Taxing Atlas:
Executive Compensation, Firm Size and Their Impact
on Optimal Top Income Tax Rates*

Laurence Ales1, Antonio Andrés Bellofatto2, and Jessie Jiaxu Wang3

1Tepper School of Business, Carnegie Mellon University
2University of Queensland
3W. P. Carey School of Business, Arizona State University

This Version: August 3, 2015
First Version: June 10, 2014

Abstract

This paper studies 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 in their talent. Effort and talent of the manager are privately observed. Managerial talent increases managers’ productivity of effort and overall firm productivity, creating a scale-of-operations effect. A tax formula for optimal taxes is derived linking optimal marginal tax rates to preferences and technology. The model is calibrated using US firm level data. Our quantitative results suggest that optimal top tax rates are in line with the current US tax code and significantly lower than previous studies ignoring the scale-of-operations effect have shown.

*We thank our discussant Adriano Rampini for valuable comments and suggestions. We also thank Emilio Bisetti, Miguel Ferreira, Brent Glover, Burton Hollifield, Finn Kydland, Christos Makridis, Pricila Maziero, Tomasz Piskorski, Florian Scheuer, Ali Shourideh, Chris Sleet, Stephen Spear, Chris Telmer, Sevin Yeltekin, and Ariel Zetlin-Jones, as well as participants at Carnegie Mellon Ph.D. Macro-Finance lunch, Econometric Society summer meetings in Minneapolis, Society for Economic Dynamics annual meeting in Toronto, and the 5th Tepper-LAEF conference for feedback. E-mail: ales@cmu.edu (Ales); andresbellofatto@gmail.com (Bellofatto); Jessie.Jiaxu.Wang@asu.edu (Wang)
1 INTRODUCTION

Income taxation of top income earners is a controversial and recurrent topic in the tax policy debate.\textsuperscript{1} However, the large literature in public finance is far from reaching a consensus on what the top tax rate should be. The range of proposed tax rates for top income earners is surprisingly large, ranging from zero (much lower than the current rate) to 80 percent (much higher than the current rate). This lack of consensus can in part be attributed to the lack of agreement on two attributes of the highly talented individuals: the magnitude of their behavioral response to tax and their prevalence in the population. This paper focuses on managers.\textsuperscript{2} By doing so we show how to use well established facts on the joint distribution of firm size and managerial reward to pin down key forces that shape optimal top tax rates. We find that optimal marginal tax rates for high income earners are in line with what we see in the US today.

In our environment managers are heterogeneous and operate a span of control technology as in Rosen (1982). Differently than Rosen, managers exert elastic effort. The contribution of the manager to the production of output is twofold: First the manager provides effective labor that interacts with hired labor (what Rosen refers to as the supervision of workers). Effective labor is the combination of talent of the manager and his elastically provided labor. This assumption implies that a more talented manager can transform hours worked into effective labor at a higher rate. Second, the manager provides what Rosen refers to as his pure management ability. This ability affects the overall productivity of the firm, creating a scale-of-operations effect (see Mayer (1960)). We assume a positive relationship between labor productivity and managerial ability. Modeling talent of the manager in this fashion has three important implications. First, more talented managers operate larger firms and receive higher levels of pre-tax income. Second, as originally shown by Rosen (1982), the scale-of-operations effect implies that managerial compensation grows at a faster rate than managerial talent. Thus, the distribution of income becomes more positively skewed relative to the distribution of talent. This has profound implications on the optimal tax problem as it impacts the inferred prevalence of high talent in the population. Third, wages of managers are endogenous; in particular, they depend on the amount of hired labor and the amount of effort exercised by managers. This aspect is also important since any tax aimed at top income earners (or any distortionary firm tax) indirectly impacts their pre-tax wages and hence their behavioral responses.

Following the standard approach of public finance literature, we assume that full redistribution of income is hindered by informational frictions. Specifically, we consider the case in which the effort and talent of the manager are private information. Firm size and firm output are

\textsuperscript{1}This is partially due to heightened concerns over the increasing inequality of the income distribution. Share of income going to the top 1\% increased from 9\% in 1970 to 23.5\% in 2007 (Diamond and Saez (2011)). From Piketty and Saez (2003), 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.

\textsuperscript{2}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\%. 
instead assumed to be observable. We characterize the (ex-ante) constrained efficient allocation and show that it can be decentralized as a competitive equilibrium with taxes levied on income and firm size. The former tax is standard in the public finance literature, the latter it is not. Firm size is measured as the size of the hired workforce. Any positive tax levied conditional on this quantity distorts firms’ decisions, forcing them to operate at a below optimal scale.

Our first result concerns firm size distortions. We show how, in general, the planner forgoes efficiency in the allocation of labor in order to relax incentive constraints and hence reduce informational rents of the managers. We also highlight that these firm size distortions do not arise absent a positive scale-of-operations effect. In our benchmark, featuring a positive scale-of-operations effect and an elasticity of substitution between hired labor and managerial effort different than one, firm size distortions do arise.

We then look at optimal income taxation. We start the analysis by providing a formula for optimal marginal tax rates. This formula links marginal tax rates to primitives of the environment. As in Saez (2001) the formula links marginal rates to the assumed distribution of talent, redistributive motives of the policy maker, and elasticity of labor supply of the manager. In addition, the span of control aspect of the production function discussed earlier introduces two novel terms. The first one looks at the impact of managerial effort on his own compensation. This term lowers the behavioral response of managers. The second term is related to how effort and marginal product grow as managerial talent grows. This term disappears once we consider a standard production function without the scale-of-operations effect. In general this is (for a given distribution of talent) a force for higher marginal tax rates.

Finally, we take the model to the data to evaluate the forces at play in the environment and to provide specific normative recommendations on optimal marginal tax rates. Calibration of the environment requires a successful identification of two primitive objects. The first object is the parameter governing the scale-of-operations effect. The second object is the distribution of managerial talent. To identify both of these objects we follow the key insights from Rosen (1982). In particular, we expand his result (adding elastic labor) showing how the scale-of-operations effect can be backed out from the data with knowledge of the elasticity of firm size (in terms of employment) to firm sales and the elasticity of managerial compensation to firm sales. We determine both of these elasticities using COMPSTAT data. The second object is the distribution of talent. Here the established public finance approach is to invert the distribution of income to uncover the distribution of talent. A similar route is possible in our environment with one key difference: our environment generates a nonlinear relationship between talent and income and between talent and firm size. This implies that as the talent of the manager increases by 1 percent, firm size and managerial income increase substantially more than that (in a standard Mirrleesian environment, keeping effort fixed, income would increase by only 1 percent). In our benchmark calibration, we instead use the distribution of firm size to invert the distribution of talent. We find that, in the presence of a positive scale-of-operations effect, the talent distribution is substantially more compact than previously assumed in the public finance literature: assuming a Pareto distribution our estimate on the tail parameter is an order of magnitude larger than what was previously identified in the literature. This is
important since the more compact the talent distribution is, the stronger the force for lower marginal rates is.

The optimal top tax rate in our benchmark calibration is 32.4 percent. To gain perspective on this number it is useful to compare it to established benchmarks in the public finance literature: at one extreme Mirrlees (1971) and Seade (1977) prescribe a zero marginal tax rate at the top; at the opposite end, using different assumptions on the distribution of talent, Saez (2001) and Diamond and Saez (2011) call for for top rates as high as 80 percent. Overall our prescribed optimal top tax rate is in the same range as what the current US tax code prescribes. The assumed production function is key in generating this result. For comparison, with the same calibration but absent a scale-of-operations effect, the optimal rate is equal to 65.4 percent. We then quantify the optimal marginal tax on firm size. We find it to be positive, progressive, and creating a markup of about 2 percent over the real wage at the top. We conclude our quantitative analysis by looking beyond the US. We extend our analysis to a panel of European countries. For these countries we also find a significant role for the scale-of-operations. Overall we find that optimal tax rate are lowered by 27 percentage points on average with respect to the case without a span of control production function.

To summarize, this paper makes three contributions. First, we provide normative grounds for why (and when) it is optimal to distort firm sizes in order to reduce informational rents of managers. Second, we derive a tax formula in a span of control environment as in Rosen (1982) and show how the distribution of talent at the top is significantly more compact than the distribution of income. Third, we illustrate how the determinants of such a tax formula can be calibrated using readily available firm level data.

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.

