How Much Does Atlas Shrug?*
(Preliminary - Do Not Circulate)

Laurence Ales **A. Andrés Bellofatto**
Jessie Jiaxu Wang†
Carnegie Mellon University

February 15, 2014

Abstract

This paper studies the optimal top income tax rate. Top income earners are modeled as managers operating a span-of-control technology as in Rosen (1982). In particular, managerial skills increase managers' productivity through both supervision and indivisible decisions, thus giving rise to a *scale-of-operations* effect (Mayer (1960)). We provide a tax formula linking the optimal top marginal tax rate to primitives of technology and preferences. The presence of a *scale-of-operations* effect prescribes higher taxes than in standard linear productivity models. We show how technology parameters can be identified using data on the elasticity of firm size with respect to sales, and the elasticity of managerial compensation with respect to sales. We estimate the technology parameters using U.S. firm level data from Computstat and the NBER-CES manufacturing industry database. Our quantitative results suggest that even for high elasticities of labor supply marginal taxes at the top should be substantially higher than those observed in the U.S. data.

*We thank Brent Glover, Burton Hollifield, Chris Sleet, Chris Telmer, Sevin Yeltekin, Ariel Zetlin-Jones and participants of the Carnegie Mellon Ph.D. Macro-Finance lunch for feedback.
†Ales: ales@cmu.edu; Bellofatto: andresb@cmu.edu; Wang: jiaxuw@andrew.cmu.edu

1 INTRODUCTION

Income taxation of top income earners is a controversial and recurrent topic in the tax policy debate. This is due to two reasons: heightened concerns over an increasing inequality of the income distribution;[1] and also for practical fiscal considerations.[2] At the same time, the large literature in public finance provides an equally large range of values on what the marginal top tax rate should be, which is partly attributed to the lack of consensus on how to quantify tax responses.[3] The main result of this paper is that even in the face of large labor supply responses, optimal marginal tax rates for high income earners are substantially higher than what we observe in the U.S. economy.[4]

In this paper, in line with the empirical evidence, we model top income earners as managers rather than workers as is standard in the public finance literature.[5,6] Managers operate a span-of-control technology similar to Rosen (1982) and exert effort which combined with hired labor generates output. Managers in our model are heterogenous in their skills, which increase both their effective effort and the overall productivity of the firm: a scale-of-operations effect (see Mayer (1960)). In this sense, output exhibits increasing returns with respect to managerial skills, unlike what is standard in the public finance literature where production technologies display constant returns to skills. In our environment, more productive managers will operate larger firms and receive larger levels of (before-tax) income. In addition, as originally shown by Rosen (1982), the scale-of-operations effect makes managerial rewards grow at a faster rate than managerial skills. Thus, the distribution of rewards becomes skewed relative to distribution of talent.

We consider the case in which effort of the manager and her skill level are private information. Firm size and output are observable. In this environment we characterize the constrained efficient allocation and show that it can be decentralized as a competitive equilibrium with taxes on income and firm size. Our analysis centers on the optimal

[1]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.

[2] Indeed in 2007 the top percentile paid 40.4% of total federal individual income tax revenues (Diamond and Saez (2011)).

[3] Mirrlees (1971) and Sadka (1976) prescribe a zero marginal tax rate at the top. Using different assumptions on the skill distribution, Saez (2001), or Diamond (1998) call for for asymptotic top rates as high as 80%. See Mankiw, Weinzierl, and Yagan (2009) and Diamond and Saez (2011) for recent surveys.

[4]According to Internal Revenue Service tax tables, in the year of 2013, U.S. taxpayers at the top could potentially face a tax bracket of 55.38% (including Federal, State and FICA taxes). Capital gains, interest and dividends, are taxed at a maximum of 20%.

[5]Business owners or self-employed account for 81% of those at the top 1% of the wealth distribution. See also Cagetti and Nardi (2006), Gentry and Hubbard (2004), Quadrini (2000).

[6]Diamond and Saez (2011) provide a recent survey on the standard public finance methodology.

marginal income tax at the top. We provide an optimal top tax formula and highlight its dependence on the span-of-control technology parameters. We show that as the scale-of-operations effect becomes more significant, taxes at the top should increase. This is due to the key insight from Rosen (1982) which was described above: as the scale-of-operations effect becomes larger the distribution of (before-tax) income becomes more right skewed. In turn, progressively fatter tails leads to higher optimal marginal tax rates.[7]

The tax formula provides important insights on the forces at play that determine tax rates but provides little guidance as to the magnitude of these effects. To quantify such effects we calibrate our environment to U.S. data. A key challenge is to determine the magnitude of the scale-of-operations effect, i.e. the marginal impact of managerial skill on the overall productivity of the firm. We show how to identify this parameter by looking at the elasticity of managerial compensation with respect to sales using firm level data from Computstat and the NBER-CES manufacturing industry database. In our benchmark environment we find optimal tax rate of approximately 65%, which is nearly 10% higher than ignoring the scale-of-operations effect. Beyond the point estimate, we observe that the top tax rate is relatively insensitive to the elasticity of labor supply. Indeed, even for when the Frisch elasticity of labor supply is equal to 3 (as in Prescott (2004)) the optimal top tax rate is still well above 60%, whereas a model without a scale-of-operations effect (such as Saez (2001)) would prescribe optimal top tax rates close to the ones observed in the U.S. data.

To summarize, this paper makes three contributions. First, we provide an optimal top tax for a span-of-control environment as in Rosen (1982). Second, we show how the determinants of such tax formula can be calibrated using readily available firm level data. Finally, we show that even for high elasticities of labor supply marginal taxes at the top should be substantially higher than those observed in the data for the U.S.

RELATED LITERATURE

To be completed.

The remainder of the paper is organized as follows. In Section 2 we describe the environment. In Section 3 we describe our notion of optimality, we study the decentralized

[7]The relationship between the "thickness" of the distribution of pre-tax income and its relationship with optimal marginal tax rates is well understood. In a nutshell, fatter right tails in the pre-tax income distribution imply that there is more individuals above any given level of pre-tax income. As a consequence, more people are taxed inframarginally (without distortion) when marginal taxes increase above any given level of pre-tax income, which makes higher marginal taxes more attractive. See, e.g., Mirrlees (1971) and Saez (2001).

environment and compute optimal tax rates. In Section 5 we describe our identification strategy and calibrate the environment. In Section 6 we display our quantitative results. Section 8 concludes. All of the proofs are relegated to the Appendix.

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. Workers and managers have quasi-linear preferences over consumption c and effort n which can be represented by the utility function:

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

where v' and v'' exist and are positive for all values of n.

Workers are homogenous and supply labor inelastically. Managers are heterogenous with respect to a productivity parameter $\theta \in \Theta$ with $\Theta = \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}_+ \setminus 0$. Managerial type, θ, is assumed to be distributed according to the cumulative distribution function $F : \Theta \to [0,1]$ with density function $f : \Theta \to \mathbb{R}_+$.

Following Rosen (1982), managers operate a span-of-control technology. Specifically, managers of type θ hire labor $L(\theta)$ and exert managerial effort $n(\theta)$ to produce final output $y(\theta)$ according to

$$y\Big(n(\theta), L(\theta); \theta\Big) = \theta^{\gamma} H\left(\theta n\left(\theta\right), L\left(\theta\right)\right), \tag{1}$$

where $\gamma > 0$, H is increasing in both arguments, has constant returns to scale and satisfies neoclassical properties.[8]

Note that managerial skills enter (1) in two ways. First, θ is managerial effort-augmenting. Second, θ is total-factor-productivity improving through $\gamma > 0$. In what follows, we refer to $\gamma > 0$ as the scale-of-operations effect (see Mayer (1960)).[9] This formulation is in line with the one in Rosen (1982) where γ captures the indivisibility of managerial decisions. The idea is that managers' actions naturally affect the productivity of all workers under their supervision (irrespective of their number). Our technological specification is more general to that in Rosen (1982), as we also incorporate managerial

[8] Lucas (1978) presents a similar span-of-control technology, but with a counterfactual elasticity of managerial rewards with respect to firm size which is crucial in our environment.

[9] As we show in Section 5, γ must be strictly positive in order to match certain firm level elasticities in the data. There is also a large literature connecting the skill of manager and the productivity and size of a firm. See for example Bartelsman and Doms (2000) and references therein.

effort n as an intensive margin.[10]

We assume that $c(\theta)$, $y(\theta)$ and $L(\theta)$ are observable, while θ, $n(\theta)$ and $\theta n(\theta)$ are each agent's private information. The assumption that θ and $n(\theta)$ are unobservable is common in the Mirrleesian literature (see Mirrlees (1971)). In contrast to this literature, though, we supppose that $\theta n(\theta)$ is not observable either. This assumption is natural in our environment because, as long as θ is TFP augmenting, $\theta n(\theta)$ does not correspond to income or effective effort as in Mirrlees (1971).

