{B.13})$$

By differentiating both sides of (B.11) with respect to θ , we have $n_{\theta y} = -\frac{y_{n\theta} n_{\theta} + y_n n_{\theta}^2}{y_n^2}$, which implies

$$\begin{aligned} -\frac{n_{\theta y}}{n_y} \theta &= \left(\frac{y_{n\theta} n}{y_n} \right) \left(\frac{n_{\theta}}{n} \theta \right) + \frac{y_n n_{\theta}}{y_n} \theta \\ &= \left[\frac{y_n n_{\theta}}{y_{\theta}} n - \frac{y_{n\theta} n}{y_n} \right] \frac{y_{\theta}}{y_n} \frac{\theta}{n} \\ &= \frac{d \ln \left(\frac{y_{\theta}}{y_n} \right)}{d \ln n} \frac{y_{\theta}}{y_n} \frac{\theta}{n}. \end{aligned} \quad (\text{B.14})$$

Substituting (B.13) and (B.14) into (B.10) gives the result.

B.3. Proof of Corollary 1

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

$$n(y, L, \theta) = \left[\frac{1}{\beta} \left(\frac{y}{\theta g(\theta)} \right)^{\rho} - \frac{1 - \beta}{\beta} \left(\frac{L}{\theta} \right)^{\rho} \right]^{\frac{1}{\rho}}. \quad (\text{B.15})$$

Taking derivatives from (B.15), 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.15), we get

$$\frac{n_{\theta}(y, L, \theta)}{n(y, L, \theta)} = -\frac{1}{\theta} \frac{\left(\frac{y}{g(\theta)} \right)^{\rho} (1 + \gamma) - (1 - \beta) L^{\rho}}{\left(\frac{y}{g(\theta)} \right)^{\rho} - (1 - \beta) L^{\rho}}. \quad (\text{B.16})$$

Also

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

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

The two equations above imply:

$$\frac{n_{\theta y}(y, L, \theta)}{n_y(y, L, \theta)} = -\frac{1}{\theta} \left[(1-\rho) \frac{\left(\frac{y}{g(\theta)} \right)^\rho (1+\gamma) - (1-\beta)L^\rho}{\left(\frac{y}{g(\theta)} \right)^\rho - (1-\beta)L^\rho} + \rho(1+\gamma) \right]. \tag{B.18}$$

By Assumption 2, we have

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

Then

$$\frac{\left(\frac{y}{g(\theta)} \right)^\rho (1+\gamma) - (1-\beta)L^\rho}{\left(\frac{y}{g(\theta)} \right)^\rho - (1-\beta)L^\rho} = 1 + \gamma \left(1 + \frac{1-\beta}{\beta} \left(\frac{L}{\theta n} \right)^\rho \right). \tag{B.19}$$

Also, Assumption 2 implies:

$$\frac{\kappa(\theta)}{1-\kappa(\theta)} = \frac{1-\beta}{\beta} \left(\frac{L}{\theta n} \right)^\rho, \tag{B.20}$$

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

Using (B.20) in (B.19) gives

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

Combining (B.13) with (B.16) and (B.21) gives the expression for $\frac{y_\theta}{y_n} \frac{\theta}{n}$:

$$\frac{y_\theta}{y_n} \frac{\theta}{n} = -\frac{n_\theta}{n} \theta = 1 + \frac{\gamma}{1-\kappa(\theta)}. \tag{B.22}$$

As for the expression for $\frac{d \ln \left(\frac{y_\theta}{y_n} \right)}{d \ln n}$, note that (B.14) implies

$$\frac{d \ln \left(\frac{y_\theta}{y_n} \right)}{d \ln n} = -\frac{n_{\theta y}}{n_y} \theta \left(-\frac{n_\theta}{n} \theta \right)^{-1} \\ = \left[(1-\rho) \left(1 + \frac{\gamma}{1-\kappa(\theta)} \right) + \rho(1+\gamma) \right] \left(1 + \frac{\gamma}{1-\kappa(\theta)} \right)^{-1} \\ = \left[1 + \frac{\gamma}{1-\kappa(\theta)} (1-\rho\kappa(\theta)) \right] \left(1 + \frac{\gamma}{1-\kappa(\theta)} \right)^{-1}, \tag{B.23}$$

where the next to last line follows from (B.18), (B.21), and (B.22). This completes the proof.