The works of 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 together with the size of the firm. In these models the manager chooses the factors of production to be purchased from the market. Terviö (2008) and Gabaix and Landier (2008), on the other hand, consider a model where the size of the firm is fixed exogenously. Here the most productive managers are assigned to the largest firms. What is key in all of these models (for the purpose of optimal taxation) is that they introduce a nonlinear mapping between compensation and talent of the manager. In particular the distribution of compensation is more positively skewed than the one for talent. Our contribution is to model the intensive margin of managerial effort (this is a necessary step in order to think about income taxation).

The literature on optimal taxation of top income earners is vast. Methodologically, our contribution with respect to this literature is twofold. To the best of our knowledge this is the

---

3Saez et al. (2012) report a top 1 percent marginal rate of approximately 42.5 percent for 2009. In the Current Population Survey in the same period we find top marginal income tax rates of 33.5% for federal and 5% for state. 4For a review refer to Mankiw et al. (2009) and Diamond and Saez (2011).
first paper that explores the distribution of underlying worker talent resulting from non-linear productions functions and its implication for taxation. In addition, our environment is one in which compensation of the agent (the manager in our case) is endogenous. This is a departure from the classical taxation environment where wages are fixed exogenously. A few exceptions to this assumption are discussed in Stiglitz (1982) where workers of different types interact within an aggregate production function, hence influencing each other’s wages. In more recent work, Slavík and Yazıcı (2014) focus on the endogenous accumulation of heterogenous forms of capital that interact differently with agents of heterogenous talent. Ales et al. (2014) focus on an assignment problem of workers with heterogenous talents to tasks with heterogenous complexity.

Our approach in this paper is to map top income earners to managers. Given this, Rothschild and Scheuer (2013) and in particular Scheuer (2014) are also related to the current work. They consider an environment where agents are characterized by a multidimensional skill/taste vector and decide whether to be a worker or a manager. In the spirit of Stiglitz (1982), a key channel that impacts optimal taxes is the spillover effect between wages of workers and wages of managers. Relative to these papers our emphasis is more quantitative. However, there are key differences in the environment that shape optimal taxation. The most important is the role of managerial talent. In our environment, as in these papers, managerial talent impacts how productive the manager is at transforming hours in effective labor that enters in the production function. Differently from Rothschild and Scheuer (2013) and Scheuer (2014), though, talent also affects the overall productivity of the production function. This in turn increases productivity of the entire firm. We show that as this second component disappears, optimal taxes revert to the Mirrleesian benchmark as in Scheuer (2014) or Saez (2001). Additionally, Piketty et al. (2014) consider a model of CEO taxation where, beyond the behavioral labor supply response, the CEO can extract surplus by imposing a negative externality on workers raising his own compensation above his marginal product. In this case, this last channel provides upward pressure on marginal tax rates as the taxes try to correct for this negative CEO externality. In our paper the approach is more in line with the empirical evidence in Kaplan and Rauh (2013) where managers are characterized as having a positive externality on workers rather than a negative one.

This paper is related to the recent papers of Ales and Sleet (2015) and Scheuer and Werning (2015). Both these paper study the optimal managerial/superstar taxation problem in competitive assignment environments. In these papers top income earners of heterogenous abilities are matched to heterogenous fixed factors (firm assets) as in Terviö (2008) and Gabaix and Landier (2008). In the present paper, firm size is endogenous and is affected by managerial ability as well as the tax code. In Ales and Sleet (2015) and Scheuer and Werning (2015) firm size is instead exogenous. One way to interpret this difference is to view Ales and Sleet (2015) and Scheuer and Werning (2015) as studying short run taxation implications where firm sizes do not adjust following changes in the taxation of managers. Instead the present paper provides a

5The environment in this paper is static. For dynamic models that consider the modeling and taxation of entrepreneurial wealth, refer to Quadrini (2000), Cagetti and Nardi (2006), Albanesi (2011) or Shourideh (2012).
framework for the design of taxation in the long run (where firm size is allowed to adjust to changes in the tax code).

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

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 $n$. Preferences 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.

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

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

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

Managerial talent enters the production function (1) in two ways. First, $\theta$ is effort-augmenting, as it multiplies $n$ within $H$. Second, $\theta$ acts as a positive total factor productivity since $\gamma > 0$. We refer to the latter effect $\gamma$ as the scale-of-operations effect, following Mayer (1960).\(^6\) This formulation is in line with the one in Rosen (1982), where managers’ actions naturally affect the productivity of all workers under their supervision (irrespective of their number). But unlike the technology in Rosen’s paper, we also incorporate elastic managerial effort $n$ as an intensive margin.

\(^6\)See Bartelsman and Doms (2000) and references therein for additional details on the relationship between managerial talent, firm size and firm productivity.
In what follows, it is convenient to define $n(y, L, \theta)$ as the effort required by a manager of talent $\theta$ to generate output $y$ when hired labor is $L$. An allocation in this economy is then defined as $(c^w, c, y, L)$, where $c^w \in \mathbb{R}_+$, $c : \Theta \to \mathbb{R}_+$, $L : \Theta \to \mathbb{R}_+$, $y : \Theta \to [0, \bar{y}]$, and $0 < \bar{y} < \infty$. We assume that $c(\theta)$, $y(\theta)$ and $L(\theta)$ are observable, while $\theta$, $n(\theta)$ and hence $\theta \cdot n(\theta)$ are private information to each $\theta$-agent.

For a given level of (exogenous) government consumption $G > 0$, an allocation is feasible if
\[ c^w + \int_{\Theta} c(\theta)dF(\theta) + G \leq \int_{\Theta} y(\theta)dF(\theta), \tag{2} \]
and
\[ \int_{\Theta} L(\theta)dF(\theta) \leq 1. \tag{3} \]
Social welfare is evaluated according to the social welfare function
\[ SWF = \Psi(c^w) + \int_{\Theta} \Psi(c(\theta) - v(n(y(\theta), L(\theta), \theta)))dF(\theta), \]
where $\Psi : \mathbb{R} \to \mathbb{R}$ is a strictly increasing, differentiable and concave function which summarizes social preferences for redistribution. In particular, we refer to $\Psi'(c(\theta) - v(n(y(\theta), L(\theta), \theta)))$ as the social marginal welfare weight on managers of talent $\theta$.

3 PARETO OPTIMALITY

In this section, we characterize Pareto optimal allocations using a direct mechanism where managers report their talent $\theta$ to a social planner and are assigned an allocation for consumption $c(\theta)$, output $y(\theta)$ and labor $L(\theta)$ accordingly. Define $n(y(\theta'), L(\theta'), \theta)$ as the level of effort exerted by a manager of talent $\theta$ who mimics a manager of talent $\theta'$. In this case, manager $\theta$ is assigned $L(\theta')$ workers and is required to produce output $y(\theta')$. An allocation is incentive-compatible when truthful revelation is optimal for all managers, which requires:
\[ c(\theta) - v(n(y(\theta), L(\theta), \theta)) \geq c(\theta') - v(n(y(\theta'), L(\theta'), \theta)), \quad \forall \theta, \theta' \in \Theta. \tag{4} \]

Pareto optimal allocations solve the following social planner’s problem:
\[
\max_{c^w, \{c(\theta), y(\theta), L(\theta)\}_{\theta \in \Theta}} \Psi(c^w) + \int_{\Theta} \Psi(c(\theta) - v(n(y(\theta), L(\theta), \theta)))dF(\theta), \tag{PO}
\]
s.t. (2), (3) and (4).

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

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

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

and

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

(5)

Proof. See Appendix A.

As it is standard in the optimal taxation literature, from here onwards we assume that
the monotonicity condition (6) is satisfied at the optimum. The condition holds, for example,
when $c, y$ and $L$ are increasing in $\theta$ and $n_{\theta L}$ is small enough. 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:

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

(PO-FOC)

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

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

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

(9)

Before presenting the decentralization and showing the properties of the optimal tax system,
we discuss the incentive constraint (5) further. The standard Mirrleesian environment features
the following relationship between output, talent and effort: $y(\theta) = \theta \cdot n(\theta)$. In this case effort
required by $\theta$ to generate output $y$ is $n(y, \theta) = y/\theta$. So that the term $n_\theta$ appearing in the right
hand side of (5) is given by $n_{\theta y}(y, \theta) = -y/\theta^2 = -n(\theta)/\theta$. In this case, the incentive constraint only
depends on the level of effort and does not depend separately on either output or employment.
Instead, our environment features a (potential) nonlinear relationship between output,
talent, and effort. It also features an additional input to production: labor ($L$).