An allocation in this economy is defined as a mapping (c, y, L) where $c : \Theta \to \mathbb{R}_+$, $n : \Theta \to \mathbb{R}_+$, $y : \Theta \to [0, \bar{y}]$, and $0 < \bar{y} < \infty$.[11] An allocation is said to be *feasible* if[12]

$$\int_{\Theta} c(\theta) dF(\theta) \leq \int_{\Theta} y(\theta) dF(\theta), \tag{2}$$

and

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

3 PARETO OPTIMALITY

In this section, we compute Pareto optimal allocations using a direct mechanism where managers report their types θ to a social planner and are assigned an allocation for consumption $c(\theta)$, output $y(\theta)$ and labor $L(\theta)$ accordingly. In what follows, we define $n(y(\theta'), L(\theta'); \theta)$ as the level of effort exerted by a manager of type θ who mimics an agent of type θ'.

An allocation (c, y, L) is said to be *incentive-compatible* when truthful revelation is optimal for the agents, that is:

$$c(\theta) - v(n(y(\theta), L(\theta); \theta)) \geq c(\theta') - v(n(y(\theta'), L(\theta'); \theta)), \tag{4}$$

for all $\theta, \theta' \in \Theta$.

Let $\Psi : \mathbb{R} \to \mathbb{R}$ be an increasing and concave function. Then Pareto optimal allocations solve

$$\max_{\{c(\theta), y(\theta), L(\theta)\}_{\theta \in \Theta}} \int_{\Theta} \Psi\left(c(\theta) - v(n(y(\theta), L(\theta); \theta))\right) dF(\theta), \tag{PO}$$

[10] Our preferred interpretation is that n is related to supervisory effort.

[11] The upper bound $\bar{y}$ makes the set of feasible allocations compact.

[12] Notice that we abstract from worker's consumption in our definition of feasibility. This will not change our results for optimal managerial taxation below, as long as the social welfare function is additively separable between worker's and manager's utility.

subject to (2), (3) and (4).

The function Ψ summarizes social preferences for redistribution across different types. Accordingly, we will refer to $\Psi'\left(c(\theta) - v(n(y(\theta), L(\theta);\theta))\right)$ as the *social marginal welfare weight* on θ-managers.

3.1 FIRST ORDER APPROACH

Define $M(\theta';\theta) \equiv c(\theta') - v(n(y(\theta'), L(\theta');\theta))$. It is clear that incentive compatibility (4) requires that for all $\theta \in \Theta$, $M(\theta';\theta)$ attain a global maximum at $\theta' = \theta$. We start by characterizing local maxima of $M(\theta';\theta)$ at $\theta' = \theta$.

Lemma 1. *Let $M(\theta';\theta) \equiv c(\theta') - v(n(\theta';\theta))$, where $n(\theta';\theta) \equiv n(y(\theta'), L(\theta');\theta)$. Then $M(\theta';\theta)$ attains a* local *maximum at $\theta' = \theta$ if and only if*

$$c'(\theta) - v'(n(\theta;\theta))\left[n_y(\theta;\theta)y'(\theta) + n_L(\theta;\theta)L'(\theta)\right] = 0, \tag{5}$$

and

$$\begin{aligned} &y'(\theta)\left[v''(n(\theta;\theta))n_y(\theta;\theta)n_\theta(\theta;\theta) + v'(n(\theta;\theta))n_{y\theta}(\theta;\theta)\right] + \\ &L'(\theta)\left[v''(n(\theta;\theta))n_L(\theta;\theta)n_\theta(\theta;\theta) + v'(n(\theta;\theta))n_{L\theta}(\theta;\theta)\right] \leq 0. \end{aligned} \tag{6}$$

Proof. See Appendix A.1. □

The next lemma shows that the local maximum conditions of Lemma 1 imply that (4) holds.

Proposition 1. *Suppose that for all $\theta \in \Theta$ (5) and (6) hold. Then (4) holds.*

Proof. See Appendix A.1. □

Let $U(\theta) \equiv c(\theta) - v(n(\theta))$ for all θ. Then by Lemma 1 we can write the planner's problem in (PO) as:

$$\max_{\{U(\theta), y(\theta), L(\theta)\}_{\theta\in\Theta}} \int_\Theta \Psi\left(U(\theta)\right) dF(\theta), \tag{PO-FOC}$$

subject to (6) and

$$U'(\theta) = -v'(n(y(\theta), L(\theta);\theta))n_\theta(y(\theta), L(\theta);\theta), \tag{7}$$

$$\int_\Theta \left[y(\theta) - U(\theta) - v(n(y(\theta), L(\theta);\theta))\right] dF(\theta) \geq 0, \tag{8}$$

$$\int_{\Theta} L(\theta) dF(\theta) \leq 1. \tag{9}$$

In what follows, we will take a *first order approach* to problem (PO-FOC). That is, we will ignore the second order condition for a global maximum (6) and verify it ex-post using simulations.[13]

Before closing this section, we show that the second best features a production inefficiency. More precisely, in the constrained efficient allocation workers' marginal products are not equalized across firms. In the decentralization of section 4 this distortions will map into a tax on firm size.

Proposition 2. *Suppose (6) is satisfied. Then at any Pareto optimum $\{U^*(\theta), y^*(\theta), L^*(\theta)\}_{\theta\in\Theta}$ workers' marginal products are not equalized across firms.*

Proof. See Appendix A.1. □

4 DECENTRALIZATION

In this section we describe a decentralization for the constrained efficient allocations in our environment and characterize the corresponding optimal income tax rates.

We begin with the definition of a competitive equilibrium with taxes.

Definition 1. *A competitive equilibrium with taxes is an allocation (c, y, L), a wage w and a tax system $\{T(\pi), T^F(wL)\}$ such that:*

1. *Given $\{w, T(\pi), T^F(wL)\}$ each θ-manager solves:*

$$\max_{c,y,L} \; c - v(n(y, L; \theta)) \tag{P1}$$

subject to:

$$c = y - wL - T^F(wL) - T(y - wL - T^F(wL)).$$

2. *Goods and labor markets clear.*

3. *Government's budget constraint is balanced.*

Using standard arguments in the literature we can show that the constrained efficient allocation can be achieved as the outcome of a competitive equilibrium with taxes. Notice that our notion of a competitive equilibrium features a tax on firm size $T^F(wL)$. This is as a consequence of the results from Proposition 2.

[13] This approach is fairly standard in the dynamic public finance literature. See Golosov, Troshkin, and Tsyvinski (2011), Pavan, Segal, and Toikka (2012) or Farhi and Werning (2013).

4.1 OPTIMAL INCOME TAXES

In what follows we center attention on the properties of the optimal marginal income taxes. Assuming differentiability of $T(\cdot)$, from the first order conditions on (P1) we get

$$1 - T'(\pi(\theta)) = v'(n(y(\theta), L(\theta); \theta))n_y(y(\theta), L(\theta); \theta), \tag{10}$$

where $\pi(\theta) \equiv y(\theta) - wL(\theta) - T^F(wL(\theta))$ corresponds to managerial income for type θ.

The next proposition characterizes the optimal marginal income tax on managers.

Proposition 3. *Suppose (6) is satisfied. Then at any Pareto optimum $\{U^*(\theta), y^*(\theta), L^*(\theta)\}_{\theta\in\Theta}$, $T'(\pi^*(\theta))$ satisfies:*

$$\frac{T'(\pi^*(\theta))}{1 - T'(\pi^*(\theta))} = \frac{1 - F(\theta)}{f(\theta)} \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \left(-\frac{v'(n^*(\theta))n^*(\theta)}{v''(n^*(\theta))} \frac{n_\theta(y^*(\theta), L^*(\theta); \theta)}{n(y^*(\theta), L^*(\theta); \theta)} - \frac{n_{\theta y}(y^*(\theta), L^*(\theta); \theta)}{n_y(y^*(\theta), L^*(\theta); \theta)}\right),$$

where $D(\theta) \equiv \frac{1}{1-F(\theta)} \int_\theta^\infty \Psi'(U^*(\theta))\, dF(\theta)$, $\pi^*(\theta) \equiv y^*(\theta) - wL^*(\theta)$ and $n^*(\theta) \equiv n(y^*(\theta), L^*(\theta); \theta)$.