B.4. Proof of Proposition 4

Combining (B.4) and (B.5) gives:

$$\frac{n_L}{n_y} = \frac{\lambda^l f - \mu v' n_{\theta L}}{\lambda^r f + \mu v' n_{\theta y}}.$$

Using that $y_L = -n_L/n_y$ ⁴⁹ and applying the manager's first order condition (20) yields:

$$w(1 + T'_L) = \frac{\lambda^l f - \mu v' n_{\theta L}}{\lambda^r f + \mu v' n_{\theta y}}. \quad (\text{B.24})$$

Consider an implementation in which $w = \lambda^l/\lambda^r$.⁵⁰ In that case,

$$1 + T'_L = \frac{1 - \frac{\mu}{w} \frac{\mu}{\lambda^r f} v' n_{\theta L}}{1 + \frac{\mu}{\lambda^r f} v' n_{\theta y}},$$

so that

$$\left(1 + \frac{\mu}{\lambda^r f} \frac{n_{\theta y}}{n_y} v' n_y\right) T'_L = -\frac{\mu}{\lambda^r f} \left(n_{\theta L} \frac{1}{w n_y} + \frac{n_{\theta y}}{n_y}\right) v' n_y. \quad (\text{B.25})$$

By the first order conditions (13) and (20), it follows that

$$v' n_y = 1 - T', \quad \text{and} \quad \frac{1}{w n_y} = -\frac{1 + T'_L}{n_L}, \quad (\text{B.26})$$

where we used that $y_L = -n_L/n_y$.

Substituting (B.26) into (B.25) produces

$$\left(1 + \frac{\mu}{\lambda^r f} \frac{n_{\theta y}}{n_y} (1 - T')\right) T'_L = -\frac{\mu}{\lambda^r f} \left(-\frac{n_{\theta L}}{n_L} (1 - T'_L) + \frac{n_{\theta y}}{n_y}\right) (1 - T').$$

This expression can be rearranged as

$$\frac{T'_L}{(1 + T'_L)(1 - T')} = \frac{\mu}{\lambda^r \theta f} \left[\frac{n_{\theta L}}{n_L} \theta - \frac{n_{\theta y}}{n_y} \theta \right]. \quad (\text{B.27})$$

We now operate on the square brackets of the right hand side. Using (B.12) we get

$$\begin{aligned} \frac{n_{\theta L}}{n_L} \theta &= \left(\frac{y_{\theta}}{y_L} \frac{\theta}{L}\right) \left(\frac{y_{\theta L}}{y_{\theta}} L - \frac{y_{nL}}{y_n} L\right) - \left(\frac{y_{\theta}}{y_n} \frac{\theta}{n}\right) \left(\frac{y_{\theta n}}{y_{\theta}} n - \frac{y_{nn}}{y_n} n\right) \\ &= \left(\frac{y_{\theta}}{y_L} \frac{\theta}{L}\right) \left(\frac{y_{\theta L}}{y_{\theta}} L - \frac{y_{nL}}{y_n} L\right) + \frac{n_{\theta y}}{n_y} \theta, \end{aligned} \quad (\text{B.28})$$

where the last equality follows from (B.14).

Plugging in (B.28) into (B.27) gives

$$\begin{aligned} \frac{T'_L}{(1 + T'_L)(1 - T')} &= \frac{\mu}{\lambda^r \theta f} \left(\frac{y_{\theta}}{y_L} \frac{\theta}{L}\right) \left[\frac{y_{\theta L}}{y_{\theta}} L - \frac{y_{nL}}{y_n} L\right] \\ &= \frac{\mu}{\lambda^r \theta f} \left(\frac{y_{\theta}}{y_L} \frac{\theta}{L}\right) \frac{d \ln\left(\frac{y_{\theta}}{y_n}\right)}{d \ln L}. \end{aligned} \quad (\text{B.29})$$

Finally, substituting $\frac{\mu}{\lambda^r \theta f} = \frac{1-F}{\theta f} \left(1 - \frac{D}{D}\right)$, as per Proposition 3, proves the result.