We next provide additional insights on the incentive constraint for this case. Suppose as in
Rosen (1982) we assume that $H$ satisfies constant returns to scale. Let $h(x) = H(x, 1)$ for all
$x \geq 0$. Since $H$ is constant return to scale we have that by definition $y = \theta^\gamma \cdot L \cdot h\left(\frac{\theta n(y, L, \theta)}{L}\right),$ so

\[7\] 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. (2013), and Farhi and Werning (2013) apply similar
techniques in dynamic environments.

\[8\] It is straightforward to verify that $n_{\theta y}, n_{\theta y} \leq 0$ and $n_{\theta L} \geq 0$.

\[9\] 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 $L$. 


that \( n(y, L, \theta) = h^{-1} \left( \frac{y}{\theta L} \right) \frac{L}{\theta} \). In this case we have

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

which simplifies to

\[
n_\theta(y, L, \theta) = -n(\theta) \frac{\theta - \gamma - \gamma H(\theta n, L)}{\theta}.
\]

The first term in (10) is the same term that appears in the Mirrleesian environment discussed above. Indeed, in the case in which \( \gamma = 0 \) it is the only term appearing. Hence, a modification of the benchmark Mirrleesian environment with a nonlinear production function featuring \( L \) with \( \gamma = 0 \), would leave the incentive constraint unchanged. The second term in (10) is novel in this environment. A notable feature of the environment here considered is that in general \( n_\theta \) and hence the right hand side of the incentive constraint (5) depends explicitly on \( L \). This implies that a particular choice of \( L \) has the effect of tightening or relaxing the incentive constraint. This is at the cost of production efficiency. The distorted choice of \( L \) in the sections below is implemented with a nonlinear tax on firm size. The ability of the planner to affect incentives by changing \( L \) depends on the particular functional form of \( H \). Suppose that \( H \) is Cobb-Douglas so that:

\[
y = \theta^\gamma (\theta \cdot n)^\alpha L^{1 - \alpha}
\]

for some \( 0 < \alpha < 1 \). We have:

\[
n_\theta(y, L, \theta) = -n(\theta) \frac{\theta - \gamma - \gamma (\theta \cdot n)^\alpha L^{1 - \alpha}}{\theta^2 \alpha (\theta \cdot n)^{\alpha - 1} L^{1 - \alpha}} = -n(\theta) \frac{1 + \gamma}{\alpha},
\]

so that in this case \( n_\theta \) assumes the same form as the standard Mirrleesian case. Below we show that the two cases discussed with \( n_\theta \) independent of \( L \) provide a similar tax recommendation as a standard Mirrleesian case. We instead focus on the a more general CES formulation for \( H \). This production function, differently than the Cobb-Douglas case features \( n_\theta \) depending directly on \( L \) and introduces a motive for the planner to distort the allocation of labor.\(^\text{10}\)

### 4 Optimal Taxation

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

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

\[
\max_{c, y, L} c - v(n(y, L, \theta)) \quad \text{(MP)}
\]

\[
s.t. \quad c \leq y - wL - T_L(wL) - T(y - wL - T_L(wL)),
\]

where \( w \in \mathbb{R}_+ \) is the real wage, \( T : \mathbb{R}_+ \rightarrow \mathbb{R} \) is a nonlinear income tax, and \( T_L : \mathbb{R}_+ \rightarrow \mathbb{R} \) is a nonlinear tax on firm size. Since workers in our environment supply labor inelastically, their problem is characterized by a simple budget constraint \( c^w = w + \phi \), where \( \phi \) is a government

\(^{10}\)See Appendix B.4 for additional details.
transfer to the worker. We can now define a competitive equilibrium for our environment:

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

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

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

(13)

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

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

**Proof.** See Appendix B.1.

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

### 4.1 Firm Size Taxation

We begin by looking at the distortions on firm size implied by the constrained efficient allocation. Such distortions provide a normative rationale for firm size taxation in the decentralization. Assuming differentiability of $T$ and $T_L$, first order conditions from the manager’s problem (MP) give:

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

(14)

where $y_L$ is the marginal product of the worker.\textsuperscript{11} Equation (14) shows that if $T'_L(wL(\theta)) \neq 0$ for some $\theta$, the worker's marginal product is not be equalized across firms. This implies a break down of the well known Diamond and Mirrlees (1971) productive efficiency result.

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

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

\textsuperscript{11}Here we use that $y_L = -n_L/n_{\phi}$, which simply follows from the implicit function theorem.
1. **Optimal marginal firm size distortions satisfy:**

   \[
   T_L'(wL(\theta)) = \frac{f(\theta) - \mu(\theta)v'(n(\theta))\frac{n\theta L(\theta)}{\lambda'}}{f(\theta) + \mu(\theta)v'(n(\theta))\frac{n\theta L(\theta)}{\lambda'}} - 1, \tag{15}
   \]

   where \(\mu(\theta) \geq 0\) is the multiplier on the incentive constraint (5), and \(\lambda' > 0\) and \(\lambda^t > 0\) are the multipliers on (8) and (7), respectively.

2. **If there is a differentiable** \(g : \mathbb{R}_+ \rightarrow \mathbb{R}_+\) **such that** \(n_\theta = g(n(\theta))\) **for all** \(\theta \in \Theta\), **then** \(T_L'(wL(\theta)) = 0\) **for all** \(\theta \in \Theta\).

   \[\Box\]

   **Proof.** See Appendix B.2.

Part 1 of Proposition 3 illustrates how in general a binding incentive constraint (with multiplier \(\mu(\theta)\)) generates nonzero firm level distortions. Part 2 provides a sufficient condition for distortions to disappear. Following the discussion at the end of Section 3, it is immediate that this condition is satisfied in the case in which \(\gamma = 0\) and \(H\) is constant return to scale, or when \(H\) is Cobb-Douglas.\(^{12}\)

### 4.2 Income Taxation

We now move to income taxation. For the rest of the analysis, we make the following standard assumption on preferences:

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

Assuming differentiability of \(T'\), first order conditions from the manager’s problem (MP) are:

\[
1 - T'(\pi(\theta)) = v'(n(y(\theta), L(\theta), \theta))n_y(y(\theta), L(\theta), \theta), \tag{16}
\]

where \(\pi(\theta) \equiv y(\theta) - wL(\theta) - T_L(L(\theta))\) corresponds managerial income of talent \(\theta\). To simplify notation let \(n(\theta) \equiv n(y(\theta), L(\theta), \theta)\) and \(y(\theta) \equiv y(n(\theta), L(\theta), \theta)\) with similar notation for the derivatives \(y_n, y_\theta, y_{nn}\) and \(y_{n\theta}\). When it does not cause confusion we also refer to \(T'(\pi(\theta))\) as \(T'(\theta)\). The next proposition characterizes the optimal marginal income tax rates.

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

\[
\frac{T'(\theta)}{1 - T'(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \cdot \left(1 - \frac{D(\theta)}{D(\theta)}\right) \cdot \left[-\frac{n_\theta}{n} \left(1 - \frac{y_{nn}}{y_n}\right) + \frac{y_{n\theta}}{y_n}\right], \tag{17}
\]

where

\[
D(\theta) \equiv \frac{1}{1 - F(\theta)} \int_\theta^\varnothing \Psi'(U(\theta)) dF(\theta).
\]

\(^{12}\)Scheuer (2014) considers an environment similar to ours with two important differences. First, the firm level production function features \(\gamma = 0\). Second, workers and managers are heterogeneous in two dimensions. He considers two cases: one in which the government can tax workers and managers differently, and one in which it cannot. This second case features firm level distortions, but it is not due to the presence of \(L\) in the incentive constraint but from the presence of an additional no-discrimination constraint absent in our environment.
Proof. See Appendix B.3.

Equation (17) reveals the main forces that shape optimal marginal income tax rates in our framework. The first two terms are well known and are present in the seminal contribution of Saez (2001). The first term looks at the effect of the shape of the talent distribution on marginal tax rates. In particular, high marginal taxes at talent level $\theta$ are attractive as the mass of managers above $\theta$, given by $(1 - F(\theta))$, is large; at the same time the resulting distortion is proportional to the mass of individuals at $\theta$ and to their productivity level, explaining the negative dependence on $\theta f(\theta)$. The second term summarizes the impact on marginal taxes of the redistributive tastes of the government, which are embedded into $D(\theta)$.

