

TAXING TOP CEO INCOMES

Laurence Ales

Christopher Sleet*

Tepper School of Business, Carnegie Mellon University

This Version: May 24, 2016

Abstract

We use a firm-CEO assignment framework to model the market for CEO effective labor. In the model's equilibrium more talented CEOs match with and supply more effort to larger firms. Taxation of CEO incomes affects the equilibrium pricing of CEO effective labor and, hence, spills over and affects firm profits. Absent the ability to tax profits or a direct concern for firm owners, a standard prescription for high marginal income taxes emerges. However, given such an ability or concern the optimal marginal tax rates are much lower. (*JEL D31, H21, H24, M12, M52*)

*Ales: ales@cmu.edu Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; Sleet: csleet@andrew.cmu.edu. Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. We thank seminar participants at Cornell University, Federal Reserve Bank of Philadelphia, Stanford University, SUNY-Stony Brook and the Wharton School. We especially thank Caroline Hoxby and Florian Scheuer for helpful comments and discussion.

1 Introduction

What should the marginal tax rate on top income earners be? Recent research suggests that it should be high, perhaps as high as 70 per cent or 80 per cent. Underpinning these numbers is the well known Diamond-Saez formula that relates the optimal marginal tax rate on top incomes to the elasticity of taxable income and a property of the right tail of the earnings distribution. This formula is derived under the assumption that the policymaker's objective is to maximize tax revenues derived from top earners. It abstracts from any positive impact of the efforts of these earners on the incomes of other agents or on tax revenues collected from other sources.¹ Our paper departs from this research by taking seriously the idea that the activities of high earning CEOs, an important group of top earners, have positive spillovers for others.² We use a firm-CEO assignment framework to model the market for CEO effective labor. [Gabaix and Landier \(2008\)](#) and [Terviö \(2008\)](#) have shown that such a framework is valuable for understanding recent growth in CEO incomes and the interaction of firm and CEO attributes in shaping this growth. We show that in an assignment model (augmented with an intensive CEO effort margin), the taxation of CEO incomes affects the equilibrium pricing of CEO effective labor and, hence, spills over and affects firm profits.³ If the policymaker has no social concern for profits, then the classic Diamond-Saez formula remains intact. Otherwise, the optimal marginal tax rate on CEO incomes is modified downwards. At our benchmark parameterization, a full reform of CEO income and profit taxation entails an optimal marginal tax on top CEO incomes of about 15 per cent.

The basic CEO to firm assignment model supposes one-to-one matching of differently talented CEOs to differently sized firms. As noted, we augment this model with an intensive CEO effort margin (and with taxes). The equilibrium features assortative matching of CEO talent with firm size. More talented CEOs match with and supply more effort (and more effective labor) to larger firms. The indivisibility

¹The literature has considered negative impacts such as rent seeking and has modified the basic Diamond-Saez formula accordingly, see [Piketty et al. \(2014\)](#). [Saez \(2001\)](#) also modifies the formula to allow for social concern for top income earners.

²[Bakija et al. \(2012\)](#) report that 60 per cent of the top 0.1 per cent of earners by income are executives, managers, supervisors and financial professionals. Our focus on CEOs is also partially motivated by the availability of high quality uncensored data on the incomes of CEOs. However, we believe that our theoretical and broad quantitative insights carry over to other “superstar” buyer-seller relationships that generate high incomes for sellers. In particular, they are applicable to hedge fund managers and other financial professionals who make complex decisions.

³As we elaborate the nature of this spillover is subtle. There is no “marginal mispricing”. Instead the infra-marginal rents accruing to a given firm are impacted by the effort of CEOs at other, smaller firms and, hence, by the taxes placed on these CEOs.

of the CEO position prevents combinations of less talented CEOs replacing more talented ones and equalizing the price for effective CEO labor across firms. On the other hand, competition amongst similarly talented CEOs for a position prevents any given CEO from extracting all of the surplus from a firm. In equilibrium the price of a unit of CEO effective labor equals the marginal product of CEO effective labor at the firm at which this unit is the last hired. Since the marginal product of CEO effective labor is increasing in firm size, the matching of more talented CEOs to larger firms enhances the dispersion of top CEO incomes. Even if there is relatively little dispersion in CEO talent, large variations in firm size can translate into large variations in top CEO incomes. However, since a CEO is only paid the marginal product of her effective labor on the last unit she sells to her firm (with infra-marginal units priced by and paid the marginal product of smaller firms), claimants to firm profits capture some surplus. In this setting, an increase in the marginal tax rate above a threshold income induces an upwards adjustment in the pricing schedule for effective labor. This in turn redistributes from firm profits to CEO incomes and, hence, CEO income tax revenues. If the policy maker is concerned only with maximizing income tax revenues, then this redistribution provides a motive for higher marginal income taxes. If, on the other hand, the policymaker is indifferent to the allocation between income tax revenues and firm profits (because the latter can be taxed, because tax receipts and firm claimants are equally valued or because depressing firm profits has adverse effects on firm creation), then no such redistribution motive for higher marginal taxes exists.

We first consider the optimal linear tax rate across a range of top CEO incomes. The classic Diamond-Saez tax formula relates this tax rate to the elasticity of taxable income and a right tail property of the income distribution. Since this formula assumes the policymaker maximizes revenues from the income taxation of top earners and is silent on how these earners generate income, it applies in our setting, but only if the policymaker attaches no social value to firm profits. An alternative Mirrleesian formula relates the optimal tax rate to the elasticity of worker effort and a right tail property of the worker talent distribution. Absent income effects on CEO effort, it holds if the policymaker places equal weight on CEO income tax revenues and firm profits (for example because these can be taxed at 100 per cent) and, thus, maximizes surplus not paid to CEOs.⁴ Since in our baseline model firm profits are pure rents, 100 per cent taxation of profits is weakly optimal. Thus, the Mirrleesian formula is relevant if a comprehensive reform of profit and

⁴It can also be interpreted as adjusting the Diamond-Saez formula to purge out adjustments in CEO income in response to tax rises that come at the expense of firm profits.

CEO income taxation is implemented. If institutional constraints or other economic frictions restrain the optimal profit tax rate and the social marginal value of profit, then neither the Diamond-Saez nor the Mirrleesian formulas are valid. In contrast, in the standard public finance setting with heterogeneous workers selling effective labor to a competitive firm, all surplus not paid out to high income workers is captured in taxes. Thus, there is no distinction between maximizing this surplus and maximizing income tax revenues and so both formulas hold.

The simpler linear tax setting just described provides intuition for the analysis of optimal nonlinear taxation. Specifically, and analogous to the linear setting, when zero social weight is placed on firm profits a conventional-looking optimal tax formula in terms of the CEO income elasticity and the local Pareto coefficient of CEO incomes emerges. In this case the elasticity is adjusted to take into account the effect of a higher tax rate at a given income on the equilibrium pricing schedule for effective labor and, hence, the incomes received by more talented CEOs earning larger amounts at bigger firms.⁵ When the policymaker weights CEO income tax revenues and firm profits equally, then a conventional Mirrleesian optimal tax formula in terms of the CEO's effort elasticity and the Pareto coefficient of CEO talent arises. We use (nonlinear) optimal tax formulas expressed in terms of the underlying (structural) talent and firm size asset distributions to quantitatively characterize optimal taxes on CEOs across a range of high incomes and firm profit weights. To that end we extend an empirical strategy of [Terviö \(2008\)](#) to allow for an intensive effort margin. If a comprehensive reform of income and firm profit taxation is implemented, then income taxes and firm profits are equally weighted and optimal marginal tax rates decline from around 20 per cent at an income of \$10 million to about 10 per cent at an income of \$100 million. If a partial reform of CEO income taxation occurs holding profit taxes close to their statutory values in the US of about 60 per cent and abstracting from direct concern for profit recipients or impact of profits on firm entry then optimal tax rates decline from 34 per cent to 27 per cent over a similar income range. In either case they are very different from the rates of 70 per cent to 80 per cent recently advocated in the literature.

The remainder of the paper proceeds as follows. Following a brief literature review, Section 2 provides our baseline assignment model of CEO incomes and firm profits. It gives an initial characterization and formulation of equilibrium suitable for tax analysis. Section 3 analyzes the optimal linear tax rate across a range of CEO top incomes. Section 4 considers optimal non-linear taxation. It provides

⁵Analogous to before, a higher marginal tax rate at a given income causes the incomes of more talented CEOs to rise at the expense of firm profits.

formulas that characterize the fully optimal non-linear tax. Section 5 uses these formulas and data on CEO compensation and firm values to calculate the optimal nonlinear tax function for CEOs over a range of incomes and weights on profits. Section 6 concludes. Appendices contain proofs and additional details.

Related literature Our paper contributes to a literature in normative public finance that considers the income tax implications of fiscal spillovers.⁶ [Stiglitz \(1982\)](#) analyzes optimal income taxation with endogenously determined wages. In this framework diminishing returns with respect to a given skill type’s labor input and imperfect substitutability between skill types implies that the wage distribution is endogenous to tax policy. [Rothschild and Scheuer \(2013\)](#) extends this analysis to a rich worker-occupation assignment setting. [Ales et al. \(2015\)](#) explore the policy implications of technical change in such a setting. In these models, many workers match with an occupation and spillovers operate through the collective effect of worker labor supply and occupation choices on occupational output prices.⁷ [Rothschild and Scheuer \(2014\)](#) extend the framework of these papers by divorcing the private return from an occupation from the social return and, hence, incorporating explicit externalities at the level of occupations. Thus, they introduce a motive for corrective Pigouvian taxation.⁸ In a different direction, [Stantcheva \(2014\)](#) provides a rich model that combines informational frictions between workers and firms as well as between workers and the government. Taxes perturb the menu of contracts offered by firms and, hence, induce redistributions amongst different worker types (rather than between CEOs and firm owners). None of the preceding papers focuses on top earners or CEOs per se. While we abstract from the externalities assumed in [Rothschild and Scheuer \(2014\)](#) and the private information between firms and workers (CEOs) assumed in [Stantcheva \(2014\)](#), we enrich the analysis with heterogeneity on the side of firms and an assignment structure.

Recent work on the taxation of top earners has emphasized positive fiscal spillovers from the taxation of top earners. In the context of high income earners, [Saez et al. \(2012\)](#) extend the basic [Diamond and Saez \(2011\)](#) formula to accommodate

⁶We use the term “fiscal spillover” broadly to describe situations in which the taxation of one population of agents or source of income spills over and affects other populations or income sources. These spillovers encompass classical general equilibrium tax incidence effects, externalities and income shifting across tax bases.

⁷Specifically, workers are perfectly substitutable within an occupation. Imperfect substitutability between workers stems from imperfect substitutability between occupational outputs and the comparative advantage of differently talented workers for different occupations which impedes occupational mobility.

⁸See also [Rothschild and Scheuer \(2015\)](#) and [Lockwood et al. \(2014\)](#).

income shifting from the personal to the corporate tax base. [Piketty et al. \(2014\)](#) further extend the formula to incorporate rent seeking by managers. In their setting, managers exert costly effort in bargaining, which to the extent that it allows them to extract resources above their marginal product imposes a negative externality on others and is discouraged by higher taxation. Their formula for the optimal marginal tax rate above a threshold income is similar to the one that we derive in Section 3. However, the economics behind these formulas and their connections to the data are quite different. We abstract from rent-seeking, but instead stress the positive role of CEOs in creating firm value which is at center of an important literature on CEO compensation.

In a celebrated result [Diamond and Mirrlees \(1971\)](#) established that Ramsey tax formulas hold (and intermediate goods taxation is undesirable) in a general production setting provided the government has access to a rich enough set of linear commodity taxes and profits are taxed at 100 per cent. This echoes our result that the Mirrlees optimal tax formulas hold when profits are taxed at 100 per cent. Subsequent contributions by [Stiglitz and Dasgupta \(1971\)](#), [Munk \(1978\)](#) and [Munk \(1980\)](#) explored the consequences of restrictions on profit taxes for optimal commodity taxation. In these papers Ramsey optimal tax formulas are modified by a concern for the impact of commodity and factor taxes on the revenues collected via the taxation of profits. This parallels our result that marginal taxes on CEO incomes are decreasing in the social weight placed on profits.