⁴⁹ See (B.11) and (B.12).

⁵⁰ By Proposition 2, it is always possible to construct such an implementation, as long as the tax system $\{T, T_L, \phi\}$ satisfies (B.3) and $\phi = c^w - w$.

B.5. Proof of Corollary 3

The proof goes along the lines of the proof of Corollary 1. In particular, by Assumption 2, it follows that

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

where $g(\theta) = \theta^\gamma$. By (B.21):

$$\begin{aligned} \frac{n_{\theta L}}{n_L} \theta &= -(1 - \rho) \left(1 + \gamma \left(1 + \frac{\kappa(\theta)}{1 - \kappa(\theta)} \right) \right) - \rho \\ &= -1 - (1 - \rho) \frac{\gamma}{1 - \kappa(\theta)}. \end{aligned} \quad (\text{B.30})$$

On the other hand, from the proof of Corollary 1, we know that

$$-\frac{n_{\theta y}}{n_y} \theta = 1 + (1 - \rho \kappa(\theta)) \frac{\gamma}{1 - \kappa(\theta)}. \quad (\text{B.31})$$

Combining (B.30) and (B.31)

$$\frac{n_{\theta L}}{n_L} \theta - \frac{n_{\theta y}}{n_y} \theta = \gamma \rho.$$

The result then follows by substituting this expression and $\frac{\mu}{x^{\rho} \theta^{\rho}} = \frac{1-F}{\theta^{\rho}} \left(1 - \frac{D}{D} \right)$ into (B.27).

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. Denote the elasticity of substitution between θn and L with $\sigma \equiv -\frac{d \ln(L/\theta n)}{d \ln(f_2/f_1)}$. We have:

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

Proof. From the definition of κ , we can write

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

Let $f(\theta n, L) = Lh(\theta n/L)$. Denote the elasticity of substitution between θn and L with $\sigma \equiv -\frac{d \ln(L/\theta n)}{d \ln(f_2/f_1)}$. 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. For notational convenience, let $g = \theta^\gamma$ and $g' = \gamma \theta^{\gamma-1}$.

Proof. We begin by determining the relationship between the manager labor supply decision and his type. Differentiating (24) and (25) with respect to θ we get, respectively,

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

and

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

Combining (C.3) and (C.4), it follows that

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

where we applied (C.1), (C.2), and the definition of g . Further rearranging and using (C.1) gives:

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

Finally, substituting the first order condition (25), 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 (26).

Now we obtain equation (27). 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 (26) and (C.6) into the above expression gives (27).

Finally, we derive equation (28). Profits are given by $\pi = y - wL$. Then

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

or

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

where we used that $wL/y = y_L L/y = \kappa$. Substituting (26) and (27) into (C.7) and rearranging gives (28). \square

C.2. Proof of Proposition 5

Proof. The following relationships are derived by combining (26)–(28):

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

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

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

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

Also, by 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) gives equation (29). \square

D. Firm distortions and tax elasticities

In this appendix we show that in the absence of firm distortions, the elasticity of taxable income of the manager equals the Frisch elasticity of labor supply ε .

We first show that, at the optimum, it is possible to write the income of managers of talent θ as $\pi(\theta) = \omega(\theta, w)n$ where $\omega(\theta, w)$ is the wage of managers of talent θ 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 θ and w , the relationship between θn 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 θ -manager as:

$$\max_{c, n} c - v(n) \quad \text{s.t.} \quad c = (1 - \tau)\omega(\theta, w)n.$$

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

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

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

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

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

E. Using the distribution of income to calculate a

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 (28) 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 (31), 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_π . We then have $\frac{1}{a_\pi} = \frac{1}{N} \sum_{i=1}^N \ln \pi(\theta^i)$. Substituting in the above, we get:

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

F. Additional robustness checks

In this appendix, we expand on the results of Section 6 by providing additional robustness checks. We perform sensitivity analyses with respect to two empirical moments: the elasticity of firm size with respect to sales $d \ln L(\theta)/d \ln y(\theta)$ and the Pareto tail parameter of the firm size distribution a_L . These moments affect the calibration of the parameters a , γ and κ .