Proof. See Appendix A.2. □

For the rest of the analysis we make the following assumptions on preferences and technology.

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

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

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

where the elasticity of substitution between $\theta n(\theta)$ and $L(\theta)$ is given by $\sigma = \frac{1}{1-\rho}$.

Under Assumption 1 and when taxes are linear, ε maps to the elasticity of managerial income to the net-of-tax rate $(1 - T')$. Refer to Appendix B for details.

The next Corollary characterizes optimal income taxes under Assumptions 1 and 2.

Corollary 1. *Suppose assumption 2 holds and that (6) is satisfied. Then at any Pareto optimum $\{U^*(\theta), y^*(\theta), L^*(\theta)\}_{\theta\in\Theta}$, $T'(\pi^*(\theta))$ satisfies*

$$\frac{T'(\pi^*(\theta))}{1 - T'(\pi^*(\theta))} = \frac{1 - F(\theta)}{\theta f(\theta)} \left(1 - \frac{D(\theta)}{D(\underline{\theta})}\right) \left(\frac{1}{\varepsilon} + 1 + \frac{\gamma}{1 - \kappa(\theta)} \left(\frac{1}{\varepsilon} + 1 - \rho\kappa(\theta)\right)\right), \tag{11}$$

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

Proof. See Appendix A.3. □

As in Saez (2001) tax distortions in (11) depend on labor supply responses, on the distribution of skills, and on the welfare criterion. What is new in our environment is that the tax formula also depends on technological parameters. In the next subsection we explore the consequences of this novel determinants for optimal marginal tax rates at the top.

4.2 ASYMPTOTIC INCOME TAXES

To highlight the properties of optimal top tax rates and their relationship with technological parameters we introduce the following asymptotic assumptions:

Assumption 3. *The skill distribution has a Pareto tail with parameter* $\frac{1}{a} > 0$*:*

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

Assumption 4. *There is zero marginal social weight at the top:* $\lim_{\theta \to \infty} D(\theta) = 0$.

We now derive the optimal marginal tax at the top.

Proposition 4. *Suppose assumptions 2-4 hold and that (6) is satisfied. Then at any Pareto optimum the marginal tax at the top* $T'(\pi^*(\infty))$ *satisfies:*

$$T'(\pi^*(\infty)) = \frac{1}{1 + a\left[\frac{1}{\varepsilon} + 1 + \frac{\gamma}{1-k(\infty)}\left(\frac{1}{\varepsilon} + 1 - \rho k(\infty)\right)\right]^{-1}}. \tag{12}$$

Proof. To show (12) we take limits of (11). Substituting Assumptions 2-4 and rearranging the result follows immediately. □

Clearly, our environment nests traditional top tax formulas first derived by Diamond (1998) or Saez (2001). This can be recovered by setting $\gamma = 0$ in (12):

$$T'(\pi^*(\infty))_{Saez} = \frac{1}{1 + a\frac{\varepsilon}{1+\varepsilon}}. \tag{13}$$

The next Corollary shows that top taxes in our environment are, ceteris paribus, higher that those in (13).

Corollary 2. $T'(\pi^*(\infty)) > T'(\pi^*(\infty))_{Saez}$.

Proof. Follows from the fact that since $\rho = (\sigma - 1)/\sigma$ then $\rho \leq 1$ and also the observation that $\kappa(\theta) \in [0, 1]$. □

The upward pressure on taxation of the previous proposition is due to two effects: The *γ-effect* (or scale-of-operations effect) and the *labor demand effect*. We explain each of these in turn.

The *γ-effect* refers to the impact of $g' > 0$ which is related to the indivisibility of management decisions. This mechanism pushes optimal top taxes *upwards*. The intuition behind this result can be easily grasped when the only input in production is managerial effort, i.e. by setting $\beta = 1$ so that $\kappa(\infty) = 0$. Then we get from (12) that:

$$T'(\pi^*(\infty))_{\gamma-effect} = \frac{1}{1 + \frac{a}{1+\gamma}\frac{\varepsilon}{1+\varepsilon}}.$$

In this particular case, one can interpret $y(\theta) = g(\theta)\theta n(\theta)$ and $g(\theta)\theta$ as the before-tax income and the (unobservable) wage of the manager, respectively. Clearly, the wage differential across managers increases as $g' > 0$, which generates fatter right tails in the before-tax income distribution. Thus, given any income threshold, higher marginal taxes increase tax revenues above such threshold (as higher marginal taxes act as an infra-marginal tax for income earners on the right of the distribution). Using Saez's terminology, when $g' > 0$ the "mechanical effect" on increasing the marginal tax rate is larger, while the "behavioral effect" remains unchanged with respect to the case when $g' = 0$. Consequently, due to the *γ-effect* one obtains higher optimal marginal taxes at the top. Another way of seeing why the *γ-effect* is pushing the top tax rates up is by noting that high types will have stronger incentives to deviate when the wage differential increases. In order to relax incentive constraints, it will be optimal to increase marginal taxes at the top.[14]

Now consider the *labor demand effect*, which arises when $\beta < 1$. This mechanism reinforces the *γ-effect* discussed above as it also pushes taxes up. The simplest way to illustrate this is by considering a Cobb-Douglas production function which corresponds

[14]Formally, take $\theta, \theta' < \infty$ at the right tail of the income distribution with $\theta > \theta'$. The incentive constraint that prevents that the θ type mimics θ' is $c(\theta) - v(n(\theta)) \geq c(\theta') - v(n(\theta')\frac{\theta' g(\theta')}{\theta g(\theta)})$. Here $n(\theta')\frac{\theta' g(\theta')}{\theta g(\theta)}$ is the required effort that the θ type have to exert to replicate θ' income, so given $n(\theta')$ such effort falls when g' grows. In order to relax this constraint, it is optimal to increase $n(\theta')$ when g' rises, which translates into higher marginal taxes for θ'. Since the same reasoning holds for any two types at the right tail of the distribution, it follows that one will observe higher average marginal taxes at the top when g' grows.

to $\rho = 0$. The optimal marginal top tax rate is now given by:

$$T'(\pi^*(\infty))_{Cobb-Douglas} = \frac{1}{1 + \frac{a}{1+\frac{\gamma}{\beta}}\frac{\varepsilon}{1+\varepsilon}}. \tag{14}$$

Compared to the case where $\beta = 1$, (14) shows that $T'(\pi^*(\infty))$ rises when $\beta < 1$. The reason is that when $\beta < 1$ production not only depends on managerial effort but also on labor demand. Hence, as marginal tax rates increase (and the marginal return on managerial effort falls) managers will be willing to substitute out their own effort for labor. As a consequence, when $\beta < 1$ income of the managers will not fall as much in response to an increase in marginal taxes. In other words, the "behavioral effect" of an increase in the marginal tax is smaller than in the standard Mirrleesian environment with only effort in the production function.

We next proceed to quantify both effects and determine the optimal tax rates.

5 IDENTIFICATION AND CALIBRATION

This section describes the calibration of the parameters in our top tax formula in equation (12). This includes two standard parameters which have already been identified in the optimal taxation literature (the elasticity of labor supply ε and the tail parameter of the Pareto distribution of talent a), two novel technology parameters (the degree of substitutability between the types of labor inputs ρ, and the scale-of-operations parameter γ), and the endogenous labor costs share implied by the production technology κ.

There is a vast literature estimating the elasticity of labor supply.[15] We follow Chetty, Guren, Manoli, and Weber (2011) and set ε equal to 0.75. We also show the effect of a wide range of labor elasticities comprising both micro and macro estimates. The distribution parameter a is set to 1.5 which falls in the range of values used by Saez (2001) and Diamond and Saez (2011). Finally for σ we follow estimates of the team production model of Adams (2006) (based on Holmstrom (1982)); it is determined that the substitution between different groups in a CES production function features $\rho \in (.5, 1)$. As such, we set $\rho = 0.75$, implying $\sigma = 4$.[16]

We require two additional parameters for the tax formula: $\{\kappa, \gamma\}$. Our strategy is

[15] For prime age males MaCurdy (1981) and Altonji (1986) estimate an elasticity between 0 to 0.54. Saez (2003) using the NBER tax panel from 1979 to 1981, estimates a labor elasticity of 0.25. Similar ranges are estimated by Blundell, Pistaferri, and Saporta-Eksten (2012) and French (2005). Chetty, Guren, Manoli, and Weber (2011) find values equal to 0.5 on the intensive margin and 0.25 on the extensive margin. In the macro literature King and Rebelo (1999) and Prescott (2004) find values between 2 to 4.