By far the most closely related paper to ours is the contemporaneous work of [Scheuer and Werning \(2015\)](#). These authors develop the implications for optimal taxation of an assignment setting similar to ours. They focus on the case in which firm profits receive the same effective social weight as income tax revenues collected from CEOs. For this case they derive formulas that permit the testing of an income tax code for Pareto optimality (with respect to different combinations of weights across CEOs). These formulas are closely related to the ones we provide. In particular, they show that in this case, the “Mirrleesian” test condition for Pareto optimality expressed in terms of the effort elasticity and the talent distribution is the same as that obtained in the conventional setting without assignment. Moreover, they show that the formula when expressed in terms of the income distribution and CEO elasticities of income is also the same provided the latter elasticities are defined appropriately (as microeconomic elasticities that hold the pricing schedule for CEO effective labor fixed). This is an important observation that potentially leads to an alternative empirical implementation of this test formula. Collectively, [Scheuer and Werning \(2015\)](#) cast these results as neutrality propositions. However,

as [Scheuer and Werning \(2015\)](#) point out, while the test formulas themselves are neutral to assignment considerations, the way these formulas are brought to the data and their potential quantitative implications for tax design are not. We focus on the calculation of optimal taxes under different weightings of income taxes and firm profits. As noted above, we obtain optimal tax rates very much lower than those obtained in conventional analyses.

[Ales et al. \(2015\)](#) consider optimal taxation in a [Rosen \(1982\)](#) span of control setting. They identify top earners with managers who match with and control teams of workers. In contrast to our setting, firms have no exogenously given factors and all variations in firm size are attributable to variations in managerial talent. Managerial productivity is enhanced by firm (team) size and this creates a novel incentive for the government to tax firm size and, hence, shape the equilibrium managerial wage distribution. When considering managerial taxation, our paper departs from [Ales et al. \(2015\)](#) and follows [Terviö \(2008\)](#) in assuming that “firms are differentiated by important indivisible characteristics that cannot be easily shuffled among(st them)”.

Two recent and highly influential papers by [Gabaix and Landier \(2008\)](#) and [Terviö \(2008\)](#) use a competitive assignment framework to understand the determination of top CEO incomes. In this framework, CEO talent and a firm’s (indivisible and non-transferrable) assets are complementary and there is assortative matching of CEOs and firms.⁹ Both [Gabaix and Landier \(2008\)](#) and [Terviö \(2008\)](#) emphasize the role of variations in the size of a firm’s assets in the determination of top CEO incomes with the former attributing the rise in these incomes to increases in firm size. Our paper augments the sort of competitive assignment models considered by [Gabaix and Landier \(2008\)](#) and [Terviö \(2008\)](#) with an intensive effort margin and income taxation. Consistent with these contributions we find evidence of a thin right tail to the talent distribution. A separate literature looks at the CEO-firm relationship as a moral hazard problem. This literature is very large, see [Gayle and Miller \(2009\)](#) and [Gayle et al. \(2015\)](#) for recent empirical contributions and [Edmans and Gabaix \(2009\)](#) and [Frydman and Jenter \(2010\)](#) for surveys. It shares with [Gabaix and Landier \(2008\)](#) and [Terviö \(2008\)](#) the idea that CEOs create value for firm owners. In contrast to [Gabaix and Landier \(2008\)](#) and [Terviö \(2008\)](#) it focuses instead on the importance of motivating CEOs to exert effort through an appropriately structured compensation package. Our paper also emphasizes the importance of CEO effort in creating value, but unlike the moral hazard literature it abstracts from the structure of CEO pay and focuses upon the level of pay and

⁹[Terviö \(2008\)](#) in particular elaborates on the nature of these assets.

the role of taxes in shaping equilibrium outcomes.

2 Competitive Assignment of CEOs and Firms

We augment an assignment game of CEOs and firms with CEO effort and taxes on CEO incomes. While our focus is upon the taxation of CEO's, our analysis applies more broadly to the taxation of (high income) sellers and buyers in an assignment setting.

CEOs and firms The population of CEOs is described by a Lebesgue measure on the interval $I = (0, 1]$. Let $h : I \rightarrow \mathbb{R}_+$ give the talent of each CEO with h a smooth and decreasing function.¹⁰ Thus, $v \in I$ provides a ranking of CEO's by talent and the inverse of h is the counter-cumulative distribution of talent. Let F denote the distribution function of talent and f its density so that $F = 1 - h^{-1}$ and $f = -1/h_v$, where h_v is the derivative of h . The amount of effective labor z^s supplied by a CEO is a multiplicative combination of talent and effort:

$$z^s = h(v)e. \quad (1)$$

We assume that a CEO can sell labor to only one firm in a period.

The utility of a CEO over consumption c and effort e is given by $U : \mathbb{R}_+ \times [0, \bar{e}] \rightarrow \mathbb{R}$, with U strictly concave, twice continuously differentiable on the interior of its domain, strictly increasing in c and strictly decreasing in e . U is assumed to satisfy the Spence-Mirrlees single crossing property, that is $-\frac{U_e(c, z^s/h)}{hU_c(c, z^s/h)}$ is assumed to be decreasing in h at each (c, z^s, h) . A CEO must pay a tax $T : \mathbb{R}_+ \rightarrow \mathbb{R}$, $T(w) \in (-\infty, w]$, on her earnings. Consequently, if the v -th ranked CEO supplies effective labor z^s and earns income w , her after-tax income and, hence, consumption is $c(w) = w - T[w]$ and her utility is:

$$U \left(w - T[w], \frac{z^s}{h(v)} \right). \quad (2)$$

Finally, we assume that all CEOs have an outside utility option of $\bar{U} > U(0, 0)$.¹¹

¹⁰To allow for unbounded CEO talent, $I = (0, 1]$ is assumed open at 0. All of our results continue to hold if I is set equal to $[0, 1]$.

¹¹The assumption of a common outside utility option \bar{U} ensures that only the least talented CEO's outside option binds in equilibrium. It is stronger than needed. Provided the outside option of CEOs does not increase too strongly with talent, the behavioral response of CEOs to the tax perturbations we consider will continue to be on the CEO effort rather than the career margin and all of our tax analysis goes through.

A population of firms is also described by a Lebesgue measure on the interval I . Firms are differentiated by the size of their productive, non-transferable and indivisible assets. These assets could be intangibles such as reputation or goodwill that are difficult to trade, they could be firm-specific intellectual property or they could capture industry specific aspects of technology that shape the scale of the firm's operations.¹² Let $S : I \rightarrow \mathbb{R}_+$ give the size of, i.e. the quantity of assets at, each firm, with S a smooth and decreasing function. In the context of firms, $v \in I$ provides a ranking by (asset) size and the inverse of S is the counter-cumulative distribution of size. Let G denote the distribution function of firm size and g its density so that $G = 1 - S^{-1}$ and $g = -1/S_v$, where S_v is the derivative of S . If the v -th firm purchases z^b units of effective labor from a CEO and pays the CEO w , then firm claimants (i.e. the owners of $S(v)$) earn profits of:

$$V(S(v), z^b) - w, \quad (3)$$

where the surplus function $V : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ is assumed to be super-modular, increasing in both arguments and continuously differentiable and concave in z^b . Note that the surplus value $V(S(v), z^b)$ is net of payments to other adjustable inputs. It can be obtained from a richer problem in which the firm buys (a vector of) additional adjustable inputs x at prices p :

$$V(S(v), z^b) := \sup_x W(S(v), z^b, x) - p \cdot x. \quad (4)$$

The input vector x could include adjustable capital and we explicitly extend the model in such a direction in our later quantitative section.^{13,14}

The market assignment game with taxes Given T and \bar{U} , CEOs and firms play an assignment game. As a precursor to later optimal tax results, we formalize this game and characterize its equilibrium. The analysis is complicated relative to that in [Terviö \(2008\)](#) and [Gabaix and Landier \(2008\)](#) by the inclusion of the intensive effort margin on the side of CEOs and taxes.¹⁵

¹²See [Terviö \(2008\)](#) and the references therein for additional discussion.

¹³Despite the exclusion of adjustable capital, we refer to $S(v)$ as firm v 's *assets* adding the qualifiers "non-transferable" or "immovable" when needed.

¹⁴The formulation (4) treats the factor prices p as exogenous to CEO behavior. If the adjustable inputs x complemented CEO effective labor and were not in perfectly elastic supply, then increases in CEO effort could raise demand for these inputs and, hence, (equilibrium) prices p . In this way tax policy that deterred CEO effort could have adverse effects beyond those emphasized in this paper.

¹⁵A tractable and important special case assumes quasilinear-constant elasticity CEO preferences $U(c, e) = c - \frac{\varepsilon}{1+\varepsilon} e^{\frac{1+\varepsilon}{\varepsilon}}$ and multiplicative firm payoffs $V(S, z) = DSz$. In this special case, absent

Let $w : I \rightarrow \mathbb{R}_+$ give the CEO income paid by each firm, $\mu : I \rightarrow I \cup \{u\}$ the talent rank of each firm's CEO (with $\mu(v) = u$ indicating that v is unmatched) and $z : I \rightarrow \mathbb{R}_+$ the quantity of CEO effective labor purchased by each firm. The functions w , μ and z are assumed to be Lebesgue measurable. In addition, the match function μ is assumed to be measure-preserving, i.e. for all Lebesgue measurable sets $B \subset \mu^{-1}(I)$, $\mathcal{M}[\mu(B)] = \mathcal{M}[B]$, where \mathcal{M} denotes Lebesgue measure. This captures the one-to-one matching of CEOs and firms.

Equilibria in the assignment game Definition 1 below defines an equilibrium for the assignment game with taxes. The definition requires that no CEO or firm can improve on its equilibrium allocation by unilaterally leaving the market and that there is no CEO-firm pair whose members can make themselves jointly better off by, if necessary, dissolving their equilibrium matches or leaving their current unmatched states, matching together and choosing a new income-labor combination.

Definition 1. A triple (μ, z, w) is an equilibrium of the firm-CEO assignment game at (T, \bar{U}) , if:

1. **(Participation)** each matched firm-CEO pair $(v, \mu(v))$ is better off at their equilibrium allocation than unmatched, i.e. for all $v \in \mu^{-1}(I)$,

$$V(S(v), z(v)) - w(v) \geq 0 \quad \text{and} \quad U\left(w(v) - T[w(v)], \frac{z(v)}{h(\mu(v))}\right) \geq \bar{U}; \quad (5)$$

2. **(Stability)** there is no firm v , CEO v' and allocation (z', w') such that both firm and CEO weakly prefer (z', w') to their equilibrium allocation and at least one strictly prefers it, i.e. there is no v, v', w' and z' such that:

- (a) if firm v is unmatched in equilibrium, then it obtains a weakly higher payoff from (z', w') than from being unmatched:

$$V(S(v), z') - w' \geq 0 \quad \text{if } \mu(v) = u \quad (6)$$

or if firm v is matched in equilibrium (with CEO $\mu(v)$), then it obtains a weakly higher payoff from (z', w') than from allocation $(z(v), w(v))$:

$$V(S(v), z') - w' \geq V(S(v), z(v)) - w(v) \quad \text{if } \mu(v) \in I \quad (7)$$

taxes, the assignment equilibrium is equivalent to one in which there is no intensive effort margin. The introduction of taxes breaks this equivalence. We thank a referee for emphasizing this point.

and

(b) if CEO v' is unmatched in equilibrium, then it obtains a weakly higher payoff from (z', w') than from being unmatched:

$$U\left(w' - T[w'], \frac{z'}{h(v')}\right) \geq \bar{U} \quad \text{if } v' \notin \mu(I) \quad (8)$$

or if CEO v' is matched in equilibrium (with firm $\hat{v} = \mu^{-1}(v')$), then it obtains a weakly higher payoff from (z', w') than from allocation $(z(\hat{v}), w(\hat{v}))$:

$$U\left(w' - T[w'], \frac{z'}{h(v')}\right) \geq U\left(w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(v')}\right) \quad \text{if } \mu(\hat{v}) = v', \quad (9)$$

with at least one of the applicable inequalities above strict;

3. **(No rents to the least talented CEO)** if $1 \in \mu(I)$, then:

$$U\left(w(\mu^{-1}(1)) - T[w(\mu^{-1}(1))], \frac{z(\mu^{-1}(1))}{h(1)}\right) = \bar{U}. \quad (10)$$

Note that an implication of the assumption $\bar{U} > U(0,0)$ is that any CEO-firm match involves the trade of a positive amount of effective labor for a positive income: there are no passive matches in which nothing is done.¹⁶ The “no rents to the least talented CEO” component of the definition may be justified informally by assuming that the interval I of CEOs are the most talented members of a population of strictly greater measure with outside option \bar{U} and that if the least talented CEO obtained a payoff in excess of \bar{U} , then a slightly less talented unmatched person could supply the same effective labor for an income slightly below $w(\mu^{-1}(1))$ to firm $\mu^{-1}(1)$ and make both this firm and herself better off.