To make the relationship between the top tax formula and $d \ln L(\theta)/d \ln y(\theta)$ and a_L more transparent, we rewrite (19) by substituting the formula for κ in (C.10), γ in (29), and a in (34). We get:

Table 5
Top tax rate: robustness.

$\frac{d \ln L(\bar{\theta})}{d \ln y(\bar{\theta})}$	$a_L = 0.8$		$a_L = 1.5$	
	1.18	0.95	1.18	0.95
$T'(\bar{\theta})$	36.8%	41.8%	23.7%	27.7%

Notes: $T'(\bar{\theta})$ denotes optimal top tax rate derived from (F.1). The range of values for $\frac{d \ln L(\bar{\theta})}{d \ln y(\bar{\theta})}$ is imputed from the observed ranges reported in Fig. 1.

Table 6
Optimal marginal tax on firm size $T'_L(wL(\theta))$.

ξ	$\bar{L}/(\bar{\theta}\bar{n})$	Firm size percentile				
		25th	50th	75th	90th	99th
0.5	2.9	2.12%	2.26%	2.34%	2.34%	2.21%
0.5	3.43	2.19%	2.33%	2.39%	2.38%	2.23%
0.7	2.9	1.68%	1.81%	1.88%	1.90%	1.84%

Notes: Firm size distortions $T'_L(wL(\theta))$ defined from equation (20).

$$T'(\pi^*(\bar{\theta})) = \frac{1}{1 + a_L \frac{\frac{d \ln L(\bar{\theta})}{d \ln y(\bar{\theta})}}{\frac{d \ln \pi(\bar{\theta})}{d \ln y(\bar{\theta})} \left(\frac{1}{\sigma} + \frac{1}{\varepsilon} \right) - \frac{1}{\sigma}}}. \quad (\text{F.1})$$

In Table 5, we compute the top marginal tax rate over a range of a_L and $\frac{d \ln L(\bar{\theta})}{d \ln y(\bar{\theta})}$.

In all cases considered the main result presented in Section 6 is robust: Optimal tax rates are significantly lower than in the case without a scale-of-operations effect.

G. Firm size taxation: a numerical illustration

As emphasized in Section 4.3, 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 (20)). In this section, we characterize optimal taxes on firm size in a calibrated example. Unlike in the body of the paper where we focused on taxation at the top, here we compute firm size taxes over the *entire* talent distribution (this allows us to study progressivity as well).

We focus on US data and take the values for the parameters a , ε , σ , and γ from Table 2. In addition, computing firm size taxes over the entire distribution of skills requires the calibration of three additional objects: the distribution of talent, the production share parameter β , and the social welfare function. We assume that the talent distribution is Pareto-Lognormal⁵¹ 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 β , we use the definition of κ which implies:

$$\kappa(\bar{\theta}) = \frac{y_L(\bar{\theta})L(\bar{\theta})}{y(\bar{\theta})} = \frac{1}{1 + \frac{\beta}{1-\beta} \left(\frac{L(\bar{\theta})}{\bar{\theta}n(\bar{\theta})} \right)^{-\rho}}. \quad (\text{G.1})$$

We proxy the ratio $L(\bar{\theta})/(\bar{\theta}n(\bar{\theta}))$ with the number of production workers per non-production workers in the NBER–CES Manufacturing Industry Database. Taking an average across all industries we find a value of 2.9 for 2011 and 3.43 for the entire sample. For robustness purposes we use both values. Combining the latter with our calibration of $\bar{\kappa} = 0.93$, (G.1) implies $\beta \in [0.15, 0.17]$. The social welfare function is parameterized as $\Psi(U) = U^\xi$. We set $\xi = 0.5$ as a benchmark, and consider the effect of increasing ξ as well. Finally, government expenses are set to $G = 0$.