The remaining term of (17) (in square brackets) represents the impact on optimal taxes resulting from the behavioral response of the manager as well as resulting from the span of control technology described in the previous section. We first provide some basic intuition on the components of this term then specialize the analysis to a benchmark case that generates the standard tax formula of Saez (2001).

The first two terms: $n_\theta$ and $y_\theta$ are respectively the elasticity of effort and elasticity marginal product with respect to managerial talent. A large absolute value of these elasticities (above one) denotes a fast growth, respectively, of effort and marginal product as we increase the talent level of the manager. This fast growth in turn maps to a fast growth in managerial income: a force for high marginal taxes at the top.

Multiplying the elasticity of hours with respect to managerial talent is a term in brackets $(1 - y_{nn} n_\theta)$. This term captures the adjustment in hours that occurs due to a change in marginal tax rates. The term $1/\varepsilon$ represents the standard labor response of agents to changes in after tax wages. A low elasticity (embodied in a low value of $\varepsilon$) translates into a small response of managers, and hence raises optimal marginal tax rates (it is easy to see that $-\frac{n_\theta}{n} > 0$). The second term in the brackets is the elasticity of marginal product of the manager to effort of the manager. A large absolute value of this elasticity ($y_{nn} \leq 0$) implies that given changes to effort there are large changes in the marginal product of the manager. This implies that as effort decreases (following an increase in tax rates), the marginal product of managerial labor increases so that the distortionary effect of marginal taxes is dampened. This effect, as the previous one, is a force for higher marginal taxes.

Further intuition can be gained by specializing the previous discussion to a simpler production function. Consider the following two assumptions: let $\gamma = 0$ and let $H$ satisfy constant return to scale. In this case, following the discussion of Section 3, we have that firms are undistorted. This implies that $H_L(\theta n, L) = w$ for all $\theta$ and some constant $w$. Since $H_L$ is a homogeneous of degree zero function we have $H_L(\theta n/L, 1) = w$ so that we can write $\theta n = L \cdot m(w)$, with $m(w) = H_L^{-1}(w, 1)$. From the above we have that $\frac{n_\theta}{n} = -1$. Differentiating the production function we have $y_n = \theta H_n \left(\frac{\theta n}{L}, 1\right)$ so that

$$\frac{y_{nn}}{y_n} = \frac{H_{nn} \theta n}{H_n}, \quad \frac{y_n \theta}{y_n} = \frac{H_{nn} \theta n}{H_n} + 1.$$  

\footnote{See also Appendix D for additional details.}
Substituting in (17) we obtain the classical tax formula from Diamond (1998) or Saez (2001) in terms of underlying structural parameters:

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

(DS)

This result shows the importance of a positive scale-of-operations parameter and of firm size distortions in generating a departure from the classic (DS) optimal tax formula.

Proposition 4 applies to quite general production functions \( H \). Our goal is to ultimately take the model to the data. To lay the groundwork for our quantitative analysis we make the following parametric assumption on the production function:

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

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

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

Note that from here on we specialize the discussion assuming a constant return to scale production function.\(^{15}\) The following corollary characterizes optimal income taxes under Assumption 2.

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

\[ \frac{T'(\theta)}{1 - T'(\theta)} = \frac{1 - F(\theta)}{\theta f(\theta)} \left( 1 - \frac{D(\theta)}{D(\theta)} \right) \left( \frac{1}{\varepsilon} + 1 + \frac{\gamma}{1 - \kappa(\theta)} \left( \frac{1}{\varepsilon} + 1 - \rho \kappa(\theta) \right) \right), \] 

where \( \kappa(\theta) = \frac{y_L(\theta)L(\theta)}{y(\theta)} \) is the share of labor costs to total sales for managers of talent \( \theta \).

**Proof.** See Appendix B.4. \( \square \)

The tax equation (18) is useful to qualitatively sign the impact on taxes of a positive span of control parameter. From the above we see that when \( \gamma > 0 \) then we have \( T'(\theta) > T'_{\gamma=0}(\theta) \) (recall \( k(\theta) < 1 \)) : for a given distribution of talent, a positive \( \gamma \) generates higher marginal tax rates than what are implied from the classical (DS) tax formula. At this point an important observation is warranted. The tax formulas described so far are expressed in terms of the underlying (primitive) talent distribution. The tax formulas in Saez (2001) and Diamond and Saez (2011) usually are a convenient re-expression of (DS) expressed in terms of observable quantities, notably the income distribution. To completely relate our tax formulas (and hence our optimal tax prescriptions) with the one in Saez (2001) we need to specify an equilibrium

\(^{14}\)The estimation procedure described in Section 5 can be extended to other production functions as long as the production function satisfies constant return to scale.

\(^{15}\)Refer to Scheuer and Werning (2015) for additional details on taxation with non constant-return-to-scale production functions.
decentralized environment mapping the distribution of talents to the distribution of incomes. This is done in the next section. We return to our comparison with Saez (2001) in Section 6.1.

In the limit, the tax formula in (18) can be simplified further by assuming the following:

**Assumption 3.**

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

\[
\lim_{\theta \to \bar{\theta}} \frac{1 - F(\theta)}{\theta f(\theta)} = \frac{1}{a}.
\]

(b) There is zero social marginal welfare weight at the top: \( \lim_{\theta \to \bar{\theta}} D(\theta) = 0 \).

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

\[
\frac{T'(\bar{\theta})}{1 - T'(\bar{\theta})} = \frac{1}{a} \left( \frac{1}{\varepsilon} + 1 \right) + \gamma \left[ \frac{1}{1 - \bar{\kappa}} \left( \frac{1}{\varepsilon} + 1 - \frac{\sigma - 1}{\sigma} \right) \right].
\]

Equation (19) is the key equation we take to the data in the next section.

5 IDENTIFICATION AND CALIBRATION

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 \( \gamma \) and the Pareto tail parameter of the talent distribution \( a \). Our main contribution in this section is to show how to estimate such parameters using firm level data.

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

5.1 FIRM LEVEL ELASTICITIES

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

**Assumption 4.** The production function \( H \) satisfies constant return to scale.

Consider a competitive equilibrium where the manager faces a linear tax \( \tau \) on her income, pays wage \( w \) (taken as given) to each unit of labor input \( L \), and gets a fraction \( \chi \in (0, 1] \) of total profits.\(^{16}\) Due to the constant returns to scale assumption, we can write the \( \theta \)-manager’s problem as

\[
\max_{\{L, n, \pi\}} (1 - \tau) \chi \pi - v(n)
\]

\(^{16}\)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.
\[
\text{s.t. } \pi = \left[ \theta^\gamma L h \left( \frac{\theta n}{L} \right) - wL \right],
\]

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 (20) are given by:

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

(21)

and

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

(22)

Equations (21) and (22) together imply the following lemma where we show how firm size, output, and profits grow with respect to managerial talent.\(^{17}\)

**Lemma 1.** Let \( \{L(\theta), n(\theta), \pi(\theta)\} \) solve the \( \theta \)-manager's problem in (20). Then the following relationships hold:

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

(23)

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

(24)

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

(25)

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

**Proof.** See Appendix C.1.

Lemma 1 reveals a key property of the scale-of-operations effect: for a given distribution of talent, the distributions of firm size, sales, and profits become more skewed as \( \gamma \) grows. In particular once we shut down the scale-of-operations effect \((\gamma = 0)\) then the growth rates in (23)-(25) are all equal to \((1 + \varepsilon)\). However, in general, \( L, y, \) and \( \pi \) respond more to differences in managerial talent when \( \gamma \) is positive.\(^{18}\)

5.2 ESTIMATING \( \gamma \)

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

**Proposition 5.** For any solution of the \( \theta \)-manager’s problem in (20) the following relationship holds:

\[
\gamma = \frac{1 - \frac{d \ln L(\theta)}{d \ln y(\theta)}}{\frac{d \ln \pi(\theta)}{d \ln y(\theta)}} - \frac{\varepsilon + \varepsilon}{1 + \varepsilon}.
\]

\(^{17}\)It is worthwhile observing that as long as the ownership share \( \chi \) is constant across talents, it does not impact the growth rates of firm size, output, and profits in Lemma 1 and, hence, it does not bias our estimation strategy.

\(^{18}\)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.
Proof. See Appendix C.2.