We now give a proposition that characterizes equilibria. It shows that there is assortative matching between CEOs and firms who choose to match and gives simple participation and incentive constraints that must be satisfied in equilibrium. The latter conditions require only that each matched CEO (resp. firm) is better off accepting its equilibrium allocation than the equilibrium allocation of another matched CEO (resp. firm). In addition, the proposition establishes that these con-

¹⁶CEOs must receive a strictly positive income to match with a firm and give up their outside option. Since firms must earn non-negative revenue, they must contract for a positive amount of effective labor from a CEO.

ditions are sufficient for stability if the tax function sufficiently penalizes income-labor allocations outside of the range of the equilibrium allocation functions (z, w) . In particular, in this case they ensure that joint deviations in which a CEO-firm pair dissolve their equilibrium matches, rematch with each other and select a new income-labor allocation cannot make both parties better off. Thus, the (more complicated) stability conditions on firms and CEOs in Definition 1 are decoupled and re-expressed as simple CEO and firm incentive conditions. This is useful for our subsequent tax analysis.

Proposition 1. *If (μ, z, w) is an equilibrium at (T, \bar{U}) , then either (a) $\mu = u$, no firm produces and all CEOs take their outside option or (b) there is a $\bar{v} \in I$ such that (i) for all $v \in (\bar{v}, 1]$, $\mu(v) = u$ and (ii) for all $v \in (0, \bar{v}]$, $\mu(v) = v$. Moreover, z and w satisfy the **participation conditions**, for all $v \in (0, \bar{v}]$,*

$$U\left(w(v) - T[w(v)], \frac{z(v)}{h(v)}\right) \geq \bar{U} \quad \text{and} \quad V(S(v), z(v)) - w(v) \geq 0, \quad (11)$$

and the **incentive conditions**, for all $v, v' \in (0, \bar{v}]$,

$$U\left(w(v) - T[w(v)], \frac{z(v)}{h(v)}\right) \geq U\left(w(v') - T[w(v')], \frac{z(v')}{h(v)}\right), \quad (12)$$

and

$$V(S(v), z(v)) - w(v) \geq V(S(v), z(v')) - w(v'). \quad (13)$$

On the other hand, given \bar{U} , if T , \bar{v} , z and w are such that (i) for all $v \in (0, \bar{v}]$, (11) to (13) hold, (ii) $U\left(w(\bar{v}) - T[w(\bar{v})], \frac{z(\bar{v})}{h(\bar{v})}\right) = \bar{U}$ and $V(S(\bar{v}), z(\bar{v})) - w(\bar{v}) \geq 0$ and (iii) for all $w' \notin w((0, \bar{v}])$, $T[w'] = w'$, then (μ, z, w) with μ such that for all $v \in (\bar{v}, 1]$, $\mu(v) = u$, and for all $v \in (0, \bar{v}]$, $\mu(v) = v$, is an equilibrium at (T, \bar{U}) .

Proof. See Appendix A. □

We also note that, by standard arguments, (see Lemma A.2 in Appendix A) equilibrium effective labor z and CEO income w are non-increasing on $(0, \bar{v}]$ as are CEO consumption $c(v) = w(v) - T[w(v)]$ and CEO and firm payoffs:

$$\Phi(v) := U\left(w - T[w], \frac{z(v)}{h(v)}\right) \quad \text{and} \quad \pi(v) := V(S(v), z(v)) - w(v).$$

Equilibrium CEO income determination If z and w are differentiable at $v \in (0, \tilde{v}]$, then (13) implies the first order condition for firms:

$$V_z(S(v), z(v))z_v(v) - w_v(v) = 0. \quad (14)$$

Integrating out (14) gives:

$$w(v) = w(\tilde{v}) + \int_v^{\tilde{v}} V_z(S(v'), z(v'))(-z_v(v'))dv'. \quad (15)$$

It follows that each increment of effective labor $-z_v(v')$ above $z(\tilde{v})$ is “priced” at $V_z(S(v'), z(v'))$, i.e. is paid its marginal product at the firm at which it is the last unit hired. In particular, CEO v receives her marginal product $V_z(S(v), z(v'))$ only on the last unit she supplies. On infra-marginal units she is paid $V_z(S(v'), z(v'))$ with firm v collecting the difference $\{V_z(S(v), z(v')) - V_z(S(v'), z(v'))\}(-z_v(v'))$ as profit. Super-modularity of V implies that the “price of effective labor” V_z is increasing in S . Since more talented firms match with larger firms, this is a source of dispersion of CEO income across talent ranks. Even if there is relatively little dispersion in CEO talent, this force can translate a large variation in firm size into a large variation in top CEO incomes. If, in addition, T is differentiable at $w(v)$ and w and z are differentiable at $v \in (0, \tilde{v}]$, then (12) implies the CEO’s first order condition:

$$U_c \left(w(v) - T[w(v)], \frac{z(v)}{h(v)} \right) w_v(v) \{1 - T_w[w(v)]\} + U_e \left(w(v) - T[w(v)], \frac{z(v)}{h(v)} \right) \frac{z_v(v)}{h(v)} = 0. \quad (16)$$

Combining (14) and (16), totally differentiating with respect to v and denoting the compensated and uncompensated effort elasticities by \mathcal{E}^c and \mathcal{E}^u gives:¹⁷

$$\frac{w_v}{w} = \left(\frac{V_z z}{w} \right) \frac{(1 + \mathcal{E}^u) \left(\frac{h_v}{h} \right) + \mathcal{E}^c \frac{V_z S}{V_z} \left(\frac{S_v}{S} \right)}{1 + \mathcal{E}^c \left\{ \frac{T_{ww} w}{1 - T_w} \frac{V_z z}{w} + \frac{V_{zz} z}{V_z} \right\}}. \quad (17)$$

Expression (17) relates equilibrium CEO income variation $\frac{w_v}{w}$ to talent $\frac{h_v}{h}$ and firm size variation $\frac{S_v}{S}$ (across rank). These last two variables contribute to CEO income variation directly and also indirectly through the incentives for greater effort that they create. In much of the paper we specialize to the case in which firm’s objective

¹⁷Here the compensated and uncompensated elasticities are given by $\mathcal{E}^c = \frac{U_e/e}{-2\frac{U_e}{U_c}U_{ce} + (\frac{U_e}{U_c})^2 U_{cc} + U_{ee}}$

and $\mathcal{E}^u = \frac{U_e/e - (\frac{U_e}{U_c})^2 U_{cc} + (\frac{U_e}{U_c}) U_{ce}}{-2\frac{U_e}{U_c}U_{ce} + (\frac{U_e}{U_c})^2 U_{cc} + U_{ee}}$ with each U_x and U_{xy} , $x, y \in \{c, e\}$, giving the relevant partial first and second derivatives of U .

is multiplicative $V(S, z) = DSz$, with D a parameter. If, in addition, the tax function T is locally linear, then equation (17) reduces to:

$$\frac{w_v}{w} = \left(\frac{DSz}{w} \right) \left\{ \frac{h_v}{h} + \mathcal{E}^u \frac{h_v}{h} + \mathcal{E}^c \frac{S_v}{S} \right\}. \quad (18)$$

Thus, CEO income variation is related via the price of effective labor DS to effective labor variation $\frac{z_v}{z} = \frac{h_v}{h} + \mathcal{E}^u \frac{h_v}{h} + \mathcal{E}^c \frac{S_v}{S}$. Equation (18) further decomposes CEO income variation into a part due to variation in the talent rent accruing to CEOs $\left(\frac{DSz}{w} \right) \frac{h_v}{h}$ and a part due to variation in CEO effort $\left(\frac{DSz}{w} \right) \frac{e_v}{e} = \left(\frac{DSz}{w} \right) \left(\mathcal{E}^u \frac{h_v}{h} + \mathcal{E}^c \frac{S_v}{S} \right)$.¹⁸ In the assignment models of [Terviö \(2008\)](#) and [Gabaix and Landier \(2008\)](#) there is no intensive effort margin, all CEO income variation is attributable to variation in talent rents and (18) further reduces to $\frac{w_v}{w} = \frac{DSz}{w} \frac{h_v}{h}$. Alternatively, in the standard labor supply model used in optimal taxation there is an intensive effort margin, but workers capture all of the surplus (with no part accruing to owners of a firm asset S). In this case (18) reduces to $\frac{w_v}{w} = \frac{h_v}{h} + \mathcal{E}^u \frac{h_v}{h}$ and CEO income variation is not enhanced by variation in firm size.

Optimal income tax formulas are often expressed as functions of the (local) Pareto coefficients for talent or income. Anticipating these formulas it is useful to re-express (18) in these terms. Recall that F and f denote the talent distribution and density and G and g the firm asset distribution and density. Let M and m denote the distribution and density of CEO incomes. The corresponding local Pareto coefficients for CEO talent, firm size and CEO income are: $\alpha_h(v) := \frac{h(v)f(h(v))}{1-F(h(v))}$, $\alpha_S(v) := \frac{S(v)g(S(v))}{1-G(S(v))}$ and $\alpha_w(v) := \frac{w(v)m(w(v))}{1-M(w(v))}$. Using $f(h(v)) = -\frac{1}{h_v}(v)$, $g(S(v)) = -\frac{1}{S_v}(v)$, $m(w(v)) = -\frac{1}{w_v}(v)$ and multiplying by the counter-cumulative distributions (18) becomes:

$$\frac{1}{\alpha_w} = \frac{DSz}{w} \left\{ (1 + \mathcal{E}^u) \frac{1}{\alpha_h} + \mathcal{E}^c \frac{1}{\alpha_S} \right\}. \quad (19)$$

Similarly, letting α_π denote the local Pareto coefficient for firm profits, the envelope condition for firms $\pi_v(v) = DS_v(v)z(v)$ can be re-written as:

$$\frac{1}{\alpha_\pi} = \frac{DSz}{\pi} \frac{1}{\alpha_S}. \quad (20)$$

Together (19) and (20) permit recovery of the local Pareto coefficients for (unobserv-

¹⁸Note that the uncompensated elasticity attaches to talent variation and the compensated elasticity to firm asset variation. Local variation in talent modifies the return to effort and CEO consumption; local variation in firm asset size modifies only the return to effort. Firm claimants not the CEO capture additional firm surplus attributable to local variation in S .

able) CEO talent and firm asset size from the corresponding coefficients for CEO incomes and firm profits. In particular, (19) and (20) imply:

$$(1 + \mathcal{E}^u) \frac{1}{\alpha_h} = \frac{w}{DSz} \frac{1}{\alpha_w} - \mathcal{E}^c \frac{\pi}{DSz} \frac{1}{\alpha_\pi}. \quad (21)$$

Larger values for the reciprocal of the Pareto coefficient in the right tail of a distribution indicate a thicker or fatter tail.¹⁹ Thus, (21) relates the tail thickness of the CEO talent distribution to those of the CEO income and firm profit distributions. Notice, in particular that an observed fat CEO income tail need not imply that the underlying CEO talent tail is fat. Mechanically from (21), $\frac{1}{\alpha_w}$ may be large (a fat CEO income tail) and $\frac{1}{\alpha_h}$ small (a thin CEO talent tail) if $\frac{w}{DSz}$ is small and/or $\frac{1}{\alpha_\pi}$ large. As described previously, CEOs may be dispersed across a large interval of high incomes not because of large variations in CEO talent, but because competition amongst firms for CEO talent translates large variations in firm size into large CEO income variation.