Table 6, 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 2.26 percent above the wage rate in the baseline specification. Second, the marginal tax on labor use decreases slightly at the top, and asymptotes at around 2 percent across all specifications.

References

- Adams, R.B., Almeida, H., Ferreira, D., 2005. Powerful CEOs and their impact on corporate performance. *The Review of Financial Studies* 18 (4), 1403–1432.
 Albanesi, S., 2011. Optimal Taxation of Entrepreneurial Capital with Private Information. Working Paper.
 Alder, S., 2016. In the wrong hands: complementarities, resource allocation, and aggregate TFP. *American Economic Journal: Macroeconomics* 8 (1), 199–241.

⁵¹ That is, $\theta \stackrel{d}{=} \theta_1 \theta_2$, where $\theta_1 \sim \text{LN}(\zeta, \iota^2)$ and $\theta_2 \sim P(a)$.

- Ales, L., Kurnaz, M., Sleet, C., 2015. Technical change, wage inequality, and taxes. *The American Economic Review* 105 (10), 3061–3101.
- Ales, L., Sleet, C., 2016. Taxing top CEO incomes. *The American Economic Review* 106 (11), 3331–3366.
- 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., Hall, B.J., 2004. CEO incentives and firm size. *Journal of Labor Economics* 22 (4), 767–798.
- Baker, G.P., Jensen, M.C., Murphy, K.J., 1988. Compensation and incentives: practice vs. theory. *Journal of Finance* 43 (3), 593–616.
- Bakija, J., Cole, A., Heim, B., 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., MacDonald, G., Yang, J., 2011. The economics of super managers. *The Review of Financial Studies* 24 (10), 3321–3368.
- Bartelsman, E.J., Doms, M., 2000. Understanding productivity: lessons from longitudinal microdata. *Journal of Economic Literature* 38 (3), 569–594.
- Bennedsen, M., Nielsen, K.M., Perez-Gonzalez, F., Wolfenzon, D., 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., Schoar, A., 2003. Managing with style: the effect of managers on firm policies. *The Quarterly Journal of Economics* 118 (4), 1169–1208.
- Blundell, R., MaCurdy, T., 1999. Labor supply: a review and alternative approaches. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. In: *Handbook of Labor Economics*, vol. 3. Elsevier, pp. 1559–1695. Chapter 27.
- Blundell, R., Pistaferri, L., Saporta-Eksten, I., 2016. Consumption inequality and family labor supply. *The American Economic Review* 106 (2), 387–435.
- Brüggemann, B., 2016. Higher Taxes at the Top: The Role of Entrepreneurs. Working Paper.
- Cagetti, M., Nardi, M.D., 2006. Entrepreneurship, frictions, and wealth. *Journal of Political Economy* 114 (5), 835–870.
- Chetty, R., Guren, A., Manoli, D., Weber, A., 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., Ferreira, M.A., Matos, P., 2013. Generalists versus specialists: lifetime work experience and chief executive officer pay. *Journal of Financial Economics* 108 (2), 471–492.
- Custódio, C., Metzger, D., 2013. How do CEOs matter? The effect of industry expertise on acquisition returns. *The Review of Financial Studies* 26 (8), 2008–2047.
- Diamond, P., 1998. Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *The American Economic Review* 88 (11), 83–95.
- Diamond, P.A., Mirrlees, J.A., 1971. Optimal taxation and public production. I: production efficiency. *The American Economic Review* 61 (1), 8–27.
- Diamond, P.A., Saez, E., 2011. The case for a progressive tax: from basic research to policy recommendations. *The Journal of Economic Perspectives* 25 (4), 165–190.
- di Giovanni, J., Levchenko, A., 2013. Firm entry, trade, and welfare in Zipf's world. *Journal of International Economics* 89 (2), 283–296.
- Edmans, A., Gabaix, X., 2011. The effect of risk on the CEO market. *The Review of Financial Studies* 24 (8), 2822–2863.
- Edmans, A., Gabaix, X., Landier, A., 2009. A multiplicative model of optimal CEO incentives in market equilibrium. *The Review of Financial Studies* 22 (12), 4881–4917.