Equation (26) forms the basis for the estimation of $\gamma$. Specifically, to evaluate $\gamma$ we require the elasticity of firm size with respect to sales ($d \ln L / d \ln y(\theta)$), the elasticity of managerial compensation with respect to sales ($d \ln \pi(\theta) / d \ln y(\theta)$), and parameters ($\sigma$) and ($\varepsilon$). Below we discuss how each of these is estimated.

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

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

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

We look at data from publicly traded US firms in COMPUSTAT. The sample is constructed at an annual frequency from 2000 to 2012.\footnote{From our sample we drop firms that report negative or zero sales and firms with duplicate CUSIP.} 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.\footnote{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.}

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. We only consider firms above (and including) the median firm size for each year. Doing this minimizes the impact of assuming a fixed number of top executives. To see this, for large firms the relationship between sales and non-managerial employees can be approximated as follows:

$$
\ln(\text{Sales}_{i,t}) = \alpha_0 + \alpha_1 \ln(\text{Employees}_{i,t} - \text{Number of top executives}_{i,t})
$$

$$
= \alpha_0 + \alpha_1 \ln(\text{Employees}_{i,t}) + \alpha_1 \ln \left( 1 - \frac{\text{Number of top executives}_{i,t}}{\text{Employees}_{i,t}} \right)
$$

$$
\approx \alpha_0 + \alpha_1 \ln(\text{Employees}_{i,t}).
$$

As a benchmark, we assume that the number of top executives is 20. From (27) we estimate a value of $\hat{\alpha}_1 = \frac{d \ln y}{d \ln L} = .951 (0.002)$, where $d \ln y / d \ln L$ denotes the average value of $d \ln y(\theta) / d \ln L(\theta)$ 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. (2013).

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 Figure 1 we also report estimates 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 as well.

![Figure 1: Estimates of \( \frac{d \ln y}{d \ln L} \) by number of top executives. “Benchmark” refers to estimates of (27) using firms above the median size, “All Sample” refers to estimates of (27) also including firms below the median firm size.](image)

**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.\(^{21}\) This literature has highlighted an empirical regularity, usually denoted as “Roberts’s Law,” which states that, on average, managerial compensation is proportional to a power of 1/3 on the own firm size. Accordingly, in our benchmark calculation we set \(d \ln \pi/d \ln y = 0.34\).

**Frisch elasticity** \((\varepsilon)\), and **elasticity of substitution** \((\sigma)\) We set the value of \(\varepsilon\) based on previous studies. Specifically, following the guidelines of Chetty et al. (2011) we set \(\varepsilon\) equal to 0.5 for our benchmark calibration.\(^{22}\) To pin down the remaining parameter \(\sigma\), we rely on equilibrium modeling restrictions.\(^{23}\) Given the values for the elasticities \(d \ln \pi/d \ln y\) and

---

\(^{21}\)See also Lewellen and Huntsman (1970), Baker et al. (1988), Gabaix and Landier (2008), Frydman and Saks (2010), or Alder (2012).

\(^{22}\)There exists an extensive literature devoted to the estimation of the elasticity of labor supply. For prime age males MacCurdy (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. (2012) and French (2005). Chetty et al. (2011) find values equal to 0.5 on the intensive margin and 0.25 on the extensive margin.

\(^{23}\)There is little direct empirical evidence on how managerial effort substitutes with hired factors of production, such as the number of workers. A possible strategy to identify \(\sigma\) is to use the aggregate behavior of wages over time. For example, Katz and Murphy (1992) use an aggregate production function and estimate an elasticity
Table 1: Estimating the Elasticity of Firm Size With Respect to Sales

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

Notes: Estimates of $\alpha_1$ in (27). 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 (27) of dummies based on Standard Industrial Classification (SIC) from Occupational Safety & Health Administration. “All time period” denotes the usage of the entire dataset up to 1950. “Decile Included” denote the sample of firms by size included in the estimation of $\alpha_1$. We report standard errors in square parentheses.

$d \ln y / d \ln L$ estimated previously, equation (26) does not return a positive value of $\gamma$ for every possible combination of $\varepsilon$ and $\sigma$. Indeed, using $\varepsilon = 0.5$ we have that $\gamma > 0$ if and only if $\sigma > 4.25$. Based on this threshold, we set $\sigma = 5$ in the benchmark calibration and perform a robustness check on this parameter in Section 6. Other values of $\varepsilon$ would yield a different feasible range for $\sigma$, as it is shown in Figure 2. Overall from Figure 2 we see that the feasible range for $\sigma$ precludes a Cobb-Douglas production function ($\sigma = 1$) for empirically relevant values of $\varepsilon$.

We can now determine $\gamma$ using equation (26) in Proposition 5. We get:

$$\gamma = \frac{1 - \frac{1}{0.951}}{1 - 0.34 \times \frac{0.5+5}{0.5+4}} = 0.30.$$  

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

5.3 Estimating $a$

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.

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

$$\frac{1}{a} = \frac{1}{N} \sum_{i=1}^{N} (\ln(\theta_i) - \ln(\bar{\theta})),$$

(28)

where $\{\theta_1, \ldots, \theta_N\}$ is a given a realization of managerial talent, and $\bar{\theta}$ is the minimum possible value of $\theta$. If we let $a_L$ denote the tail parameter of the Pareto distribution of firm size, the analogue to (28) yields:

$$\frac{1}{a_L} = \frac{1}{N} \sum_{i=1}^{N} (\ln L(\theta_i) - \ln(L(\bar{\theta}))).$$

(29)

From equation (23) in Lemma 1 we have that:

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

(30)

Finally, combining (28)-(30) and taking limits we can link the tail parameters on the talent and firm size distributions as follows:25

**Proposition 6.** For any solution of the $\theta$-manager’s problem in (20) we have:

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

(31)

Using equation (31) and the parameters of Section 5.2, the Pareto tail parameter $a$ can be inferred from observed value for $a_L$, and the share of labor costs at the top $\bar{\kappa}$. We pin down

---


25We assume that regularity conditions necessary for consistency of maximum likelihood estimates hold.
these 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 (23)-(25) from Lemma 1 we obtain:

$$1 - \kappa(\theta) = 1 - \frac{d\ln L(\theta)}{d\ln y(\theta)} - \frac{d\ln \pi(\theta)}{d\ln y(\theta)}, \quad (32)$$

using the estimated elasticities $d\ln L/d\ln y$ and $d\ln \pi/d\ln y$ into (32) we get $\bar{\kappa} \approx 0.93$. Plugging in our estimates into equation (31), 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 $a_L$. Fundamentally, this big difference relies on a positive scale-of-operations effect $\gamma$.

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

Table 2: Benchmark Parameter Calibration.

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

Notes: $\varepsilon$ and $\sigma$ are imposed exogenously. Firm elasticities are estimated from the data as described in Section 5.2. $\gamma$, $a$, and $\kappa$ are computed using (26), (31), and (32), respectively.

---

26See equation (C.10) in Appendix C.2.

27This 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 $\frac{0.08}{0.93} \approx 86.2$. 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 PAYSALCE (see http://www.payscale.com/data-packages/ceo-income/full-list). The average across these top firms is 87.4, which closely matches what the estimated $\kappa$ implies.

28The same conclusion holds when calibrating $a$ using the income distribution, as discussed in Appendix E.
6 OPTIMAL TOP INCOME TAX RATES

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:

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

The above prescribed value for top marginal rates relies crucially on the estimated value of \( \gamma \). To see this it is instructive to compare the result with the case in which \( \gamma = 0 \). In this case \( a_{\gamma=0} = 1.59 \) (Diamond and Saez (2011) use a value of \( a = 1.5 \)) and the corresponding tax rate is almost double our benchmark at 65.4 percent.

We next proceed to study the effect of the elasticity of labor supply (\( \varepsilon \)) and the degree of substitutability across inputs (\( \sigma \)) on the optimal top marginal tax rates. These are reported in Table 3. As expected, we observe that tax rates are decreasing in \( \varepsilon \) (see also Figure 3(a)).

Table 3: Top Marginal Tax Rates.