2.1 The Effect of Taxes on CEO Income and Profits

We now describe how the marginal tax rate impacts equilibrium CEO incomes and firm profits. To develop intuition we start with a simple setting in which taxes are linear above a threshold income w_0 :

$$T[w] := T[w_0] + \tau(w - w_0) \quad w \in [w_0, \infty). \quad (22)$$

We consider the consequences of variation in τ (keeping w_0 , $T[w_0]$ and \bar{U} fixed throughout). We refer to $1 - \tau$ as the *retention rate* (above w_0). To decompose the effects of changes in the marginal tax on CEO incomes and to consider different equilibria parameterized by the retention rate, it will be convenient to re-express CEO income as a function of effective labor and to sometimes make its dependence on taxes explicit. Thus, we define $\omega(z; 1 - \tau)$ to be the income of a CEO supplying effective labor z when the retention rate is $1 - \tau$.²⁰ Similarly, we sometimes make the dependence of equilibrium effective labor on the tax rate explicit and define $z(v; 1 - \tau)$ to be the equilibrium effective labor of CEO v given retention rate $1 - \tau$.

For simplicity, assume that firm surplus is multiplicative in S and z , i.e. $V(S, z) =$

¹⁹A distribution has a heavy (right) tail if its density has no more than limiting exponential decay and is fat tailed if its density has limiting geometric decay. Fat tailed distributions have finite limiting Pareto coefficients. Thinner tailed distributions have infinite limiting Pareto coefficients.

²⁰Function ω is well defined, see Appendix A.

DSz . Then, as described previously, in equilibrium firm v pays its marginal product $DS(v)$ for the last (and only for the last) unit of effective labor it hires. Since larger firms have larger marginal products and hire more effective CEO labor, it follows that the “prices” paid for successively higher increments of effective labor are higher. Moreover, the total income earned by a CEO supplying effective labor z depends upon the prices paid for each incremental unit of effective labor up to z and this in turn depends upon the identity of the firms for whom these incremental units were the last hired. Specifically, in equilibrium CEO v earns:

$$\omega(z(v; 1 - \tau); 1 - \tau) = \omega(z_0; 1 - \tau) + \int_{z_0}^{z(v; 1 - \tau)} DS(v(z'; 1 - \tau)) dz', \quad (23)$$

where z_0 is the effective labor of the least talented CEO earning income weakly more than w_0 and $v(\cdot, 1 - \tau)$ is the inverse of $z(\cdot; 1 - \tau)$ with $v(z'; 1 - \tau)$ giving the equilibrium rank of the CEO exerting effective labor z' . The overall impact of a rise in $1 - \tau$ on the income of CEO v is thus:²¹

$$\begin{aligned} \frac{d\omega}{d(1 - \tau)}(v) &= \frac{\partial \omega}{\partial z}(z(v)) \frac{\partial z}{\partial(1 - \tau)}(v) + \frac{\partial \omega}{\partial(1 - \tau)}(z(v)) \\ &= DS(v) \frac{\partial z}{\partial(1 - \tau)}(v) + \int_{z_0}^{z(v)} \left\{ DS_v(v(z')) \frac{\partial v}{\partial(1 - \tau)}(z') \right\} dz', \quad (24) \end{aligned}$$

where to simplify notation dependence on $1 - \tau$ is omitted from the functions. The first term on the right hand side of both equalities gives the impact of the retention rate change on CEO v 's effective labor and, hence, income holding the equilibrium pricing schedule for effective labor fixed at $\omega(\cdot; 1 - \tau)$. Provided income effects on CEO effort are not too strong, this term is positive.²² The second term gives the impact of the tax change on the income paid to the CEO supplying effective labor $z(v; 1 - \tau)$. This term is negative (again provided income effects on CEO effort are not too strong). To see why consider the impact of a rise in $1 - \tau$ that induces CEOs to work harder. Each incremental unit of effective labor between z_0 and a given $z(v; 1 - \tau)$ then becomes associated with a less talented CEO matched to a smaller firm with a lower marginal product of effective labor. As a result, the prices paid for these incremental units falls, as does the income paid to the CEO supplying $z(v; 1 - \tau)$.

²¹ Equation (24) uses $\frac{\partial \omega}{\partial(1 - \tau)}(z_0, 1 - \tau) = 0$, see Appendix B for the derivation.

²²More precisely, if CEOs' uncompensated behavioral elasticity of effort is positive, then a rise in $1 - \tau$ induces CEOs to work harder, see Appendix B.

Next consider the impact of a retention rate rise on the profit of firm v , $\pi(v; 1 - \tau) = DS(v)z(v; 1 - \tau) - \omega(z(v; 1 - \tau); 1 - \tau)$. Application of an envelope theorem to the firm's problem implies that:

$$\frac{d\pi}{d(1 - \tau)}(v) = -\frac{\partial\omega}{\partial(1 - \tau)}(z(v)) = -\int_{z_0}^{z(v)} \left\{ DS_v(v(z')) \frac{\partial v}{\partial(1 - \tau)}(z') \right\} dz' > 0. \quad (25)$$

A retention rate rise induces a downwards adjustment in the schedule $\omega(\cdot; 1 - \tau)$ causing firm profits to rise. As our discussion below highlights, this spillover from the CEO income tax rate to firm profits is a force for lower optimal tax rates relative to conventional formulas.

Let $\mathcal{E}_w(v)$ and $\mathcal{E}_\pi(v)$ denote the elasticities of CEO income and firm profit with respect to the retention rate (for the v -th ranked CEO and firm at retention rate $1 - \tau$). In Appendix B, we use (24) and (25) to derive explicit expressions for these elasticities. In general, these expressions are complicated. However, they are much simplified in the case of quasilinear/constant elasticity CEO preferences. Specifically, if CEO preferences are $U(c, w) = c - \frac{\mathcal{E}}{1 + \mathcal{E}} e^{\frac{1 + \mathcal{E}}{\mathcal{E}}}$, with $\mathcal{E} > 0$ the elasticity of CEO effort, then $\mathcal{E}_w(v)$ and $\mathcal{E}_\pi(v)$ are given by:

$$\mathcal{E}_w(v) = \frac{DS(v)z(v)}{w(v)} \mathcal{E} - \left(\frac{\pi(v) - \pi_0}{w(v)} \right) \mathcal{E} \quad (26a)$$

$$\mathcal{E}_\pi(v) = \left(\frac{\pi(v) - \pi_0}{\pi(v)} \right) \mathcal{E}, \quad (26b)$$

where π_0 is the profit of the smallest firm paying its CEO at least w_0 , $w(v) = \omega(z(v))$ and $\frac{\partial\omega}{\partial(1 - \tau)} = -\left(\frac{\pi(v) - \pi_0}{\pi(v)} \right) \mathcal{E}$. Note that (26a) can be rewritten as: $\mathcal{E}_w(v) = \left(1 + \frac{\pi_0}{w(v)} \right) \mathcal{E} > \mathcal{E}$ and that $\mathcal{E}_\pi(v) = \left(\frac{\pi(v) - \pi_0}{\pi(v)} \right) \mathcal{E} \in (0, \mathcal{E})$.²³

3 Optimal Linear Taxes

As a precursor to analysis of optimal *nonlinear* taxes, we consider the problem of a policymaker selecting an optimal linear tax function over a range of top incomes.

²³In the language of [Scheuer and Werning \(2015\)](#), $\mathcal{E}_w(v)$ and $\mathcal{E}_\pi(v)$ are “macro-elasticities” that describe the equilibrium response of a given CEO's income to changes in the retention rate. Aggregates of these elasticities feature in the optimal tax equations of the next section. In particular, the elasticity of taxable income that is emphasized in the public finance literature and that features in the classical Diamond-Saez optimal tax equation is an aggregate of the (macro)elasticities $\mathcal{E}_w(v)$. [Scheuer and Werning \(2015\)](#) make the distinction between such macro-elasticities and micro-elasticities that isolate the response of an individual's income to a tax change holding everyone else's behavior fixed. They show that test conditions for the Pareto optimality of a nonlinear tax function are of a standard form when cast in terms of such micro-elasticities.

The simple tax formulas in this case directly connect our results to related formulas in [Saez \(2001\)](#), [Diamond and Saez \(2011\)](#) and [Piketty et al. \(2014\)](#). They also highlight the role of spillovers from CEO income taxation to firm profits and, hence, that of the social marginal value of profit in shaping and modifying conventional optimal tax formulas.

The policy maker’s problem Assume that the policymaker is restricted to CEO income tax functions in the class (22) and that she selects a marginal tax rate τ to maximize a weighted sum of income tax revenues and firm profits:

$$\sup_{\tau \in [0,1]} \tau \int_0^{v_0} \{w(v; 1 - \tau) - w_0\} dv + \chi \int_0^{v_0} \pi(v; 1 - \tau) dv, \quad (27)$$

keeping \bar{U} and w_0 fixed.²⁴

Interpreting χ The weight χ is the social marginal value of (aggregate) profit and an important parameter in our analysis. If the planner cares only about tax revenues and the social value of profit stems entirely from its role as a source of such revenues, then χ corresponds to the profit tax rate τ^F . Under this interpretation, the policymaker in (27) is simply selecting the CEO income tax rate τ to maximize overall tax revenues given the profit tax. In our assignment model, profits are pure rent and placing taxes upon them is non-distortionary. Consequently, the fully optimal tax system in which both CEO and profit taxes are chosen is one in which profits are taxed at 100% and $\chi = 1$. Analysis of optimal CEO income taxation with χ fixed at a value of less than one corresponds in our environment to a partial reform of the tax system (with the profit tax rate fixed at a sub-optimal level). Alternatively, it reflects a fully optimal outcome in an extended model featuring economic frictions or institutional constraints that restrict both profit taxes and the social marginal value of profit.²⁵

²⁴Here, v_0 denotes the rank of the least talented CEO earning at least w_0 . Note that the choice of τ does not affect the incomes of CEOs earning less than w_0 , which evolve from $w(\bar{v})$ according to (18) with $w(\bar{v})$ and $z(\bar{v})$ determined to ensure that firm \bar{v} maximizes profits subject to the \bar{v} CEO receiving utility \bar{U} . Hence, it does not affect v_0 .

²⁵A literature exploring the optimal structure of commodity and other taxation under exogenous restrictions on profit taxes developed in the 1970s and 1980s. Prominent contributions include [Stiglitz and Dasgupta \(1971\)](#), [Munk \(1978\)](#) and [Munk \(1980\)](#). In these papers Ramsey optimal tax formulas are modified by a concern for the impact of commodity and factor taxes on the revenues collected from the taxation of profits. In particular, if the profit tax is constrained to be less than 100 per cent, this concern is less important and a higher commodity tax may be warranted. This parallels our result that if $\chi < 1$, then marginal taxes on CEO incomes are optimally higher.