[16] The results for the optimal top tax rates reported below are not very sensitive to the value of ρ.

to estimate κ directly from data and estimate the novel γ-effect parameter using the equilibrium restrictions generated by our decentralized environment. We next describe our identification strategy.

5.1 FIRM LEVEL ELASTICITIES

In this section we follow Rosen (1982) and derive equilibrium restriction that relate γ to observables. Relative to Rosen's derivation we include income taxation and an elastic margin for effort. First note that given that the technology in (1) has constant returns to scale, we can write

$$y(\theta) = g(\theta)L(\theta)h\left(\frac{\theta n(\theta)}{L(\theta)}\right), \tag{15}$$

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

Now consider a competitive equilibrium where the manager faces a linear tax τ on her income, pays a real wage w to each unit of labor L, and gets a fraction $\chi \in [0, 1]$ of total profits. As a special case, if $\chi = 1$ there is no separation between ownership and control. This is the case considered in Rosen (1982). We can write the θ-manager's problem as:

$$\max_{L,n,\pi} \; \pi - v\left(n\right) \tag{16}$$

subject to:

$$\pi = (1-\tau)\chi\left[g\left(\theta\right)Lh\left(\frac{\theta n}{L}\right) - wL\right].$$

The first order conditions of the problem in (16) are

$$g\left(\theta\right)\left[h\left(\frac{\theta n}{L}\right) - \frac{\theta n}{L}h'\left(\frac{\theta n}{L}\right)\right] = w, \tag{17}$$

and

$$(1-\tau)\chi g\left(\theta\right)\theta h'\left(\frac{\theta n}{L}\right) = v'\left(n\right). \tag{18}$$

By manipulating (17) and (18) we can prove the following Lemma.

Lemma 2. *Let $\{L(\theta), n(\theta), \pi(\theta)\}$ solve the θ-manager's problem in (16). Then the following holds:*

$$\frac{d\ln \pi(\theta)}{d\ln\theta} = 1 + \frac{\gamma}{1-\kappa(\theta)}, \tag{19}$$

$$\frac{d\ln n(\theta)}{d\ln\theta} = \varepsilon\left(1 + \frac{\gamma}{1-\kappa(\theta)}\right), \tag{20}$$

$$\frac{d\ln L(\theta)}{d\ln\theta} = 1 + \frac{\gamma\sigma}{1-\kappa(\theta)} + \varepsilon\left(1 + \frac{\gamma}{1-\kappa(\theta)}\right), \tag{21}$$

$$\frac{d\ln y(\theta)}{d\ln\theta} = 1 + \gamma + \varepsilon\left(1 + \frac{\gamma}{1-\kappa(\theta)}\right) + \frac{\kappa(\theta)}{1-\kappa(\theta)}\gamma\sigma, \tag{22}$$

where $\kappa(\theta) \equiv wL(\theta)/y(\theta)$ is the share of labor costs to total sales, and $\sigma = \frac{1}{1-\rho}$ is the elasticity of substitution between θn and L.

Proof. See Appendix A.4. □

Equations (19)-(22) are instrumental in the derivation of our estimation equations in the next Proportion. But Lemma 2 also reveals some key properties of the scale-of-operations effect which are essential in our environment. In particular, equation (19) shows that as long as $\gamma > 0$, the distribution of rewards becomes skewed relative to distribution of talent. Also, according to (21) and (22), an analogous relationship holds for the distributions of firm size and and total sales, respectively.

Proposition 5. *The following holds:*

$$\frac{d\ln L(\theta)}{d\ln y(\theta)} = \frac{(1-\kappa(\theta))(1+\varepsilon) + \gamma(\sigma+\varepsilon)}{(1-\kappa(\theta))(1+\gamma+\varepsilon) + \gamma(\kappa(\theta)\sigma+\varepsilon)}, \tag{23}$$

$$\frac{d\ln \pi(\theta)}{d\ln y(\theta)} = \frac{1-\kappa(\theta)+\gamma}{(1-\kappa(\theta))(1+\gamma+\varepsilon) + \gamma(\kappa(\theta)\sigma+\varepsilon)}. \tag{24}$$

The two equations above relate the elasticities of firm size and firm profits with respect to sales to technology parameters. Our calibration of the key parameter γ relies on using share of workers compensation (κ) and elasticity of firm size with respect to sales $\frac{d\ln L(\theta)}{d\ln y(\theta)}$ which are easily determined in the data. To this end we will use the following result derived from rearranging Equation (23).

Corollary 3. *The* γ-effect parameter satisfies,

$$\gamma = \frac{\left(1 - \frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\ln y(\theta)}\right)(1-\kappa)(1+\varepsilon)}{\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\ln y(\theta)}(1-\kappa+\kappa\sigma+\varepsilon) - (\sigma+\varepsilon)}. \tag{25}$$

We now move to the estimation procedure.

5.2 ESTIMATING κ AND γ

We first estimate the labor cost share $\kappa(\theta)$ directly from the data. Next we use the above results to estimate γ by matching the elasticity of firm size with respect to sales

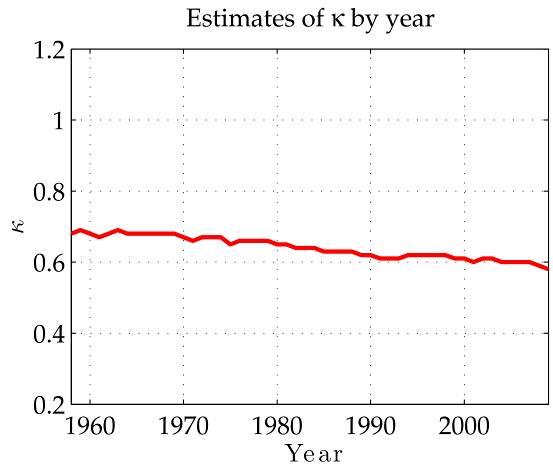


Figure 1: Estimates of κ by year.

$\mathrm{d}\ln y(\theta)/\mathrm{d}\ln L(\theta)$ to data.

Throughout the remainder of the analysis we treat $\kappa(\theta)$ as being constant across θ. This assumption is in part justified on the grounds that our estimates for κ do not vary much across industries or years (see below).

To estimate the share of labor costs κ, we need compensation data for both production workers and managerial employees. We take the NBER-CES Manufacturing Industry Database. This dataset has annual industry-level data on output, employment, and payroll for both production workers and non-production workers from 1958 to 2009. We map the production workers to workers and non-production workers to managers in our model.[17] We compute κ as the faction of wages of production workers over total payroll: *Production workers wages / Total payroll*. We estimate a value of $\widehat{\kappa} = 0.640$, with a confidence interval of $(0.580, .702)$.

Figure 1, plots the time trend of κ estimates by year. The–small–declining trend is consistent with the stylized fact that labor share is declining over time.[18] The estimate is also quite stable across industries, as shown in Figure 2. *Textile and Lumber* displays the highest values of κ close around 0.74, whereas *Measuring, Chemicals, and Machinery Computer* display the smallest values.

Next, we use firm level data from Compustat to estimate the elasticity of firm size

[17]From Berman, Bound, and Griliches (1994), production workers are defined as "workers (up through the working foreman level) engaged in fabricating, processing, assembling, inspecting and other manufacturing." This excludes supervisors above the line-supervisor lever, clerical, sales, office, professional, and technical workers. Non-production workers are defined as "personnel, including those engaged in supervision (above the working foreman level), installation and servicing of own product, sales, delivery, professional, technological, administrative etc."

[18]See, Samuel and Gilles (2003) and Karabarbounis and Neiman (2014).