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( \varepsilon = 0.25 )</th>
<th>( \varepsilon = 0.50 )</th>
<th>( \varepsilon = 0.75 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T'(\bar{\theta}) ) (%)</td>
<td>51.8%</td>
<td>51.9%</td>
<td>32.3%</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.16</td>
<td>0.13</td>
<td>0.35</td>
</tr>
<tr>
<td>( a )</td>
<td>13.5</td>
<td>11.7</td>
<td>29.3</td>
</tr>
</tbody>
</table>

| \( T_{\gamma=0}'(\bar{\theta}) \) (%) | 79.1% | 79.1% | 65.4% | 65.4% | 55.7% | 55.7% |
| \( a_{\gamma=0} \) | 1.3 | 1.3 | 1.6 | 1.6 | 1.8 | 1.8 |

Notes: \( T'(\bar{\theta}) \) denotes optimal tax rate as imputed by (19). The values for \( \gamma \) and \( a \) are respectively computed using (26) and (31). \( T_{\gamma=0}'(\bar{\theta}) \) denotes the optimal tax rates with the exogenous constraint of \( \gamma = 0 \).

When looking at changes of the elasticity of substitution between effort of the manager and of the worker we see that marginal tax rates are increasing in \( \sigma \). This is also displayed in Figure 3(b) where marginal taxes are plotted over a wider range of \( \sigma \). In this figure the shaded region represents the subset of the parameter space consistent with a positive value of \( \gamma \). When changing values of either \( \varepsilon \) and \( \sigma \) given equation (26) and (31) we re-compute the values for \( \gamma \) and \( a \). Table 3 displays the re-estimated values. Also displayed are the values of marginal taxes when \( \gamma = 0 \). Compared to our benchmark, marginal taxes are higher. However, the difference between the two (in relative terms) is decreasing as the manager labor supply becomes more inelastic.

To conclude this benchmark it is useful to compare our benchmark optimal tax rate of 32.4 percent with what we see in US data. We look at the March edition of the CPS from 2000 to 2010. For every individual in our sample we compute federal and state taxes of labor income
using the NBER TAXSIM calculator.\textsuperscript{29} For the top 99\textsuperscript{th} percentile we find an effective marginal federal tax rate of 33.5 percent and a marginal state tax rate of 5 percent. \textit{Saez et al. (2012)} report, for the top 1 percent of workers post 2000, a marginal rate of approximately 50 percent. With these values in mind our benchmark prescribes a tax rate of the same magnitude of what we currently see in data.

6.1 RELATIONSHIP WITH \textit{SAEZ (2001)}

Section 4 concluded showing how the classical (DS) tax formula is recovered in the case with $\gamma = 0$. That formula was expressed in terms of primitives of the environment, namely the distribution of managerial talent and the structural parameters of the production function. We now relate our formula to the one in \textit{Saez (2001)} expressed in terms of the income distribution.

In our environment for a given value of $a$ we can recover $a_y$, the tail parameter for the distribution of income (see Appendix E for a derivation):

$$a = \left(1 + \frac{\gamma}{1 - \kappa}\right) (1 + \varepsilon)a_y,$$

substituting in (19) we get:

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

It is immediate that in the case with no scale-of-operations effect ($\gamma = 0$) or in the Cobb-Douglas

\textsuperscript{29}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 \textit{Ales et al. (2014)} for further details.
case \((\rho = 0)\) then we get \(T' = 1/(1 + a_\rho \varepsilon)\) as in Saez (2001) and Diamond and Saez (2011).

Equation (34) allows us to understand further the forces present in this environment compared to the benchmark case of Saez. As discussed in Diamond and Saez (2011) marginal taxes at the top can be understood looking at the tail parameter of the distribution of talent \((a)\) and the income elasticity of the after tax rate \(e = \frac{\partial \log \pi(\theta)}{\partial \log(1 - \tau)}\). In equation (33) we see how the presence of the scale-of-operations effect creates a wedge between the distribution of income and the distribution of managerial talent. This wedge is a force for lower taxes (since it points towards a higher value of \(a\)). At the same time firm level distortions that emerge with the scale-of-operations effect (as long as \(\rho \neq 0\)) generate a lower response of income to taxes. This is a force for higher marginal tax rates. In the two cases discussed above, these two forces cancel each other perfectly. This is, however, not true in general as our benchmark environment has demonstrated.

**Remark 1.** The benchmark environment we have taken to the data features a lower income elasticity of the after tax rate than the one documented in the data (see, for example Saez et al. (2012) and Piketty et al. (2014)). However this elasticity, in our environment, is endogenous to policy (with firm level distortions having a particular strong effect). Indeed, (as we show in Appendix D), absent firm size distortions the value of this elasticity is given by the Frisch elasticity of labor supply. Hence, in this case the value of “\(e\)” generated by the model would be consistent with the one estimated in the data.

### 7 Optimal Firm Size Taxation

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

The values for the parameters \(a, \varepsilon, \sigma,\) and \(\gamma\) are taken from Table 2. In addition here we assume that the talent distribution is Pareto-Lognormal\(^{30}\) with \(\theta \sim PLN(\zeta, \iota^2, a)\), and following Mankiw et al. (2009) we set \(\zeta = 2.76\) and \(\iota = 0.56\). To calibrate \(\beta\) (the share parameter on the production function), we use the definition of \(\kappa\) which implies:

\[
\kappa(\theta) = \frac{MP_L(\theta)L(\theta)}{y(\theta)} = \frac{1}{1 + \frac{\beta}{\gamma+\beta} \left(\frac{\theta}{T}\right)^\rho}.
\]

From the NBER-CES Manufacturing Industry Database, the average team size defined by the number of production workers per non-production worker is estimated to be 3.58. Using this number to proxy \(\left(\frac{\theta}{T}\right)^{-1}\), the equation above and our benchmark value of \(\kappa \approx 0.93\) together imply \(\beta \approx 0.13\). Finally, the social welfare function (refer to (PO)) is \(\Psi(U) = U^{1/2}\).

\(^{30}\)That is, \(\theta \sim \theta_1, \theta_2\), where \(\theta_1 \sim LN(\zeta, \iota^2)\) and \(\theta_2 \sim P(a)\).
Table 4 reports optimal marginal firm size taxes across firm size percentiles.

Table 4: Optimal Marginal Tax on Firm Size.

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

Two facts stand out. First, computed firm distortions are positive and economically significant. For the median firm, for example, these distortions raise the marginal labor cost by 1.75 percent above the wage rate. Second, the marginal tax on labor use is progressive in the range of firm sizes reported. In particular, $T'_L$ increases by 30 percent between the 25th percentile and the top, where it asymptotes at around 2 percent.

It is useful to benchmark the aforementioned prescriptions against actual firm size distortions in the data. As for the shape of such wedges, our model generates progressive firm distortions, which is a property that is ubiquitous across countries (see, Guner et al. (2008)). When comparing levels, on the other hand, one faces a significant obstacle: in practice policies distorting firm sizes include a variety of labor regulations which cannot be easily summarized into a “tax equivalent” measure as in our model. However, a recent study by Garicano et al. (2013) circumvents this problem by structurally estimating the tax equivalent of French labor regulations, which affect firms with more than 50 employees. Using a Lucas (1978) span of control model, the authors find that such labor legislation increase the cost of labor by around 1.3% of the wage. A value in line with the numbers in Table 4.

8 CROSS-COUNTRY RESULTS

A contribution of this paper is to make operational the study of optimal top income tax rates with information available from firm data. This is appealing since, in most countries, this data is publicly available given the regulatory requirements on publicly traded firms. In Table 5 we extend the analysis of Section 6 to a panel of European countries. The first observation is that the main result of Section 6 is also confirmed once we look across countries: optimal tax rates are lower once we consider the scale-of-operations effect as opposed to the case without. Looking at the raw data across countries we see that there is a large variation of firm size (Column 1) and the pay-size elasticity (Column 2). This in turn implies that optimal top tax rates differ across countries. The US together with the Netherlands is close to the bottom when looking at top marginal rates while Switzerland and France are close to the top. This is due to differences in the magnitude of the scale-of-operations parameter as well as differences in the distribution of managerial talent.