- Eisfeldt, A.L., Kuhn, C.M., 2013. CEO turnover in a competitive assignment framework. *Journal of Financial Economics* 109 (2), 351–372.
- Farhi, E., Werning, I., 2013. Insurance and taxation over the life cycle. *The Review of Economic Studies* 80 (2), 596–635.
- Fernandes, N., Ferreira, M.A., Matos, P., Murphy, K.J., 2013. Are U.S. CEOs paid more? New international evidence. *The Review of Financial Studies* 26 (2), 323–367.
- French, E., 2005. The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies* 72 (2), 395–427.
- Frydman, C., Saks, R.E., 2010. Executive compensation: a new view from a long-term perspective, 1936–2005. *The Review of Financial Studies* 23 (5), 2099–2138.
- Gabaix, X., Landier, A., 2008. Why has CEO pay increased so much? *The Quarterly Journal of Economics* 123 (1), 49–100.
- Golosov, M., Troshkin, M., Tsyvinski, A., 2016. Redistribution and social insurance. *The American Economic Review* 106 (2), 359–386.
- Gouveia, M., Strauss, R., 1994. Effective individual income tax functions: an exploratory empirical analysis. *National Tax Journal* 47 (2), 317–339.
- Guesnerie, R., 1981. On Taxation and Incentives: Further Reflections on the Limits to Redistribution. Discussion Paper No. 89, Sonderforschungsbereich 21, University of Bonn.
- Huggett, M., Badel, A., 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., Klebanov, M.M., Sorensen, M., 2012. Which CEO characteristics and abilities matter? *Journal of Finance* 67 (3), 973–1007.
- Kaplan, S.N., Rauh, J., 2013. It's the market: the broad-based rise in the return to top talent. *The Journal of Economic Perspectives* 27 (3), 35–56.
- Lazear, E.P., Shaw, K.L., Stanton, C., 2016. Making do with less: working harder during recessions. *Journal of Labor Economics* 34 (1), 333–360.
- Lewellen, W.G., Huntsman, B., 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. *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. *The Journal of Economic Perspectives* 27 (3), 21–34.
- Mankiw, G., Weinzierl, M., Yagan, D., 2009. Optimal taxation in theory and practice. *The Journal of Economic Perspectives* 23 (4), 147–174.
- Mayer, T., 1960. The distributions of ability and earnings. *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., Saez, E., 2003. Income inequality in the United States, 1913–1998. *The Quarterly Journal of Economics* 118 (1), 1–39.
- Piketty, T., Saez, E., Stantcheva, S., 2014. Optimal taxation of top labor incomes: a tale of three elasticities. *American Economic Journal: Economic Policy* 6 (1), 230–271.
- Quadrini, V., 2000. Entrepreneurship, saving and social mobility. *Review of Economic Dynamics* 3 (1), 1–40.
- Rand, A., 1957. *Atlas Shrugged*. Signet, New York.
- 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. *Bell Journal of Economics* 13 (2), 311–323.
- Rothschild, C., Scheuer, F., 2013. Redistributive taxation in the Roy model. *The Quarterly Journal of Economics* 128 (2), 623–668.
- Rothschild, C., Scheuer, F., 2016. Optimal taxation with rent-seeking. *The Review of Economic Studies* 83 (3), 1225–1262.
- 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., Slemrod, J., Giertz, S.H., 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., Stantcheva, S., 2016. Generalized social welfare weights for optimal tax theory. *The American Economic Review* 106 (1), 24–45.

- Scheuer, F., 2014. Entrepreneurial taxation with endogenous entry. *American Economic Journal: Economic Policy* 6 (2), 126–163.
- Scheuer, F., Werning, I., 2017. The taxation of superstars. *The Quarterly Journal of Economics* 132 (1), 211–270.
- Shourideh, A., 2012. Optimal Taxation of Entrepreneurial Income: a Mirrleesian Approach to Capital Accumulation. Working Paper.
- Slavík, C., Yazıcı, H., 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. *The 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.