We briefly elaborate the implications of such frictions for the interpretation of χ and the determination of profit taxes. Assume first that firms can divert profits to owners before taxation, but that for every pre-tax dollar diverted the owner receives only $1 - t$ dollars. For example, suppose that a firm can realize profits in a foreign tax jurisdiction through transfer pricing. If foreign corporate and dividend income taxes equal t and owners do not repatriate profit income (because, say, they partly reside and consume in the foreign country), then the firm can divert profits to owners at the rate $1 - t$.²⁶ Alternatively, suppose that a firm can conceal profit transfers to owners as a deductible business expense at a cost of t dollars lost per dollar concealed. In these situations, each firm v , after determination of $y(v)$ and $w(v)$, will select the pre-tax transfer p to owners to solve:

$$\max_{p \in [0, V(S(v), y(v)) - w(v)]} (1 - t)p + (1 - \chi) \{V(S(v), y(v)) - w(v) - p\}.$$

If the profit tax exceeds t , then the firm will divert all profit to owners prior to the application of the profit tax and, consequently, this tax and χ are effectively bounded by t . If the amount received by firm owners as a function of the amount diverted to them is concave rather than linear,²⁷ then the social marginal value of profit χ at the optimum no longer equals the profit tax τ^F . Rather $\chi = \tau^F(1 - \theta(\tau^F)) \leq \tau^F \leq 1$, where $\theta(\tau^F)$ is the fraction of profit optimally transferred to owners prior to taxation at the profit tax.

It is natural to conjecture that the disincentive to create firms might also deter high rates of profit taxation and, hence, lower the social marginal value of profit χ to a planner concerned only with maximizing tax revenues. If, however, profits relax firm entry conditions, then profit has social value beyond its direct role as a base for taxation. In some situations firm entry is consistent with the optimal social marginal value of profit equalling that of tax revenues (even if the tax on firm profit is optimally set below one). We develop the implications of the firm entry margin for the effective social marginal value of profit in Appendix D.1.²⁸

²⁶Zucman (2014) describes and attempts to measure the extent of such tax avoidance and evasion under current tax regimes. He estimates that about 20% of US corporate profits are realized overseas and about 80% of offshore accounts held by US citizens are not declared to US tax authorities.

²⁷This concavity could reflect progressive foreign taxation, high costs of overseas consumption for some owners or convex costs of concealing transfers.

²⁸In our baseline assignment model, a fixed population of firms learns their asset size $S(v)$ and chooses whether or not to hire a CEO and operate. Only the profit of the smallest entering firm is relevant for the firm entry condition. Profits at larger firms are not and the social marginal value of profit equals the profit tax. Further, all entry costs are absorbed into V and are, thus, implicitly treated as being tax deductible.

The weight χ may also incorporate direct social concern for firm claimants. For example, if χ^F is the welfare weight placed on firm claimant incomes (inclusive of any lump sum transfers) and a unit welfare weight is placed on tax revenues, then, absent other economic frictions, the effective weight placed on profits is: $\chi = \tau^F + \chi^F(1 - \tau^F)$. If τ^F is (constrained to be) less than one, then the effective social weight on profits is enhanced by direct concern for firm claimants and is entirely due to such concern if $\tau^F = 0$.

Further assumptions The objective in (27) places no direct weight on CEOs. In our quantitative analysis in Section 5, we focus upon CEOs earning incomes above \$500,000. The zero welfare weight placed on CEOs parallels the treatment of high income earners in the optimal tax analyses of [Diamond and Saez \(2011\)](#) and [Piketty et al. \(2014\)](#). In Appendix B we extend our analysis to allow for positive weighting of CEOs. We also abstract from any effects of τ on the after-tax labor income of (socially valued) non-CEO workers. Implicitly, this supposes that such workers do not pay the tax either because it is specific to CEOs or because they earn incomes below the threshold w_0 and that the incomes earned by non-CEO workers are not affected by the effort of CEOs. Thus, non-CEOs are only affected by τ to the extent that they benefit from tax revenues or have a claim on firm profits.²⁹

Aggregate elasticities and tail coefficients To state the optimal tax formulas implied by (27), it is useful to introduce a number of definitions. Let W denote the total income of CEOs earning more than the threshold income w_0 when the retention rate is $1 - \tau$ and Π the corresponding total firm profit. For notational ease we suppress the dependence of W , Π and other variables on $1 - \tau$ for the remainder of this section. Define the “aggregate” elasticities at the optimum by:

$$\mathcal{E}_W = \frac{1 - \tau}{W} \frac{\partial W}{\partial(1 - \tau)} \quad \text{and} \quad \mathcal{E}_\Pi = \frac{1 - \tau}{\Pi} \frac{\partial \Pi}{\partial(1 - \tau)}. \quad (28)$$

To the extent that high income earners are identified with CEOs, \mathcal{E}_W is the model counterpart of the elasticity of taxable income (ETI) for high earners emphasized in the empirical public finance literature.³⁰ These elasticities are aggregates of the

²⁹If (4) holds, non-CEO labor is an adjustable input that complements CEO effective labor and the elasticity of such labor supply is not perfect, then higher tax rates on CEOs that deter CEO effort depress demand for non-CEO labor and, hence, the equilibrium non-CEO wage making non-CEOs worse off. Including such an effect would create an additional force for lower marginal taxation of CEOs.

³⁰See [Saez et al. \(2012\)](#) for an extensive discussion of the role of ETI in public finance.

individual level elasticities given in the preceding section: $\mathcal{E}_W = \int_0^{v_0} \frac{w}{W} \mathcal{E}_w(v) dv$ and $\mathcal{E}_\Pi = \int_0^{v_0} \frac{\pi}{\Pi} \mathcal{E}_\pi(v) dv$. Both incorporate adjustment in the equilibrium CEO income and effective labor schedules. By our previous discussion, \mathcal{E}_Π is positive under reasonable economic restrictions. In particular, in the quasilinear/constant elasticity setting using (26) and (28) we have:

$$\mathcal{E}_W = \left\{ 1 + \frac{\pi_0}{W} \right\} \mathcal{E} \quad \text{and} \quad \mathcal{E}_\Pi = \frac{1}{A_\Pi} \mathcal{E}, \quad (29)$$

where $A_\Pi = \frac{\Pi}{\Delta\Pi}$, with $\Delta\Pi = \Pi - \pi_0$, and as before \mathcal{E} is the CEO's effort elasticity. Let $A_W := \frac{W}{\Delta W}$ where $\Delta W := W - w_0$. We refer to A_W and A_Π as the tail coefficients of CEO income and profit. It is easy to verify that:

$$\frac{1}{A_W} = \int_0^{v_0} \frac{w(v)}{W} \frac{1}{\alpha_w(v)} dv.$$

In particular, if CEO income has a Paretian right tail above w_0 , then $A_W = \alpha_w$, where α_w is the constant Pareto coefficient.

Optimal tax formulas With the preceding definitions in place, a simple formula for the optimal marginal income tax rate τ^* is available. Rearranging the first order condition from (27) and using the definitions given above yields:

$$\tau^* = \frac{1 - \chi A_W^* \frac{\Pi^*}{W^*} \mathcal{E}_\Pi^*}{1 + A_W^* \mathcal{E}_W^*}, \quad (30)$$

where *'s denote optimal values. Formula (30) contrasts with the standard expression derived by [Saez \(2001\)](#) and emphasized by [Diamond and Saez \(2011\)](#):

$$\tau^{\text{Saez}} = \frac{1}{1 + A_W^* \mathcal{E}_W^*}. \quad (31)$$

The logic behind (30) extends that behind (31) to include concern for the spillover from CEO income taxes to profits.³¹

Given $A_W^* \frac{\Pi^*}{W^*} \mathcal{E}_\Pi^* > 0$, formulas (30) and (31) give the same value for the optimal tax rate only if profits receive no weight in the policymaker objective ($\chi = 0$). More generally, if $\chi > 0$, then the depressing effect of τ on profits creates a motive for

³¹Note formula (30) does not rely on the assignment framework. It is valid whenever there is a spillover from the CEO income tax rate to firm profits and, in this sense, is quite general. The assignment model supplies the mechanism underlying this spillover and facilitates our subsequent empirical strategy for evaluating optimal nonlinear taxes.

lower marginal taxes.

Optimal taxes when $\chi = 0$. Diamond and Saez (2011) and Saez et al. (2012) use formula (31) to provide guidance on the optimal taxation of top earners. As noted, this optimal tax formula emerges as the appropriate one in our assignment model as well (only) if $\chi = 0$. Following the discussion earlier in this section, this is an extreme case requiring no ability to tax profits, no impact of profits on firm entry conditions and no concern for the recipients of firm profits. However, as a benchmark for subsequent calculations, we evaluate (31).

Various authors proceed as if $A_W \mathcal{E}_W$ is relatively stable in the face of marginal tax rate changes and use empirical evaluations of $A_W \mathcal{E}_W$ in US data (and at a prevailing allocation) to determine or at least approximate its value at the optimum. Our model is consistent with such a strategy if CEO preferences are of the quasi-linear/constant elasticity form, firm surplus is multiplicative and the tax rate is linear above a threshold (see Appendix B). Based on prior empirical analyses of the general population (of top earners), Diamond and Saez (2011) and Saez et al. (2012) set A_W equal to 1.5 and \mathcal{E}_W to 0.25 implying a top tax rate of 72.7 per cent.³² It is possible that the population of top earning CEOs is different from the general population of top earners. In Appendix G, we estimate A_W for the CEO population using the Standard and Poor's ExecuComp database and find that it is stable above an income of about \$12 million (2011 USD) and equal to 2.1. There is limited direct evidence on \mathcal{E}_W for top CEOs. Frydman and Molloy (2011) report a strong negative correlation between top marginal tax rates and aggregate CEO incomes in the US. However, they also estimate a small contemporaneous response of CEO incomes to tax reforms in the cross section. Based on this, they reject a value of \mathcal{E}_W for CEOs above 0.2; their largest point estimate is about 0.094. Goolsbee (2000) studies data from 1991 to 1995 and rejects an elasticity above 0.4. If \mathcal{E}_W is set equal to the 0.25 value proposed by Diamond and Saez (2011) and Saez et al. (2012), and A_W is set to 2.1, then (31) implies a top tax rate of 65.6 per cent. Lower values for \mathcal{E}_W would imply higher marginal tax rates. In particular, if $\mathcal{E}_W = 0.1$, then the top tax rate is 82.6 per cent.

Fiscal spillovers Saez et al. (2012) and Piketty et al. (2014) emphasize positive fiscal spillovers from income tax rates to other tax bases and modify formula (31)

³²Saez et al. (2012) argue that the best available estimates for the (long run) elasticity of earnings with respect to the retention rate range between 0.12 and 0.4. They select a value of 0.25 as the midpoint of these.

accordingly. These spillovers create motives for even higher marginal tax rates than those reported above. Specifically, [Saez et al. \(2012\)](#) consider shifts of income from the personal to corporate tax bases in response to higher marginal tax rates. They assume that 50 per cent of income is shifted and that this shifted income is taxed at 30 per cent. Their modified version of (31) then implies that the optimal tax rate rises from 72.7 per cent to 76.8 per cent. [Piketty et al. \(2014\)](#) assume that income earners can engage in rent-seeking at the expense of tax revenues. Higher marginal tax rates deter such rent seeking. They suggest that the elasticity of earnings from rent seeking with respect to the retention rate might be at least 0.3 implying an optimal top marginal tax rate of 83 per cent.³³

Much of the CEO literature is, however, consistent with a positive impact of CEO effective labor on firm profits and, hence, a negative spillover from top income taxation to firm profits. In particular, as we have stressed, this is true of the CEO assignment model augmented with CEO effort. We focus upon this.³⁴ At the same time, the literature has given little guidance on the size of this spillover and, in particular, on the magnitude of \mathcal{E}_{Π} . In Appendix C we provide simple regressions of the dividend-GDP ratio on top marginal income tax rates (and corporate tax rates) for the US that are consistent with a small positive value for \mathcal{E}_{Π} .³⁵ These regressions do not, of course, establish a casual relationship between income tax rates and firm profits. Moreover, in the assignment model profit is defined to be the rents accruing to owners of the asset S . These rents exclude payments to adjustable capital and may not be realized contemporaneously with the application of CEO effective labor. Empirical measures of profit must be appropriately adjusted. In Section 5, we pursue a different approach and obtain implications of a calibrated assignment model for optimal (nonlinear) CEO income taxation. This approach implicitly characterizes spillovers to profit.