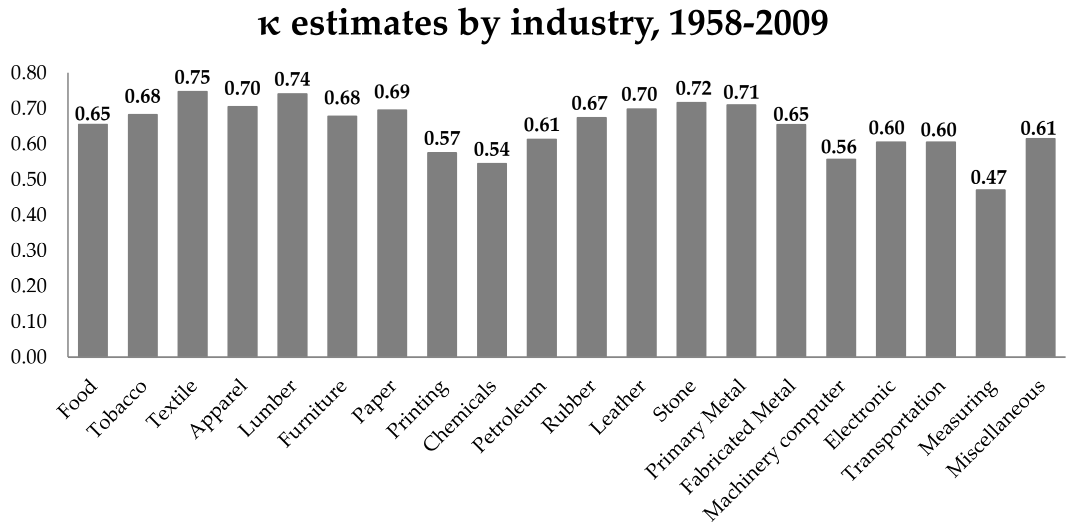


Figure 2: Estimates of κ by industry.

with respect to sales $\mathrm{d}\ln y(\theta)/\mathrm{d}\ln L(\theta)$. Our sample is constructed from U.S. COMPUSTAT firms that have non-missing values for sales and employee numbers between 1950 and 2012. Specifically, data of firm sales is taken from *Gross Sales* in the Income Statement; data on the total number of employees is taken from the *Employees* item. We restrict firms to have size not smaller than 30 employees so that there a sizable team for each manager to supervise. The total firm-year observations is 284,396. Among the sample, the average firm employee size is 9,326. We assume that the average size of managerial workforce is of 20 individuals.[19] Hence firm size in terms of number of non-managerial people (corresponding to our model $L(\theta)$) is defined as L = Total number of employees - 20. We estimate the following linear model:

$$\ln y_t(\theta_i) = \alpha_0 + \alpha_1 \ln L_t(\theta_i) + \varepsilon_{i,t}. \tag{26}$$

We estimate a value of $\alpha_1 = \widehat{\frac{\mathrm{d}\ln y(\theta)}{\mathrm{d}\ln L(\theta)}} = .902\ (.899,\ .904)$. The estimated elasticity is consistent with the known *making-do-with-less* effect which implies a coefficient smaller

[19]According to Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook, 2014-15 Edition, Top Executives, the number of employment for company chief executives with SOC 11-1011 is 330,500 in the year of 2012. Since we only focus on public firms, and suppose that the private firms and public firms share the same number of top executive positions, we estimate that per firm number of top executives is approximately 330,500 /2 /8250 = 20. For example, from the company annual reports, General Motor has 29 officers and directors in total; Amazon.com has 17; and Bank of America has 20. In addition, note that any proportional value of the managerial size relative to the workforce, will not bias the estimates.

than one.[20] The estimates are robust if we run the estimation by year and/or by industry. Figure 3 shows the estimates by year.

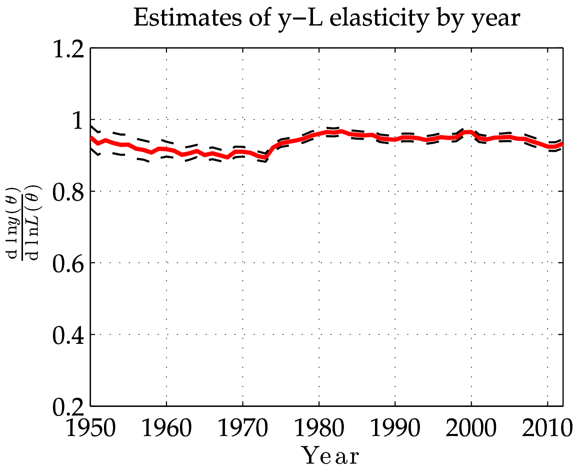


Figure 3: Estimates of $\widehat{\frac{\mathrm{d}\ln y(\theta)}{\mathrm{d}\ln L(\theta)}}$ by year.

Next we back out γ using Equation (25).

$$\widehat{\gamma} = \frac{\left(1 - \widehat{\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\ln y(\theta)}}\right)(1-\widehat{\kappa})(1+\varepsilon)}{\widehat{\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\ln y(\theta)}}(1-\widehat{\kappa}+\widehat{\kappa}\sigma+\varepsilon)-(\sigma+\varepsilon)} = 0.101. \tag{27}$$

Under the above estimated values of $\{\kappa, \gamma\}$, we match the estimated elasticity of firm size with respect to sales in Equation (23). Plugging into Equations (23) and (24), we obtain the implied elasticity of managerial compensation with respect to size as

$$\frac{d\,\widehat{\ln \pi(\theta)}}{d\ln L(\theta)} = \frac{1-\widehat{\kappa}+\widehat{\gamma}}{(1-\widehat{\kappa})(1+\varepsilon)+\widehat{\gamma}(\sigma+\varepsilon)} = 0.415. \tag{28}$$

In our environment, firm profits (or a fraction of them) are mapped to managerial compensation. In this way, our estimates are consistent with micro-level evidence that the elasticity of CEO pay with respect to firm size is approximately 0.30.[21]

Lastly, we estimate β from the definition of $\kappa(\theta)$.

[20] See Lazear, Shaw, and Stanton (2013)

[21] Roberts (1956), Fox (1978), Baker, Jensen, and Murphy (1988), Frydman and Saks (2010)); See also Lewellen and Huntsman (1970) and Conyon and Murphy (2000). Gabaix and Landier (2008) use ExecuComp data from 1992 to 2004 and find Elasticity of CEO pay to firm size estimates of approximately equal to 0.33.

$$\kappa(\theta) = \frac{wL(\theta)}{y(\theta)} = \frac{MP_{L(\theta)}L(\theta)}{y(\theta)} = \frac{1}{1 + \frac{\beta}{1-\beta}\left(\frac{n\theta}{L}\right)^{\rho}} = 0.64. \tag{29}$$

From NBER-CES Manufacturing Industry Database, the average team size defined by the number of production worker per non-production worker is estimated to be 3.58. We normalize the managers' effective effort level $n\theta = 1$, i.e. the data counterpart of $\frac{n\theta}{L}$ is $1/3.58$. Equation (29) thus implies a unique value $\beta = 0.594$. Table 1 summarizes the parameter calibration of the benchmark model.

Table 1: Benchmark Parameter Calibration.

Parameter	Symbol	Value
Frisch Elasticity	ε	0.75
Inverse of Pareto tail parameter of θ	a	1.5
Degree of substitutability of inputs	ρ	0.75
Scale-of-operations effect	γ	0.101
Share of labor costs to total sales	κ	0.640
Labor share parameter in production	β	0.594

We next compute the values for optimal taxes.

6 OPTIMAL TOP TAX RATES

Plugging in the parameters of the benchmark calibration from Table 1 into our top tax formula (12), we obtain that the optimal tax rate implied in our environment:

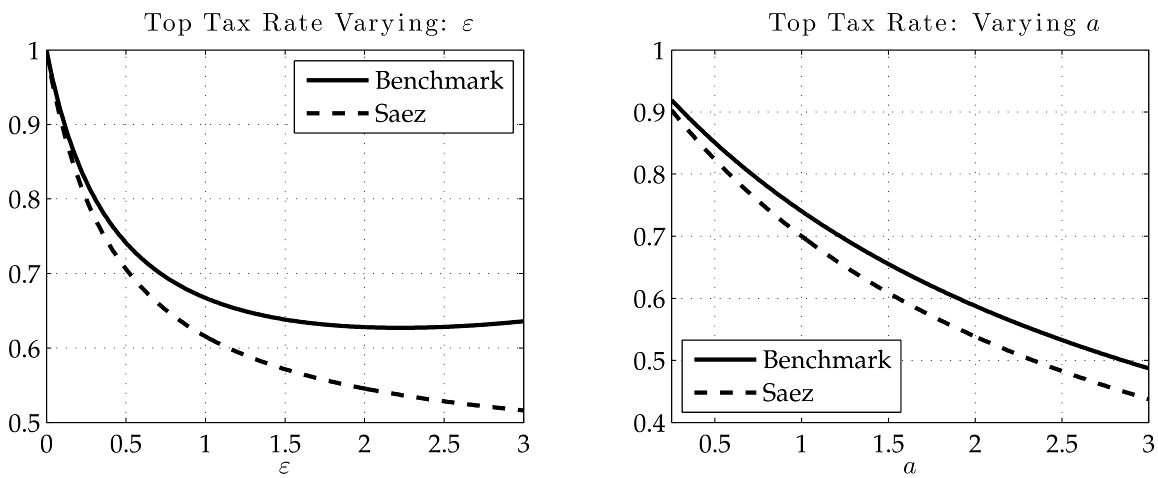
$$\widehat{\tau} = \frac{1}{1 + a\left[\frac{1}{\varepsilon} + 1 + \frac{\widehat{\gamma}}{1-\widehat{\kappa}}\left(\frac{1}{\varepsilon} + 1 - \rho\widehat{\kappa}\right)\right]^{-1}} = 65.52\%. \tag{30}$$

Table 2 presents results of implied optimal tax rates for $\gamma = \widehat{\gamma}$ in our environment. Under different key parameter values of a and ε which are common in the literature.[22] For comparison, we also include the case when $\gamma = 0$, which corresponds to the tax formula in Saez (2001). Notice that the impact of the scale-of-operations effect is large: optimal top tax rates are nearly 10% higher with respect to the case where $\gamma = 0$.