---

31 We thank Miguel Ferreira for sharing the estimates of CEO pay-sales elasticities in non-US countries.
### Table 5: Top Tax Rates: Cross-Country Results

<table>
<thead>
<tr>
<th>Country</th>
<th>$a_L$</th>
<th>$\frac{d \ln \pi}{d \ln y}$</th>
<th>$\frac{d \ln y}{d \ln L}$</th>
<th>$T' (\bar{\theta})$</th>
<th>$T'_{\gamma=0} (\bar{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>0.84</td>
<td>0.37</td>
<td>0.92</td>
<td>0.40</td>
<td>0.71</td>
</tr>
<tr>
<td>France</td>
<td>0.90</td>
<td>0.51</td>
<td>0.89</td>
<td>0.48</td>
<td>0.69</td>
</tr>
<tr>
<td>Germany</td>
<td>0.89</td>
<td>0.39</td>
<td>0.93</td>
<td>0.41</td>
<td>0.69</td>
</tr>
<tr>
<td>Italy</td>
<td>1.10</td>
<td>0.52</td>
<td>0.80</td>
<td>0.41</td>
<td>0.65</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1.03</td>
<td>0.30</td>
<td>0.95</td>
<td>0.30</td>
<td>0.66</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.73</td>
<td>0.31</td>
<td>0.92</td>
<td>0.38</td>
<td>0.73</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.88</td>
<td>0.60</td>
<td>0.97</td>
<td>0.55</td>
<td>0.69</td>
</tr>
<tr>
<td>UK</td>
<td>1.00</td>
<td>0.42</td>
<td>0.96</td>
<td>0.41</td>
<td>0.67</td>
</tr>
<tr>
<td>US</td>
<td>1.06</td>
<td>0.33</td>
<td>0.95</td>
<td>0.32</td>
<td>0.65</td>
</tr>
</tbody>
</table>

*Notes:* Optimal top marginal tax rates for France, UK, Italy, Netherlands, Germany, Sweden, Belgium, and Switzerland. The firm size Pareto tail parameter $a_L$ of France, UK, and Italy are taken from Fujiwara et al. (2004). $a_L$ of the Netherlands is from Marsili (2005). For Germany, Sweden, Belgium, and Switzerland, we take the estimates of firm sales Pareto tail parameter $a_y$ from Ramsden and Kiss-Haypäl (2000) and compute $a_L = \frac{d \ln y}{d \ln L} a_y$. The pay size elasticity $\frac{d \ln \pi}{d \ln y}$ of France is taken from Llense (2010), and for the other non-US countries are from Fernandes et al. (2013). The sales size elasticity $\frac{d \ln y}{d \ln L}$ is estimated using GLOBAL COMPUSTAT data for 2000-2012 using the methodology described in Section 5.2. The results are computed based on $\sigma = 5$ and $\varepsilon = 0.5$.

### 9 CONCLUSION

The title of this paper is a reference to the thought provoking novel of Rand (1957). In this dystopian novel what we would call “top income earners” reduce their labor effort in response to high taxes. In the novel, this action is described as having dramatic and long lasting effects for the economy. In this paper, we quantify such effects. Top income earners are modeled as managers whose effort and talent jointly contribute with hired labor to generate output. Our key finding is that tax rates should be substantially lower than what previous literature with unbounded distributions on talent has found. This result is robust to a large range of parameter specifications. The scale-of-operations parameter emerges as a key driver for determining tax schedules.

Methodologically, this paper has shown how to use firm level data (as opposed to using surveys or censuses eliciting workers’ income) to determine key parameters relevant for income taxation. The logical next step involves taking a close look at the impact of managers on production. Two extensions come to mind: the first one is to determine the impact of a hierarchical organizations rather than a single manager technology for optimal taxes. The second extension is to consider the case in which managerial ability and the managerial labor productivity are separate and drawn independently rather than being jointly determined by a unique parameter.

---

32The excellent survey of Slemrod (2000) also features a similar title.
REFERENCES


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 $\theta \in \Theta$:

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

and

\[
    y'(\theta) \left[v''(n(\theta))n_y(\theta)n_\theta(\theta) + v'(n(\theta))n_{\theta y}(\theta)\right] + \\
    + L'(\theta) \left[v''(n(\theta))n_L(\theta)n_\theta(\theta) + v'(n(\theta))n_{\theta L}(\theta)\right] \leq 0, \tag{A.2}
\]

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

**Proof.** The first order condition for $\theta' = \theta$ to be a local maximum of $M(\theta', \theta)$ is $M_1(\theta', \theta) = 0$.\(^{33}\) This is equivalent to \((A.1)\). Differentiating the first order condition $M_1(\theta', \theta) = 0$ with respect to $\theta$ gives $M_{11}(\theta) + M_{12}(\theta) = 0$. Hence, the second order condition $M_{11}(\theta) \leq 0$ can be written as $-M_{12}(\theta) \leq 0$, which gives \((A.2)\).

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

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

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

\[
    c'(\theta') = v'(n(\theta')) \left[n_y(\theta')y'(\theta') + n_L(\theta')L'(\theta')\right]. \tag{A.4}
\]

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

\[
    M_1(\theta', \theta) = J(\theta', \theta') - J(\theta', \theta), \tag{A.5}
\]

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

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

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

\(^{33}\)The subscript $i = 1, 2$ denotes derivative with respect to the first or second argument.
B PROOFS OF SECTION 4

B.1 PROOF OF PROPOSITION 2

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

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

and the budget set

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

Now consider the following claim:

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

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

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

$$c(\hat{\theta}) > R(y(\hat{\theta}), L(\hat{\theta}))$$

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

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

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

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

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

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

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

Clearly, many different tax-wage combinations satisfy the relationship in (B.3).\footnote{For this reason, the levels of $T, T_L$ and $w$ are not determined in the decentralization.} Claim 1 then implies that each of those combinations, $\{c(\theta), y(\theta), L(\theta)\}$ solves the $\theta$-manager’s problem in (MP).

To complete the proof of the decentralization, define the transfer $\phi \equiv c^w - w$, where $c^w$ is the consumption of the worker at the optimum. Given this level $\phi$, the worker’s budget constraint holds with equality at the optimal allocation. The optimum also satisfies market clearing conditions by construction, while the government’s budget is balanced by Walras’ law.
We compute the Pareto optimal allocation by solving the optimal control problem (PO-FOC) where $y(\theta)$ and $L(\theta)$ are the controls and $U(\theta)$ is the state variable. After integrating by parts, the Lagrangian to the planner’s problem is (suppressing dependencies with respect to $\theta, y$ and $L$):

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

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

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

$$
\begin{align}
\lambda^r (1 - v'(n)n_y)f + \mu(v''(n)n_n n_\theta + v' n_{\theta y}) &= 0, \\
-\lambda^l f - \lambda^r v'(n)n_L f + \mu(v''(n)n_n n_\theta + v' n_{\theta L}) &= 0
\end{align}
$$

and the costate equation is

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

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

$$
y_L = -n_L/n_y. \tag{B.7}
$$

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

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

and

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

so that

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

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

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

$$
\frac{n_{\theta L}}{n_{\theta y}} = \frac{n_L}{n_y}. \tag{B.9}
$$

Rearranging (B.8) and substituting (B.9) we get $-\frac{n_{\theta L}}{n_{\theta y}} \lambda^r = \lambda^l$. Simplifying and substituting (B.9) we get $-n_L/n_y = \lambda^l/\lambda^r$, so that $T_L' = 0$.

B.3 Proof of Proposition 4

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

$$
\mu(\theta) = \int_\tilde{\theta}^{\theta} (\lambda^r - \Psi'(U(\theta))) f(\theta) d\theta. \tag{B.10}
$$
Evaluating (B.10) at \( \theta \) gives

\[
\lambda^r = \int_{\Theta} \Psi'(U(\theta)) f(\theta) d\theta. \tag{B.11}
\]

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

\[
\mu(\theta) = (1 - F(\theta)) (D(\theta) - D(\theta)), \tag{B.12}
\]

from (B.4)

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

Substituting into the above yields

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

By Assumption 1 and (B.11) we get

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

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

\[
n_L = -\frac{y_L}{y_n}, \tag{B.14}
\]

and

\[
n_\theta = -\frac{y_\theta}{y_n}. \tag{B.15}
\]

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

\[
-\frac{n_{\theta y}}{n_y} = \left( \frac{y_n}{y_n} \right) \left( \frac{n_{\theta}}{n} \right) + \frac{y_{n \theta}}{y_n}. \tag{B.16}
\]

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

B.4 PROOF OF COROLLARY 1

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

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

Taking derivatives from (B.17) we obtain

\[
n_{\theta}(y, L, \theta) = \left( -\frac{1}{\theta} \right) \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho - \frac{1 - \beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right]^{\frac{1}{\rho} - 1} \left[ \frac{1}{\beta} \left( \frac{y}{\theta g(\theta)} \right)^\rho (1 + \gamma) - \frac{1 - \beta}{\beta} \left( \frac{L}{\theta} \right)^\rho \right].
\]
using the above and \((B.17)\) we get
\[
\frac{n_\theta(y, L, \theta)}{n(y, L, \theta)} = -\frac{1}{\theta} \left( \frac{y}{\theta} \right)^\rho (1 + \gamma) - (1 - \beta)L^\rho - (1 - \beta)L^\rho.
\]