A restatement of the optimal tax formula. We now restrict attention to the case of quasilinear and constant elasticity CEO preferences and a multiplicative firm

³³[Stantcheva \(2014\)](#) considers a model in which neither the policymaker nor firms observe worker types. Firms offer menus of contracts and workers of different types select amongst them. Taxes perturb this menu, hence inducing further redistributions amongst worker types (rather than between CEOs and firm owners as occurs in our setting). This introduces a term, analogous to $-\chi \frac{\Pi^*}{W^*} \mathcal{E}_{\Pi}^*$, that captures the social desirability of such redistribution into her optimal tax formulas.

³⁴In doing so, we abstract from sensitivity of rent seeking to taxation. Equally, we abstract from other positive benefits of CEO effort such as job creation for non-CEOs.

³⁵Note that even if \mathcal{E}_{Π}^* is small, the impact of τ on profits can be large *relative to its impact on CEO income tax revenues* if $\frac{\Pi^*}{W^*}$ is large enough. In this case, for a large enough value of χ , the depressing effect of τ on firm profits can still translate into a significant depressing effect on optimal tax rates.

surplus function. In this setting, we derive an alternative version of the optimal tax formula in terms of the primitive talent distribution and effort elasticity for the case $\chi = 1$. This formula anticipates the optimal non-linear tax formulas derived in the next section (without the assumption of quasilinearity or a constant elasticity).

The first order condition for τ^* in (27) may be organized as:

$$\frac{\tau^*}{1 - \tau^*} (W^* \mathcal{E}_W^* + \Pi^* \mathcal{E}_\Pi^*) + \left(\frac{\chi - \tau^*}{1 - \tau^*} \right) \Pi^* \mathcal{E}_\Pi^* - \Delta W^* = 0. \quad (32)$$

The first term gives the marginal behavioral impact on tax revenues of a change in $1 - \tau$ if the *entire* firm surplus $R^* = \int_0^{v_0} r^* dv$, $r^*(v) = DS(v)z(v; 1 - \tau^*)$, is taxed at the rate τ^* . The second term adjusts the first to take into account the fact that profits are only taxed at (or valued at) the rate χ . The third term gives the mechanical effect on tax revenues of a change in $1 - \tau$ holding the distribution of CEO incomes fixed. If $\chi = 1$ and CEO preferences are of the quasilinear and constant elasticity form, then $\Delta W^* - \left(\frac{\chi - \tau^*}{1 - \tau^*} \right) \Pi^* \mathcal{E}_\Pi^* = \Delta W^* - \Pi^* \mathcal{E}_\Pi^* = (1 + \mathcal{E}) \int_0^{v_0} r^* \frac{1}{\alpha_h} dv$, where, as before, α_h is the local Pareto coefficient of the talent distribution and \mathcal{E} the effort elasticity. In addition, the first term in (32) is simply $\frac{\tau^*}{1 - \tau^*} R^* \mathcal{E}$. Consequently, when $\chi = 1$ and CEO preferences are of the quasilinear and constant elasticity form, equation (32) implies:

$$\tau^* = \frac{1}{1 + \frac{\mathcal{E}}{1 + \mathcal{E}} \frac{1}{\int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv}}. \quad (33)$$

If the talent distribution has a Paretian right tail, then (33) further reduces to:

$$\tau^* = \frac{1}{1 + \frac{\mathcal{E}}{1 + \mathcal{E}} \alpha_h}. \quad (34)$$

Formulas (33) and (34) anticipate optimal non-linear tax formulas derived in the next section (without the assumption of quasilinearity or a constant elasticity). Strikingly these formulas hold in standard (i.e. non-assignment) labor market settings in which there are no spillovers to profits. Thus, when $\chi = 1$ the standard optimal tax formula expressed in terms of talents (33) holds, but the standard formula expressed in terms of incomes (31) does not and when $\chi = 0$ the situation is reversed. For alternative χ values neither standard formula holds. The logic behind these results is as follows. The standard formula expressed in terms of incomes (31) is valid when the policymaker seeks to maximize income tax revenues; the standard formula expressed in terms of talents (34) is valid when the policymaker seeks to maximize the total surplus not captured by CEOs (or, more generally, high income

earners). In the conventional optimal tax setting, total income tax revenue equals total surplus not captured by high income earners and so both formulas hold. But in the assignment setting some surplus is paid as profit to firm claimants. Hence, either the policymaker is maximizing income tax revenues or she is maximizing total surplus not captured by CEOs (or she is maximizing a weighted sum of the two) and so at most one of the formulas holds.

Note that in the conventional labor supply setting (33) is consistent with a high optimal marginal tax rate. In particular, in this setting $r^* = w^*$ and $(1 + \mathcal{E})\frac{1}{\alpha_h} = \frac{1}{\alpha_w^*}$ implying $\frac{1}{1+\mathcal{E}} \frac{1}{\int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv} = \frac{1}{\int_0^{v_0} \frac{w^*}{W^*} \frac{1}{\alpha_w^*} dv} = A_W^*$. In addition, $\mathcal{E} = \mathcal{E}_W^*$ and so $\frac{\mathcal{E}}{1+\mathcal{E}} \frac{1}{\int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv} = A_W^* \mathcal{E}_W^*$. As noted previously, empirical evaluations of $A_W^* \mathcal{E}_W^*$ are relatively small implying a high tax rate. In contrast, in the assignment setting the talent distribution has a thinner right tail than the income distribution (see equation (21)) and so α_h is larger than α_w^* (and larger than would be implied by attempts to infer α_h from the income data using the standard labor market model). Thus, in the assignment model, $\frac{1}{1+\mathcal{E}} \frac{1}{\int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv} > \frac{1}{\int_0^{v_0} \frac{w^*}{W^*} \frac{1}{\alpha_w^*} dv} = A_W^*$ and (33) is consistent with a lower optimal tax rate than (31).

4 Optimal Non-linear Taxation

We now generalize our earlier results and consider the policymaker's optimal choice of non-linear tax function over (all) CEO incomes. In the general non-linear setting, the simplest and most direct way of deriving optimal nonlinear tax formulas is to formulate the policymaker's problem as a mechanism design problem and then recover optimal taxes from the associated first order conditions. This gives optimal formulas in terms of effort elasticities and (local) Pareto coefficients of the talent and firm asset distributions. We subsequently derive formulas in terms of the CEO income and firm profit distributions and the CEO income elasticity via direct perturbation of the tax function. Note that the latter is complicated relative to [Saez \(2001\)](#) by the endogeneity of the CEO income schedule.

The policymaker's mechanism design problem It is convenient to reformulate a tax equilibrium in terms of a tuple (\tilde{v}, z, w, Φ) , where Φ gives the CEO's utility with $\Phi(v) := U(c(v), z(v)/h(v))$. Let $\mathcal{C}(\phi, z/h)$ be the consumption of a CEO when her utility is ϕ and her effort z/h . We focus on smooth allocations and relax the global CEO and firm incentive constraints (12) and (13), replacing them with, respectively, the

CEO's envelope condition and the firm's first order condition.³⁶ To further simplify matters and to align our work with [Diamond and Saez \(2011\)](#), in the main text we (continue to) focus on the case in which CEOs receive zero welfare weight.³⁷ Let T^0 denote tax revenues (or more generally social surplus) generated by unmatched CEOs. The policymaker's problem can then be formulated as the optimal control problem:³⁸

$$\sup_{\tilde{v}, \Phi, z, w} \int_0^{\tilde{v}} \{ \chi V(S(v), z(v)) + (1 - \chi)w(v) - \mathcal{C}[\Phi(v), z(v)/h(v)] \} dv + T^0 \int_{\tilde{v}}^1 dv \quad (35)$$

subject to $\tilde{v} \in I$, $\Phi(\tilde{v}) = \bar{U}$, $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) \geq 0$, and for $v \in (0, \tilde{v}]$,

$$\Phi_v(v) = -U_e \left(\mathcal{C} \left[\Phi(v), \frac{z(v)}{h(v)} \right], \frac{z(v)}{h(v)} \right) \frac{z(v)}{h(v)} \frac{h_v(v)}{h(v)} \quad (36)$$

$$w_v(v) = V_z(S(v), z(v))z_v(v). \quad (37)$$

The optimal tax formula in terms of primitive distributions After manipulating the first order and co-state equations from the optimal control problem (35) (see [Appendix D](#) for details), the following optimality condition emerges for all $v \in (0, \tilde{v}]$:³⁹

$$V_z + \frac{U_e/h}{U_c} = -\frac{p^\Phi}{h} \left\{ \left[U_{ec} \left(-\frac{U_e}{U_c} \right) + U_{ee} \right] \frac{z}{h} + U_e \right\} \frac{-h_v}{h} + (1 - \chi)vV_{zS}(-S_v), \quad (38)$$

where p^Φ is the co-state associated with CEO utility Φ . This condition captures the marginal benefits and costs associated with a small change in CEO v 's effective labor (holding her utility fixed). It has a very natural interpretation. The left hand side of (38) gives the marginal benefit of the CEO's (compensated) labor supply increase. It consists of the marginal increase in firm surplus V_z less the additional consumption needed to maintain the CEO at her previous utility level $U_e h / U_c$. Relative to the standard Mirrlees model the only modification is that the marginal product of effective labor equals V_z rather than one. The terms on the right hand side capture two sources of welfare loss associated with a small increase in CEO v 's effective labor. The first is the (standard) increment to the information rent paid to

³⁶In our later calculations we check for monotonicity and an absence of bunching ex post at the optimum, ensuring these local conditions are sufficient for incentive compatibility.

³⁷The general case in which CEOs receive non-negative welfare weights is analyzed in [Appendix D](#).

³⁸The case $\chi = 1$ is somewhat easier to solve since w no longer appears in (35). It can be formulated as an optimal control problem with two rather than three state variables.

³⁹All functions below are understood to be evaluated at (the allocation associated with) the CEO's rank v . To economize on notation the v argument is dropped.

more productive CEOs in order that the CEOs' incentive compatibility conditions continue to hold. The second term is novel to the assignment setting. It can be interpreted as the welfare loss associated with a redistribution from CEO income tax revenues to firm profits at firms ranked above v . The economic forces underlying this redistribution, though viewed from the perspective of a local perturbation in labor supply, rather than a global perturbation in the marginal tax rate, are essentially the same as in the linear tax setting. Recall that a unit of effective labor is "priced" by (i.e. paid the marginal product of) the firm for which it is the last unit hired. An increase in CEO effective labor at firm v implies that the additional effective labor is priced by firm v rather than a slightly higher ranked and more productive firm. Thus, its (shadow) price falls and the incomes paid to all greater CEO effective labor supplies are correspondingly reduced. In the associated tax equilibrium, firm profits rise, tax revenues collected from CEOs fall and, if $\chi < 1$, welfare falls. If, on the other hand, $\chi = 1$, because, for example, the policymaker can tax firm profits at 100 per cent, then this redistributive effect has no impact on welfare. In this case, the policy objective is to maximize surplus extracted from CEOs, and (38) reduces to the standard optimality condition in Mirrlees models (with zero weight on top earners and modulo the change to the marginal product of effective labor on the left hand side).