[22]Note that for any choice of ε the parameter γ needs to be recomputed as discussed in the previous section.

Table 2: Top Tax Rates.

	$a = 1.5$		$a = 2.00$	
ε	0.75	1.00	0.75	1.00
$\gamma > 0$	0.65	0.62	0.59	0.55
$\gamma = 0$	0.61	0.57	0.54	0.50
$\gamma > 0 / \gamma = 0$	1.08	1.08	1.09	1.10



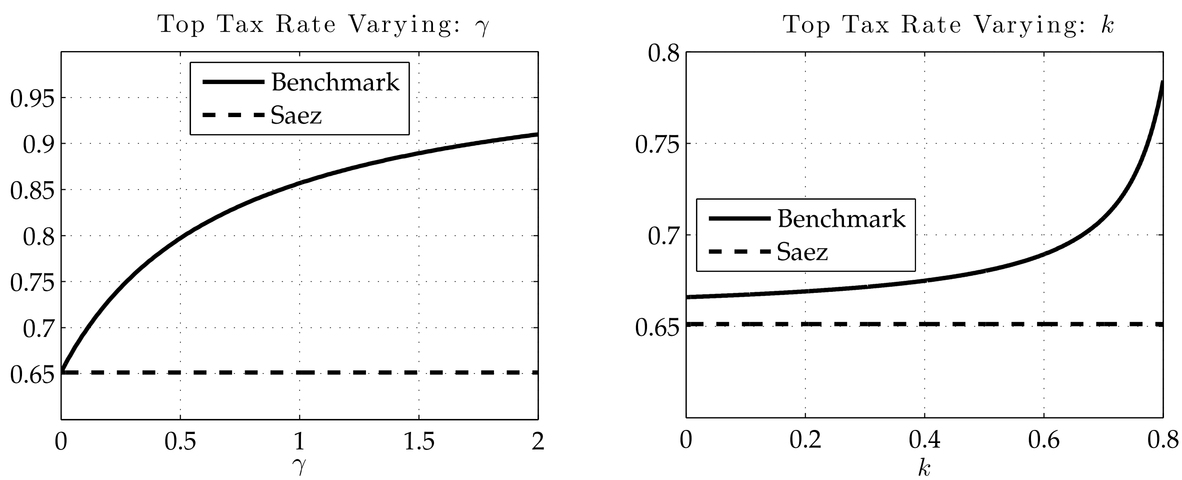
(a) Varying ε.

(b) Varying a.

Figure 4: Comparative Statics of Optimal Top Tax Rates: Benchmark Environment ($\gamma > 0$) vs. Saez (2001) ($\gamma = 0$).

Figures 4(a) and 4(b) expand the analysis of Table 2 regarding the effect of ε and a. For expositional purposes, we identify the case of $\gamma = 0$ with Saez (2001). First, we observe that differences between our benchmark environment and a standard environment with $\gamma = 0$ increase as ε increases. In particular, consider a value of ε close to 3 as in Prescott (2004). In this case, the $\gamma = 0$ tax approaches optimal tax rate observed in the U.S. while the environment with a positive scale-of-operations displays marginal rate well above 60%.

In Figure 4(b) we look at the effect of a, the tail of the Pareto distribution on optimal rates. Analogously to the comparative statics for ε, we observe that the tax with a $\gamma > 0$ displays somewhat slightly less sensitivity than for the $\gamma = 0$ case. For instance, for $a = 1.5$ a model in line with Saez (2001) would prescribe a top tax rate equal to 60%. On the other hand, such tax rate is only optimal in our environment with $a = 1.8$, i.e. with a thinner tail of the skill distribution.



(a) Varying γ.

(b) Varying κ.

Figure 5: Comparative Statics of Optimal Top Tax Rates: Benchmark Environment ($\gamma > 0$) vs Saez (2001) ($\gamma = 0$)

.

Finally, in Figures 5(a) and 5(b) we consider the sensitivity of our tax rates to the technological parameters γ and κ, respectively. Note that tax rates are increasing in these parameters. Also, optimal tax rates are particularly sensitive to small variations in κ and γ around our benchmark estimates.

6.1 ROBUSTNESS AND COUNTERFACTUALS

`To be Completed.`

7 SIMULATIONS FOR THE ENTIRE TAX CODE

`To be Completed.`

8 CONCLUSION

The title of this paper is a reference to the tough provoking novel of Rand (1957).[23] 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

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

earners are modeled as managers whose effort and skill contribute–together with hired labor–to generate output. Our key finding is that tax rates should be substantially higher than what we see in the data as well as in previous estimates in the literature. This result is robust to a large range of parameter specifications. Methodologically this paper highlights how to consider firm level data to determine key parameters relevant for income taxation. There are two logical next steps. The first one is to determine the impact of a hierarchical organizations rather than a single manager technology for optimal taxes. The second step is to push further in identifying the distribution of managerial skill given the observed distribution of firm size.

REFERENCES

Adams, C. P. (2006): "Optimal team incentives with CES production," *Economics Letters*, 92(1), 143–148.

Altonji, J. G. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, 94(3), S176–S215.

Baker, G. P., M. C. Jensen, and K. J. Murphy (1988): "Compensation and Incentives: Practice vs. Theory," *The Journal of Finance*, 43(3), 593–616.

Bartelsman, E. J., and M. Doms (2000): "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3), 569–594.

Berman, E., J. Bound, and Z. Griliches (1994): "Changes in the demand for skilled labor within US manufacturing: evidence from the annual survey of manufactures," *The Quarterly Journal of Economics*, 109(2), 367–397.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2012): "Consumption inequality and family labor supply," Discussion paper, National Bureau of Economic Research.

Cagetti, M., and M. D. Nardi (2006): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 114(5), 835–870.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins," *The American Economic Review*, 101(3), 471–475.

Conyon, M. J., and K. J. Murphy (2000): "The Prince and the Pauper? CEO Pay in the United States and United Kingdom," *The Economic Journal*, 110(467), 640–671.

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

Diamond, P. A., and E. Saez (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25(4), 165–90.

Farhi, E., and I. Werning (2013): "Insurance and taxation over the life cycle," *The Review of Economic Studies*, 80(2), 596–635.

French, E. (2005): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, 72(2), 395–427.

Frydman, C., and R. E. Saks (2010): "Executive Compensation: A New View from a Long-Term Perspective, 1936âĂŞ2005," *Review of Financial Studies*, 23(5), 2099–2138.

Gabaix, X., and A. Landier (2008): "Why has CEO Pay Increased So Much?," *The Quarterly Journal of Economics*, 123(1), 49–100.

Gentry, W. M., and R. G. Hubbard (2004): "The effects of progressive income taxation on job turnover," *Journal of Public Economics*, 88(11), 2301–2322.

Golosov, M., M. Troshkin, and A. Tsyvinski (2011): "Optimal dynamic taxes," Discussion paper, National Bureau of Economic Research.

Holmstrom, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13(2), 324–340.

Karabarbounis, L., and B. Neiman (2014): "The global decline of the labor share," *The Quarterly Journal of Economics*, 129(1), 61–103.

King, R. G., and S. T. Rebelo (1999): "Resuscitating real business cycles," *Handbook of macroeconomics*, 1, 927–1007.

Lazear, E. P., K. L. Shaw, and C. Stanton (2013): "Making Do With Less: Working Harder During Recessions," Working Paper 19328, National Bureau of Economic Research.

Lewellen, W. G., and B. Huntsman (1970): "Managerial Pay and Corporate Performance," *The American Economic Review*, 60(4), 710–720.

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9(2), 508–523.

MaCurdy, T. E. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting," *The Journal of Political Economy*, 89(6), 1059–1085.

MANKIW, G., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), 147–174.