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

The two above imply:
\[
\frac{n_{\theta y}(y, L, \theta)}{n_y(y, L, \theta)} = -\frac{1}{\theta} \left[ (1 - \rho) \frac{\left( \frac{y}{\theta} \right)^\rho (1 + \gamma) - (1 - \beta)L^\rho}{\left( \frac{y}{\theta} \right)^\rho - (1 - \beta)L^\rho} + \rho(1 + \gamma) \right].
\]

By Assumption 2 we have
\[
\left( \frac{y(\theta)}{g(\theta)} \right)^\rho = \beta(\theta n)^\rho + (1 - \beta)L^\rho.
\]

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

Also, Assumption 2 implies:
\[
\frac{\kappa(\theta)}{1 - \kappa(\theta)} = \frac{1 - \beta}{\beta} \left( \frac{L}{\theta n} \right)^\rho,
\]

where \(\kappa(\theta) \equiv y_L(\theta)/y(\theta)\) denotes the share of labor costs to total sales for manager \(\theta\). Using \((B.22)\) in \((B.21)\) gives
\[
\frac{\left( \frac{y}{g(\theta)} \right)^\rho (1 + \gamma) - (1 - \beta)L^\rho}{\left( \frac{y}{g(\theta)} \right)^\rho - (1 - \beta)L^\rho} = 1 + \gamma \left( 1 + \frac{\kappa(\theta)}{1 - \kappa(\theta)} \right).
\]

Using \((B.18)\), \((B.20)\) and \((B.23)\) into \((B.13)\) gives the result.

**C Proofs of Section 5**

**C.1 Proof of Lemma 1**

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

**Lemma 3.**
\[
\frac{1}{\sigma} = - \left( \frac{\theta n}{L} \right) \frac{h''}{h' \kappa}.
\]

\(\text{(C.1)}\)
Proof. From the definition of $\kappa$ we can write
\[
1 - \kappa = \frac{\theta_n h'}{L h}.
\] (C.2)

Let $f(\theta_n, L) = Lh(\theta_n/L)$. Define the elasticity of substitution between $\theta_n$ and $L$ as $\sigma = -\frac{\partial \ln(f_2/f_1)}{\partial \ln(\theta_n)}$. 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
\[
\frac{1}{\sigma} = -\frac{\partial \ln(f_2/f_1)}{\partial \ln(L/\theta_n)} = -\left(\frac{h}{h'} - \frac{\theta_n}{L}\right)^{-1} \frac{L}{\theta_n} \frac{d}{dL} \left(\frac{h}{h'} - \frac{\theta_n}{L}\right).
\]

Differentiating and re-arranging, we get:
\[
\frac{1}{\sigma} = -\left(\frac{\theta_n}{L}\right) \frac{h''}{h'} \left(1 - \frac{\theta_n h'}{L h}\right)^{-1}.
\]
Substituting (C.2) in the above we obtain the result. 

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

Proof. Differentiating (21) and (22) we get
\[
\frac{\partial \ln L}{\partial \ln \theta} - \frac{\partial \ln n}{\partial \ln \theta} = 1 - \theta g' \frac{L}{\theta n} \left(1 - \frac{\theta_n h'}{L h}\right), \quad (C.3)
\]
\[
\frac{n}{L} g'' \frac{d\ln L}{d\ln \theta} + \left[\frac{n}{\theta} \frac{v''}{(1-\tau) \chi} - \frac{\theta_n}{L} g h''\right] \frac{d\ln n}{d\ln \theta} = g' h' + gh' + gh'' \frac{n \theta}{L}. \quad (C.4)
\]

Combining (C.3) and (C.4),
\[
\frac{n}{\theta} \frac{v''}{(1-\tau) \chi} \frac{d\ln n}{d\ln \theta} + \frac{n}{L} g h'' \left(1 + \frac{\gamma \sigma}{1-\kappa}\right) = g' h' + gh' + gh'' \frac{n \theta}{L},
\]
where we applied (C.1), (C.2) and the definition of $\gamma$ on the last term. Further rearranging (C.1) and (C.2) gives
\[
\frac{v''}{(1-\tau) \chi} \frac{d\ln n}{d\ln \theta} = gh' \left(1 + \frac{\gamma}{1-\kappa}\right). \quad (C.5)
\]
From the first order condition (22) we have
\[
\frac{d\ln n}{d\ln \theta} = \varepsilon \left(1 + \frac{\gamma}{1-\kappa}\right), \quad (C.6)
\]
where we used that $\frac{v'}{v''} = \varepsilon$.

Plugging in (C.6) into (C.3),
\[
\frac{d\ln L}{d\ln \theta} = \varepsilon \left(1 + \frac{\gamma}{1-\kappa}\right) + 1 - \gamma \frac{L}{\theta n} \left(1 - \frac{\theta_n h'}{L h}\right). \quad (C.7)
\]
Then applying (C.1) and (C.2) and rearranging gives (23).
Now we obtain equation (24). By constant returns to scale we can write $y = gLh$, so that

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

Substituting (C.2) in the above gives

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

So substituting (23) and (C.6) into the above expression gives (24).

Finally, we derive equation (25). Profits are given by $\pi = y - wL$. Then

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

or

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

where $\kappa = wL/y$. Substituting (23) and (24) into (C.7) and rearranging gives (25).

C.2 PROOF OF PROPOSITION 5

Proof. The following relationships are derived by combining (23)-(25):

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

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

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

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

rearranging equation (C.8), we have:

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

Substituting (C.10) into equation (C.11), we obtain equation (26).

D FIRM DISTORTIONS AND TAX ELASTICITIES

In this section we show that if there are no firm level distortions then a wage rate exists for the effective effort of the manager and the income elasticity of the after tax rate equals the Frish elasticity of labor supply.
We first show that at the optimum it is possible to write the income of managers of talent \( \theta \) as \( \pi(\theta) = \omega(\theta, w)n \) where \( \omega(\theta, w) \) is the wage of managers of talent \( \theta \) exercising effort \( n \). The first order condition with respect to \( L \) is: \( \theta \gamma H_L(\theta n, L) = w. \) Since \( H_L \) is a homogeneous of degree zero function we have \( \theta \gamma H_L(\theta n/L, 1) = H_L^{-1}(\frac{w}{\theta \gamma}, 1) \). This relationship implies that for a given \( \theta \) and \( w \) the relationship between \( \theta n \) and \( L \) is linear. Define \( m(\theta, w) = \frac{1}{H_L^{-1}(w/\theta \gamma, 1)} \). So that \( L = m(\theta, w) \theta n. \) Substituting in the expression for profits we have:

\[
\pi(\theta, n) = \theta \gamma H(\theta n, m(\theta, w) \theta n) - w \theta n.
\]

This relationship implies that for a given \( \theta \) and \( w \) the relationship between \( \theta n \) and \( L \) is linear. Define \( m(\theta, w) = \frac{1}{H_L^{-1}(w/\theta \gamma, 1)} \). So that \( L = m(\theta, w) \theta n. \) Substituting in the expression for profits we have:

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

We can now write the problem of the manager as:

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

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

\[
\frac{\partial n}{\partial (1 - \tau)} = \frac{1}{v''(n(\theta))} \cdot \omega(\theta, w) = \frac{v'(n(\theta))}{v''(n(\theta))} \cdot \frac{1}{(1 - \tau)}, \tag{D.1}
\]

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

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

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

**E Estimating \( a \) from the Income Distribution**

In Subsection 5.3 we estimated \( a \) using the distribution of firm sizes. In this section we proceed similarly but we focus instead on the distribution of incomes. From equation (25) in Lemma 1 we have that

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

As before, approximating \( k(\theta) = \hat{\kappa} \) and substituting in (28) we get

\[
\frac{1}{a} = \frac{1}{(1 + \frac{\gamma}{1 - \hat{\kappa}})} (1 + \varepsilon) \frac{1}{N} \sum_{i=1}^{N} \ln \pi(\theta^i).
\]

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

\[
a = \left( 1 + \frac{\gamma}{1 - \hat{\kappa}} \right) (1 + \varepsilon)a_y. \tag{E.1}
\]

Taking \( a_y = 2 \) from Saez (2001) with the above estimates of \( \kappa \) and \( \gamma \) we get \( a = 15.62. \) In this case the top tax rate is equal to 44.4%.