Together the CEOs' and the firms' first order conditions (14) and (16) imply:

$$(1 - T_w[w])V_z h U_c = -U_e. \quad (39)$$

The optimal marginal tax rate must be set to align a CEO's private return to effort with the social return. Since the CEO's effective wage coincides with the firm's marginal product, this reduces to ensuring that her pre-tax return on effort equates to the right hand side of (38). If $\chi = 1$, then the right hand side of (38) equals the usual marginal informational rents term from the Mirrlees model and combining (38) and (39) and the definition of α_h the standard formula for optimal marginal tax rates obtains:

$$T_w[w] = \frac{1}{1 + \frac{1}{\tilde{p}^\Phi} \frac{\varepsilon^c}{1 + \varepsilon^u} \alpha_h}, \quad (40)$$

where \tilde{p}^Φ is the normalized co-state $\frac{U_c}{1-F(h)} p^\Phi$. More generally, using (38) and (39)

and the definitions of α_h and α_S , the optimal marginal tax rate is:

$$T_w[w] = \frac{1 + (1 - \chi) \frac{1}{\bar{p}^\Phi} \frac{\mathcal{E}^c}{1 + \mathcal{E}^u} \frac{V_{zS}}{V_z} \frac{\alpha_h}{\alpha_S}}{1 + \frac{1}{\bar{p}^\Phi} \frac{\mathcal{E}^c}{1 + \mathcal{E}^u} \alpha_h}. \quad (41)$$

Intuitively, when $\chi < 1$ and the policymaker places more weight on CEO income tax revenues than firm profits, then the second marginal cost term in (38) becomes relevant. In this case, consistent with the intuition described above, the policymaker sets a higher marginal tax rate on a CEO's income in order to reduce that CEO's effective labor supply, modify the (implicit) pricing schedule for effective labor and, hence, redistribute from firm profits to the incomes of more talented CEOs. In doing so the policymaker achieves her goal of collecting more tax revenues from CEOs.

The optimal tax formula in terms of the income distribution In the linear tax setting we presented an optimal tax formula in terms of the elasticities and tail coefficients of CEO incomes and firm profits. This formula was obtained from a direct perturbation of the optimal tax function. Such perturbations are more complicated in the nonlinear setting. We pursue such a perturbation now and, hence, relate equations (40) and (41) to nonlinear tax formulas in terms of induced CEO incomes and (spillovers to) firm profits. We focus on the quasilinear/ constant elasticity CEO preferences and multiplicative firm objective case. Additional details are given in Appendix F.

The perturbation we consider involves a small modification to the marginal tax rate over a small interval of incomes starting at income w_0 . Following Saez (2001), its impact on tax revenues can be decomposed into “mechanical” and “behavioral” parts. The first of these equals $1 - M(w_0)$; it gives the revenue response to the perturbation holding the CEO income schedule w fixed. The second component is the behavioral part. This is more complicated than in standard models since it incorporates the impact of the tax change on the equilibrium schedule of incomes. It may be expressed compactly as:

$$-\frac{T_w[w_0]}{1 - T_w[w_0]} m(w_0) w_0 \tilde{\mathcal{E}}_w(w_0)$$

where $\tilde{\mathcal{E}}_w(w_0)$ is a weighted elasticity of CEO incomes at and above w_0 with respect to the local retention rate $1 - T_w[w_0]$. The precise formula for this elasticity is given in Appendix F. In addition to the usual “local” effect on tax revenues caused by

the CEO at w_0 working harder in response to a higher retention rate, this elasticity also incorporates a “global” effect on revenues collected from CEOs earning more than w_0 . This is caused by the downward shift in the equilibrium income schedule discussed previously.

When $\chi = 0$, the policymaker is concerned only with maximizing CEO income tax revenues. In this case the sum of the mechanical and behavioral impacts derived above must equal zero at the optimum and so combining terms:

$$T_w^*[w_0] = \frac{1}{1 + \alpha_w(v_0)\tilde{\mathcal{E}}_w(w_0)},$$

where v_0 is the rank of the CEO earning w_0 and $\alpha_w(v_0)$ is the corresponding local Pareto coefficient of income.

For χ values greater than 0, the policymaker is also concerned with the spillover of marginal tax rates to firm profits. Consequently, the behavioral term is augmented with an extra component that captures the enhancing effect of a higher retention rate at w_0 on the profits of firms ranked above v_0 . Again this enhancement is due to the reduction in CEO incomes at each effective labor supply above z_0 . It is given by:

$$\begin{aligned} & -\chi \frac{1}{1 - T_w[w_0]} \frac{\mathcal{E} \frac{S_0 z_0}{w_0}}{1 + \mathcal{E} \mathcal{T}(w_0) \frac{S_0 z_0}{w_0}} \frac{\alpha_w(v_0)}{\alpha_S(v_0)} \\ & \times \int_0^{v_0} \exp \left\{ - \int_v^{v_0} \frac{\frac{S_v(v')}{S(v')} \mathcal{E} \mathcal{T}(w^*(v')) \frac{S(v') z^*(v')}{w(v')}}{1 + \mathcal{E} \mathcal{T}(w^*(v')) \frac{S(v') z^*(v')}{w^*(v')}} dv' \right\} dv, \end{aligned} \quad (42)$$

where $\mathcal{T}(w) = \frac{w T_{ww}[w]}{1 - T_w[w]}$ is the elasticity of the marginal retention function. The term (42) is the analogue for the nonlinear setting of the spillover term $\chi \frac{\Pi}{1-\tau} \mathcal{E}_\Pi$ in the derivation of (30). Adding it to the other behavioral and mechanical terms leads to an optimal tax equation for the non-linear setting analogous to (30).

5 Quantitative Evaluation of Optimal Non-Linear Taxes

Equation (41) may be used to compute optimal marginal taxes over a range of CEO incomes. We specialize the analysis to quasi-linear/constant elasticity CEO preferences and a multiplicative firm surplus function. In this case (41) may be

rewritten as:

$$T_w^*[w^*(v)] = \frac{1 + (1 - \chi) \frac{\mathcal{E}}{1 + \mathcal{E}} \frac{\alpha_h(v)}{\alpha_s(v)}}{1 + \frac{\mathcal{E}}{1 + \mathcal{E}} \alpha_h(v)}. \quad (43)$$

To quantitatively evaluate the implications of this formula for optimal taxes values for the weight χ , the Pareto coefficients α_h and α_s and the elasticity \mathcal{E} are required. We discuss the choice of χ first.

5.1 Selecting Values for χ

As described in Section 3 the weight χ is the social marginal value of profit. In our baseline model, it corresponds simply to the profit tax rate, τ^F . We consider a range of values for χ between the empirically observed profit tax rate in the US and the optimal profit rate of 1. In the US, corporate profits over much of the tax schedule are taxed at 35 per cent. These are augmented by state level taxes and by dividend taxes placed on disbursed profits. The OECD reports that the overall tax rates on dividend income in the US was 57.6 per cent in 2015.⁴⁰ Since it is possible that statutory rates overstate the effective tax rate on profits, we consider values of χ as low as 0.4. Note that χ also incorporates direct social concern for profit claimants and, in more elaborate models, the social value of relaxing constraints on firm entry. These considerations raise χ above the profit tax rate and in some models increase χ to one.

The extent of social concern for firm claimants is a normative and ethical question, though one likely influenced by the incomes of firm claimants. Ownership of corporate equities extends well beyond top CEOs (to whom we continue to attach zero weight) and to those with much lower incomes. To see this, we use data extracted from the *Survey of Consumer Finances* (SCF) for 2013. The SCF provides a measure of directly and indirectly held equities (equities in stock mutual funds, IRAs/Keoghs and other managed assets). The total value of equities held by households in the SCF equals \$15,904 billion.⁴¹ Overall, 65 per cent of the reported

⁴⁰This value includes the marginal statutory corporate income tax rate on distributed profits and the sum of the maximum federal personal and average state income tax rate on dividends. Data taken from OECD Tax Database, Corporate and Capital Income Taxes: Table II.4. Historical values for the US range from 51.7 per cent in 2003 to 88.7 in 1981. The average value for OECD countries in 2015 is 43.1 per cent.

⁴¹ The *Flow of Funds* reports that in 2013 the total market value of domestic US corporations is equal to 27,183 billions of US dollars (Table L.223 - Lines 2+3). Total direct and indirect ownership of the household sector amounts to 19,502 billions (Table L.223 - Lines 6+17). Additional amounts are held by pension funds (4,888 billions) (Table L.223 - Lines 14+15+16) and insurance and life

value of equity in the SCF is in the hands of households with incomes of less than \$500,000. The median household in the SCF reports an income of about \$51,700 and equity (held directly and indirectly) valued at \$33,000 and constituting 40 per cent of its financial wealth.

5.2 Recovering Measures of α_h and α_S

In our firm-CEO assignment model, equations (19) and (20) relate the Pareto coefficients for the CEO talent distribution and firm assets to those for CEO incomes and firm profits. We seek to use these equations to determine α_h and α_S . There are two complications in doing so. First use of (19) and (20) requires measurement of DSz , i.e. measurement of firm surplus after payment to (non-CEO) adjustable inputs, and, hence, measurement of economic profit. A firm's market capitalization combines the capitalized value of such surpluses (net of payments to its CEO) with the value of the firm's adjustable capital. Recovery of the value of firm surpluses, thus requires disentangling these from the value of adjustable capital. Furthermore, the surplus from a given application of CEO effective labor may be realized over time. Specifically, a given set of CEO decisions may have a long lasting impact on a firm's stream of surpluses. DSz corresponds to the value of this stream rather than the contemporaneous value of surplus. To handle these issues, we follow the procedure of Terviö which requires introducing two parameters describing the share of gross surplus paid to adjustable capital and the rate of decay of CEO effective labor on surplus. We select the parameter values suggested by Terviö (2008) and undertake sensitivity analysis around them. A second issue in applying (19) and (20) concerns the decomposition of effective labor into its talent and effort components. This hinges on the elasticity of CEO effort. We make conservative choices in this regard and undertake sensitivity analysis around them. Our choices are consistent with modest values for the elasticity of taxable income and with moderate variation in effort across the population of CEOs.

Connecting α_h and α_S to firm market capitalization and CEO income data

Along the lines of Terviö (2008), we first provide a simple dynamic extension of the environment in Section 2 that accommodates productivity growth and long-lived effects of CEO effort. Specifically, we assume that (i) firms are infinitely lived and firm productivity grows at a constant and common rate g , (ii) CEOs marginal utility of consumption decays at a steady rate g over time and (iii) the CEO's outside util-

insurance companies (2,053 billions) (Table L.223 - Lines 12+13).

ity option is constant. In addition, we assume that the tax function in successive periods is linear above a productivity-adjusted threshold income $(1+g)^t w_0$ and that the tax levied at this threshold grows at a rate g , i.e. $T_t[(1+g)^t w_0] = (1+g)^t T[w_0]$. This assumption implies that if a CEO's income grows at a rate g and has an initial value in excess of w_0 , then the CEO's tax liabilities also grow at this rate. In addition, we assume that the impact of CEO effective labor exerted at t has a long-lived effect on firm surplus that decays at rate λ . The latter assumption implies that if a firm buys a stream of effective labor $\{z_r\}_{r=-\infty}^{\infty}$ from a sequence of CEOs, the effective labor applied at date t within the firm is:

$$Z_t = \lambda \sum_{i=0}^{\infty} \frac{z_{t-i}}{(1+\lambda)^{i+1}}. \quad (44)$$

These assumptions correspond to (and extend) the “strong stationarity conditions” of [Terviö \(2008\)](#). Combined with quasilinear-constant elasticity CEO preferences and a multiplicative firm surplus, they ensure a stationary equilibrium in which firm surpluses and CEO incomes scale up by a factor of $1+g$ in each period and the effective labor supplies of CEOs remain constant.