MAYER, T. (1960): "The Distributions of Ability and Earnings," *Review of Economics and Statistics*, pp. 189–195.

MIRRLEES, J. A. (1971): "An exploration in the theory of optimum income taxation," *The review of economic studies*, 38(2), 175–208.

PAVAN, A., I. SEGAL, AND J. TOIKKA (2012): "Dynamic mechanism design," *Unpublished man*.

PIKETTY, T., AND E. SAEZ (2003): "Income Inequality In The United States, 1913-1998," *The Quarterly Journal of Economics*, 118(1), 1–39.

PRESCOTT, E. C. (2004): "Why Do Americans Work So Much More Than Europeans?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1), 2–13.

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

RAND, A. (1957): *Atlas Shrugged*. Penguin.

ROBERTS, D. R. (1956): "A General Theory of Executive Compensation Based on Statistically Tested Propositions," *The Quarterly Journal of Economics*, 70(2), 270–294.

ROSEN, S. (1982): "Authority, control, and the distribution of earnings," *The Bell Journal of Economics*, pp. 311–323.

SADKA, E. (1976): "On Income Distribution, Incentive Effects and Optimal Income Taxation," *Review of Economic Studies*, 43(1), 261–268.

SAEZ, E. (2001): "Using elasticities to derive optimal income tax rates," *The Review of Economic Studies*, 68(1), 205–229.

SAEZ, E. (2003): "The effect of marginal tax rates on income: a panel study of 'bracket creep'," *Journal of Public Economics*, 87(5-6), 1231–1258.

SAMUEL, B., AND S.-P. GILLES (2003): "Explaining Movements in the Labor Share," *The B.E. Journal of Macroeconomics*, 3(1), 1–33.

SLEMROD, J. (2000): *Does Atlas Shrug?: The Economic Consequences of Taxing the Rich*. Harvard University Press.

APPENDIX

A PROOFS

A.1 PROOFS OF SECTION 3.1

We begin with the proof of Lemma 1.

Proof. The first order condition for $\theta' = \theta$ to be a local maximum of $M(\theta';\theta)$ is $M_1(\theta;\theta) = 0$. This is equivalent to (5). Differentiating the first order condition $M_1(\theta;\theta) = 0$ with respect to θ gives

$$M_{11}(\theta;\theta) + M_{12}(\theta;\theta) = 0.$$

Hence, the second order condition $M_{11}(\theta;\theta) \leq 0$ can be written as $-M_{12}(\theta;\theta) \leq 0$, which gives (6). □

We now look at the proof of Proposition 1.

Proof. We want to show that if (5) and (6) hold, then $M_1(\theta';\theta)$ has the sign of $(\theta - \theta')$. This obviously implies that $M(\theta';\theta)$ attains a global maximum at $\theta' = \theta$ so that (4) holds. First note that

$$M_1(\theta';\theta) = c'(\theta') - v'(n(\theta';\theta)) \left[n_y(\theta';\theta) y'(\theta') + n_L(\theta';\theta) L'(\theta') \right]. \tag{31}$$

We also have that (5) at θ' gives

$$c'(\theta') = v'(n(\theta';\theta')) \left[n_y(\theta';\theta') y'(\theta') + n_L(\theta';\theta') L'(\theta') \right]. \tag{32}$$

Using (32) into (31) gives

$$M_1(\theta';\theta) = J(\theta';\theta') - J(\theta';\theta), \tag{33}$$

where $J(\theta';\theta) \equiv v'(n(\theta';\theta)) \left[n_y(\theta';\theta) y'(\theta') + n_L(\theta';\theta) L'(\theta') \right]$. Clearly

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

$$L'(\theta') \left[v''(n(\theta';\theta')) n_L(\theta';\theta') n_\theta(\theta';\theta') + v'(n(\theta';\theta')) n_{L\theta}(\theta';\theta') \right],$$

so that (6) gives that $J_2(\theta';\theta') \leq 0$. Then (33) implies that $M_1(\theta';\theta) \geq 0$ if and only if $\theta' \leq \theta$, which completes the proof. □

Finally, we present the proof for Proposition 2.

Proof. `To be Completed.` □

A.2 PROOF OF PROPOSITION 3

After integrating by parts, the Lagrangian to the planner's problem can be written as

$$\mathcal{L} = \int_{\Theta} \Psi\left(U(\theta)\right) f(\theta) d\theta$$

$$- \int_{\Theta} \left[U(\theta)\mu'(\theta) - \mu(\theta)v'(n(\theta;\theta))n_{\theta}(\theta;\theta)\right] d\theta$$

$$+\lambda^{RC}\left(\int_{\Theta} \left[y(\theta) - U(\theta) - v(n(\theta;\theta))\right] f(\theta) d\theta\right)$$

$$+\lambda^{LM}\left(1 - \int_{\Theta} L(\theta) f(\theta) d\theta\right),$$

where $\mu(\theta)$ is the multiplier on (7), λ^{RC} is the multiplier on (8) and λ^{LM} is the multiplier on (9). First order conditions are

$$\mu'(\theta) = \left(\Psi'(U(\theta)) - \lambda^{RC}\right) f(\theta), \tag{34}$$

$$\mu(\theta)\left[-v''(n(\theta;\theta))n_y(\theta;\theta)n_{\theta}(\theta;\theta) - v'(n(\theta;\theta))n_{\theta y}(\theta;\theta)\right] = \lambda^{RC}\left[1 - v'(n(\theta;\theta))n_y(\theta;\theta)\right] f(\theta), \tag{35}$$

$$\mu(\theta)\left[-v''(n(\theta;\theta))n_L(\theta;\theta)n_{\theta}(\theta;\theta) - v'(n(\theta;\theta))n_{\theta L}(\theta;\theta)\right] = -\lambda^{RC}v'(n(\theta;\theta))n_L(\theta;\theta)f(\theta) - \lambda^{LM} f(\theta). \tag{36}$$

Integrating (34) between θ and $\bar{\theta}$ and using the transversality condition $\mu(\bar{\theta}) = 0$ we get

$$\mu(\theta) = \int_{\theta}^{\bar{\theta}} \left(\lambda^{RC} - \Psi'(U(\theta))\right) f(\theta) d\theta. \tag{37}$$

Evaluating (37) at $\underline{\theta}$ gives

$$\lambda^{RC} = \int_{\Theta} \Psi'(U(\theta)) f(\theta) d\theta. \tag{38}$$

So using (38) into (37) we obtain

$$\mu(\theta) = (1 - F(\theta))\left(D(\underline{\theta}) - D(\theta)\right), \tag{39}$$

where $D(\theta) \equiv \frac{1}{1-F(\theta)} \int_{\theta}^{\infty} \Psi'\left(U(\theta)\right) dF(\theta)$. Also, (35) can be written as

$$\mu(\theta)v'(n(\theta;\theta))n_y(\theta;\theta)\left[-\frac{v''(n(\theta;\theta))}{v'(n(\theta;\theta))}n_{\theta}(\theta;\theta) - \frac{n_{\theta y}(\theta;\theta)}{n_y(\theta;\theta)}\right] = \lambda^{RC}\left[1 - v'(n(\theta;\theta))n_y(\theta;\theta)\right] f(\theta). \tag{40}$$

So using (38), (39) and (10) into (40) gives the result.

A.3 PROOF OF COROLLARY 1

First note that given Assumption 2 we have

$$n(y,L;\theta) = \left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho} - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right]^{\frac{1}{\rho}}. \tag{41}$$

Taking derivatives from (41) we obtain

$$n_\theta(y,L;\theta) = \left(-\frac{1}{\theta}\right)\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho} - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right]^{\frac{1}{\rho}-1}\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}(1+\gamma) - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right], \tag{42}$$

$$n_y(y,L;\theta) = \left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho} - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right]^{\frac{1}{\rho}-1}\frac{1}{\beta y}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}, \tag{43}$$

$$\begin{aligned} n_{\theta y}(y,L;\theta) = &\left(-\frac{1-\rho}{\theta}\right)\frac{1}{\beta y}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho} - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right]^{\frac{1}{\rho}-2} \\ &\times\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}(1+\gamma) - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right] \\ &-\frac{\rho}{\theta}\left[\frac{1}{\beta}\left(\frac{y}{\theta g(\theta)}\right)^{\rho} - \frac{1-\beta}{\beta}\left(\frac{L}{\theta}\right)^{\rho}\right]^{\frac{1}{\rho}-1}\frac{1}{\beta y}\left(\frac{y}{\theta g(\theta)}\right)^{\rho}(1+\gamma). \end{aligned} \tag{44}$$

Using (41) and (42) we get

$$\frac{n_\theta(y,L;\theta)}{n(y,L;\theta)} = -\frac{1}{\theta}\frac{\left(\frac{y}{g(\theta)}\right)^{\rho}(1+\gamma) - (1-\beta)L^{\rho}}{\left(\frac{y}{g(\theta)}\right)^{\rho} - (1-\beta)L^{\rho}}, \tag{45}$$

and

$$\frac{n_{\theta y}(y,L;\theta)}{n_y(y,L;\theta)} = -\frac{1}{\theta}\left[(1-\rho)\frac{\left(\frac{y}{g(\theta)}\right)^{\rho}(1+\gamma) - (1-\beta)L^{\rho}}{\left(\frac{y}{g(\theta)}\right)^{\rho} - (1-\beta)L^{\rho}} + \rho(1+\gamma)\right]. \tag{46}$$

Now notice that by Assumption 2 we have

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

Then

$$\frac{\left(\frac{y}{g(\theta)}\right)^{\rho}(1+\gamma) - (1-\beta)L^{\rho}}{\left(\frac{y}{g(\theta)}\right)^{\rho} - (1-\beta)L^{\rho}} = 1 + \gamma\left(1 + \frac{1-\beta}{\beta}\left(\frac{L}{\theta n}\right)^{\rho}\right) \tag{47}$$

Also notice that Assumption 2 implies

$$\frac{y_L(\theta)L(\theta)/y(\theta)}{y_{\theta n}\theta n/y(\theta)} = \frac{1-\beta}{\beta}\left(\frac{L}{\theta n}\right)^{\rho},$$

or

$$\frac{\kappa(\theta)}{1-\kappa(\theta)} = \frac{1-\beta}{\beta}\left(\frac{L}{\theta n}\right)^{\rho}, \tag{48}$$

where $\kappa(\theta) \equiv y_L(\theta)L(\theta)/y(\theta)$ denotes the *share of labor costs to total sales* for manager θ. Using (48) in (47) gives

$$\frac{\left(\frac{y}{g(\theta)}\right)^{\rho}(1+\gamma)-(1-\beta)L^{\rho}}{\left(\frac{y}{g(\theta)}\right)^{\rho}-(1-\beta)L^{\rho}} = 1+\gamma\left(1+\frac{\kappa(\theta)}{1-\kappa(\theta)}\right). \tag{49}$$

So using (45), (46) and (49) into the formula from Proposition 3 gives the result.

A.4 PROOF OF LEMMA 2

We start with some preliminary results. Using (17) and the definition of k we can write

$$1-\kappa = \frac{\theta n}{L}\frac{h'}{h}. \tag{50}$$

Also, let $f(\theta n, L) = Lh(\theta n/L)$. Define the elasticity of substitution between θn and L as

$$\sigma \equiv -\frac{d\ln(L/\theta n)}{d\ln(f_2/f_1)}. \tag{51}$$

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

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

Straightforward algebra leads to:

$$\frac{1}{\sigma} = -\left(\frac{\theta n}{L}\right)\frac{h''}{h'}\left(1-\frac{\theta n}{L}\frac{h'}{h}\right)^{-1}. \tag{52}$$

So combining (50) and (52) we obtain

$$\frac{1}{\sigma} = -\left(\frac{\theta n}{L}\right)\frac{h''}{h'}\frac{1}{\kappa}. \tag{53}$$

For now on we will work with (53) for the elasticity of substitution.

We now move to the proof of Lemma 2. First we look at equation (19). Let $M(\theta)$ be the value function to problem (16). Applying the envelope theorem we get:

$$\begin{aligned}\frac{dM(\theta)}{d\theta} &= (1-\tau)\chi\frac{d\pi(\theta)}{d\theta} \\ &= (1-\tau)\chi\left[g'(\theta)Lh + g(\theta)h'n\right],\end{aligned}$$

so that

$$\begin{aligned}\frac{d\pi(\theta)}{d\theta}\frac{\theta}{\pi} &= \frac{\theta}{\pi}\left[g'(\theta)Lh + g(\theta)h'n\right] \\ &= \frac{g(\theta)Lh}{\pi}\left[\frac{g'(\theta)}{g(\theta)}\theta + \frac{\theta n}{L}\frac{h'}{h}\right]\end{aligned}$$

by definition of κ and using (50) we get $\frac{d\ln\pi}{d\ln\theta} = \frac{1}{1-\kappa}(\gamma + 1 - \kappa)$ which gives (19). We now move to equation (20). Differentiating the system defined by (17) and (18) we get

$$\frac{d\ln L}{d\ln\theta} - \frac{d\ln n}{d\ln\theta} = 1 - \theta\frac{g'}{g}\left(\frac{L}{\theta n}\right)^2\frac{1}{h''}\left(h - \frac{\theta n}{L}h'\right), \tag{54}$$

$$\frac{n\theta}{L}gh''\frac{d\ln L}{d\ln\theta} + \left[\frac{n}{\theta}\frac{v''}{(1-\tau)\chi} - \frac{\theta n}{L}gh''\right]\frac{d\ln n}{d\ln\theta} = \tag{55}$$

$$g'\theta h' + gh' + gh''\frac{n\theta}{L}. \tag{56}$$

Combining (54) and (55),

$$\frac{n}{\theta}\frac{v''}{(1-\tau)\chi}\frac{d\ln n}{d\ln\theta} + \frac{n\theta}{L}gh''\left(1 + \frac{\gamma\sigma}{1-\kappa}\right) = \tag{57}$$

$$g'\theta h' + gh' + gh''\frac{n\theta}{L}, \tag{58}$$

where we used (50), (53) and the definition of γ on the last term. Further rearranging using (50) and (53) gives

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

Using the first order condition (18) we have

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

From Assumption 1 we derive (20). For Equation (21), using (20) into (54) gives

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

Then applying (50) and (53) and rearranging gives (21). Finally for equation (22): by (15) we have

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

Using (50) gives

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

Plugging in (22) and (21) gives the result.

B ON THE ELASTICITY OF MANAGERIAL INCOME

In this section we derive an expression for the elasticity of managerial income π with respect to the after-tax top rate $(1-\tau)$.

Proposition 6. *Let* $\widehat{e}\equiv\frac{\partial\ln\pi}{\partial\ln(1-\tau)}$. *Then* $\widehat{e}=1+\varepsilon$.

Proof. By definition π we have that

$$\frac{\partial\pi}{\partial(1-\tau)} = \chi\left[gLh-wL\right]+(1-\tau)\chi\left[gh-gh'\frac{\theta n}{L}-w\right]\frac{\partial L}{\partial(1-\tau)}+(1-\tau)\chi g\theta h'\frac{\partial n}{\partial(1-\tau)}. \tag{59}$$

Differentiating the first order conditions from the manager's problem in (17) and (18) with respect to n, L and $(1-\tau)$ we get

$$\frac{\partial n}{\partial(1-\tau)}-\frac{n}{L}\frac{\partial L}{\partial(1-\tau)} = 0, \tag{60}$$

and

$$\left[\frac{v''}{\chi g\theta h'}-(1-\tau)\frac{\theta}{L}\frac{h''}{h'}\right]\frac{\partial n}{\partial(1-\tau)}+(1-\tau)\frac{\theta n}{L^2}\frac{h''}{h'}\frac{\partial L}{\partial(1-\tau)} = 1. \tag{61}$$

Combining (60) and (61) we obtain

$$\frac{\partial L}{\partial(1-\tau)} = \frac{L}{n}\frac{\chi g\theta h'}{v''}, \tag{62}$$

$$\frac{\partial n}{\partial(1-\tau)} = \frac{\chi g\theta h'}{v''}. \tag{63}$$

Using (62), (63) and (18) into (59) gives

$$\frac{\partial \pi}{\partial(1-\tau)} = \chi\left[gLh - wL\right] + \chi\left[ghL - gh'\theta n - wL\right]\frac{v'}{v''n} + \chi g\theta h'\frac{v'}{v''},$$

or

$$\frac{\partial \pi}{\partial(1-\tau)} = \chi\left[gLh - wL\right] + \chi\left[ghL - gh'\theta n - wL\right]\varepsilon + \chi g\theta h'\varepsilon n.$$

Rearranging gives

$$\frac{\partial \pi}{\partial(1-\tau)} = \chi\left(gLh - wL\right)(1+\varepsilon). \tag{64}$$

So using the definition of π once more gives

$$\frac{\partial \pi}{\partial(1-\tau)}\frac{1-\tau}{\pi} = 1+\varepsilon.$$

as we wanted to show. □