Assume that $\{CSZ_t(1+g)^t\}^{1-\theta} k_t^\theta$ gives the firm payoff at date t net of payment to and after maximization over all adjustable inputs except capital. If r denotes the (after-tax) rental cost of adjustable capital, then a firm with asset S using effective labor Z_t obtains a surplus at t of:

$$\hat{V}_t(S, Z_t) := \max_{k_t \geq 0} \{CSZ_t(1+g)^t\}^{1-\theta} k_t^\theta - rk_t, \quad (45)$$

After maximization over k_t (and with appropriate choice of the constant C and units for S), the date t surplus has the multiplicative form $\hat{V}_t(S, Z_t) = (1+g)^t SZ_t$. Hence, the present discounted value of firm surpluses at date t is:

$$(1+g)^t \sum_{j=0}^{\infty} B^j SZ_{t+j}, \quad (46)$$

where $B = \frac{1+g}{1+r}$ is the growth-adjusted firm discount factor. Since the effective labor z_t supplied by a CEO at t has a long-lasting effect, the (present discounted value of the) surplus it generates is:

$$V_t(S, z_t) := (1+g)^t DSz_t, \quad \text{with } D := \frac{\lambda}{1+\lambda-B}. \quad (47)$$

In a stationary equilibrium, the income received by the v -th CEO grows at the rate g , i.e. $w_t(v) = (1 + g)^t w(v)$ and the v -th firm chooses v' to maximize $(1 + g)^t \{DS(v)z(v') - w(v')\}$. As in previous sections, its first order condition implies:

$$\frac{w_v}{w} = \left(\frac{DSz}{w} \right) \frac{z_v}{z}. \quad (48)$$

Equation (48) is central to our two step identification strategy. The first step (as in [Terviö \(2008\)](#)) is to relate firm surplus DSz to (observed) market capitalization. The second step (new to this paper) is to isolate the component of effective labor variation z_v/z that is due to talent variation and substitute out that part which is not. We describe these steps next.

In the stationary equilibrium, the v -th firm matches with the v -th CEO and hires a constant amount of effective labor $z(v)$. Thus, the effective labor used by the firm at t equals that supplied by the CEO at t and $Z_t(v) = z(v)$. Consequently, the capitalized value of the v -th firm's profits at date 0 is:

$$P(v) := \frac{S(v)z(v) - w(v)}{1 - B}. \quad (49)$$

The market capitalization of the v -th firm at date 0 augments $P(v)$ with the date 0 adjustable capital choice $k_0(v)$ and is given by $Q(v) := k_0(v) + P(v)$. Date 0 adjustable capital $k_0(v)$ may be recovered from (45). Substituting this optimal choice and the definition of $P(v)$ in (49) into the definition for the market capitalization $Q(v)$ implies:

$$Q(v) = \left(\frac{1}{\Xi} \right) \frac{S(v)z(v)}{1 - B} - \frac{w(v)}{1 - B}, \quad (50)$$

where $\Xi := \frac{1-\theta}{1-\theta+\frac{\theta}{r}(1-B)}$ inflates the firm surplus to account for adjustable capital. Hence, the firm's first order condition (48) can be re-expressed as:

$$\frac{w_v}{w} = \left(\frac{DSz}{w} \right) \frac{z_v}{z} = D\Xi \left(\frac{w + (1 - B)Q}{w} \right) \frac{z_v}{z}, \quad (51)$$

where (50) is used to replace the unobserved firm surplus Sz with observed market capitalization Q and obtain the second equality. In our setting, CEO effective labor variation z_v/z is attributable partly to variation in talent h_v/h and partly to variation in effort e_v/e . The latter in turn is related to variation in the return to effort which depends on CEO talent and firm asset size. Specifically, and as in the static case, if the tax system in period 0 is linear across high incomes, then the v -th CEO's first

order condition implies:

$$\frac{z_v}{z} = \frac{h_v}{h} + \frac{e_v}{e} = (1 + \mathcal{E}) \frac{h_v}{h} + \mathcal{E} \frac{S_v}{S}. \quad (52)$$

Together (51) and (52) relate CEO income variation w_v/w to CEO talent h_v/h and firm asset size variation S_v/S . It remains to eliminate $\mathcal{E}S_v/S$ from this, i.e. to eliminate that part of effort variation attributable to (unobserved) firm asset variation. This is done by totally differentiating (50) to obtain an expression for S_v/S and combining the result with (51) and (52) to give a differential equation for CEO talent in terms of observable CEO income and firm market capitalization. Re-expressing this in terms of local Pareto coefficients (with α_Q the local Pareto coefficient for firm market capitalization), yields:

$$\frac{1}{\alpha_h} = \mathcal{N} \left(\frac{w}{w + (1 - B)Q} \right) \frac{1}{\alpha_w} + \mathcal{P} \left(\frac{(1 - B)Q}{w + (1 - B)Q} \right) \frac{1}{\alpha_Q}, \quad (53)$$

where $\mathcal{N} > 0 > \mathcal{P}$ are constants depending on the parameters λ , r , g , θ and \mathcal{E} . Equation (53) is the empirically operational version of (21). It implies that the reciprocal Pareto coefficient for CEO talent is a weighted sum of the coefficients for CEO income and firm market capitalization (rather than CEO income and profit as in (21)). It is distinct from analogous expressions in [Terriö \(2008\)](#) in that it purges out local variation in CEO income due to variation in effort. In particular, and analogous to the weighting of the profit Pareto coefficient in (21), the weight $\mathcal{P} = -\frac{\mathcal{E}}{1+\mathcal{E}}$ on the reciprocal Pareto coefficient for firm market capitalization, $\frac{1}{\alpha_Q}$, is negative. Heuristically, greater variation in market capitalization is associated with greater variation in firm asset size and a correspondingly greater contribution of CEO effort variation to firm effective labor variation. This in turn implies a smaller role for CEO talent variation in explaining CEO effective labor and income variation.

Estimates of α_S are also needed. A derivation very similar to that underpinning (53) yields:

$$\frac{1}{\alpha_S} = \mathcal{M} \left(\frac{w}{w + (1 - B)Q} \right) \frac{1}{\alpha_w} + \left(\frac{(1 - B)Q}{w + (1 - B)Q} \right) \frac{1}{\alpha_Q}, \quad (54)$$

where $\mathcal{M} < 0$ depends upon parameters.

Connecting CEO income and firm market capitalization data We use (53) and (54) to recover estimates of the local Pareto coefficients for talent and firm size. This requires prior estimation of the local Pareto coefficients for CEO compensation and firm market capitalization along with calculation of the parameters

B , \mathcal{N} , \mathcal{P} and \mathcal{M} . We use CEO compensation data for the year 2011 taken from the Standard and Poor’s ExecuComp database. The measure of compensation considered includes the amounts received by a CEO (within a fiscal year) from salary, bonus, restricted stock grants and an evaluation of long term incentive pay. The value of options received as compensation is calculated by determining the profit obtained at the time the options are exercised. We restrict our sample to CEOs with reported income above \$500,000 2011 US dollars. The final sample contains 1683 CEOs. We compute firm market capitalizations using data on the number of shares outstanding and average monthly share price contained in the Center for Research in Security Prices (CRSP) database. The model predicts a perfect ordering between CEO income and market capitalization. This relationship, known as Roberts’ Law, is robust in the data, but is obviously not perfect. To bring the model to the data, we order by CEO income. We then impute corresponding market capitalizations by estimating the following log-linear relationship:

$$\log Q_i = \beta_0 + \beta_1 \log w_i + \varepsilon_i,$$

where w_i is the income of the i -th CEO and Q_i is the market capitalization of the firm the i -th CEO manages. Using the estimated coefficients $(\hat{\beta}_0, \hat{\beta}_1)$, market capitalization values are set equal to:⁴²

$$\log \hat{Q}_i = \hat{\beta}_0 + \hat{\beta}_1 \log w_i.$$

We find that the right tail of the CEO income and imputed market capitalization distributions are well described by Pareto distributions. We fit Pareto distributions to both and, hence, recover estimates of α_w and α_Q .

Selecting parameter values We follow [Terviö \(2008\)](#) and select values for $g = 0.025$, $r = 0.05$, $\lambda = 0.5$ and $\theta = 0.4$. Robustness tests of our results with respect to these parameters are performed in Subsection 5.4. We choose a benchmark value of $\mathcal{E} = 1/15$, but also consider other elasticity values in the range 1/20 to 1/10. Collectively, these choices pin down the parameters B , \mathcal{N} and \mathcal{P} in (53) and \mathcal{M} in (54). Note that the data and the requirement that α_h be positive place some restrictions on parameter choices. In particular, given the benchmark parameterization

⁴²We experimented with a wide variety of smoothing techniques, in all cases the magnitude and behavior of the estimated α_h and the computed optimal tax rates are little changed. Ordering by firm market capitalization and then smoothing to obtain imputed CEO compensation leads to somewhat higher values for the local Pareto coefficients of talent. This, in turn, implies even lower values for optimal tax rates than we obtain.

of [Terviö \(2008\)](#), positivity of the talent Pareto coefficient requires that the elasticity \mathcal{E} be below 0.12.⁴³ Our choices for \mathcal{E} are consistent with this bound. If, in addition to our CEO preference assumptions, the tax rate is linear above w_0 , then $\mathcal{E}_W = \Gamma \mathcal{E}$, where $\Gamma = 1 + \frac{D\Xi(1-B)Q(0)+(D\Xi-1)w(0)}{W}$. Our data and benchmark parameter choices imply a value of $\Gamma = 1.53$ and, hence, a value of \mathcal{E}_W approximately equal to 0.1. The alternative values for \mathcal{E} we consider imply a range for \mathcal{E}_W between 0.0765 and 0.153. These values align with the estimates of [Frydman and Molloy \(2011\)](#). They obtain point estimates of \mathcal{E}_W below 0.1 and reject values above 0.2. Our implied values for \mathcal{E}_W are also in line with those of [Gruber and Saez \(2002\)](#) who obtain an elasticity of broad income (before deductions) of 0.12 for the general population and 0.17 for high income earners. They are more conservative than the 0.25 value chosen by [Diamond and Saez \(2011\)](#). In Appendix H we compute the entire optimal equilibrium allocation under our benchmark parameterization. Our calculations imply that effort variation from the lowest to the highest ranked CEO is of the order of 40 per cent. [Bandiera et al. \(2011\)](#) is one of the few studies to explore effort variation amongst CEOs. They study a sample of Italian CEOs and report variation in hours worked of about 50 per cent between the CEO's at the 90th and the 10th percentile rank by hours worked. They also confirm a positive relationship between CEO hours and firm productivity. Our benchmark parameterization, thus, generates variation in effort of a similar order of magnitude to the variation in hours found by [Bandiera et al. \(2011\)](#) albeit at the optimum rather than the prevailing (Italian) tax system.^{44,45}

Recall from (29) that under our assumptions, $\mathcal{E}_\Pi = \frac{1}{A_\Pi} \mathcal{E}$. We recover profits from (49) and estimate a value for $\frac{1}{A_\Pi}$ equal to 0.53. Hence, under our benchmark \mathcal{E} choice, $\mathcal{E}_\Pi = 0.035$. This small positive elasticity is of a similar order of magnitude to our profit elasticity estimates for the US in Appendix C. Overall, our assumptions regarding the effort elasticity are relatively conservative by the standards of the literature, but generate implications for effort variation and the aggregate profit

⁴³Intuitively, the relatively fat empirical right tail for market capitalization implies a relatively rapid increase in S and, hence, price of effective labor across rank. The empirical right tail for CEO income then requires that CEO effective labor (the product of talent and effort) does not increase too quickly with rank. In particular, for the talent Pareto coefficient to remain positive (and talent to decrease with v), effort cannot be too responsive to the return to effort.

⁴⁴[Bandiera et al. \(2011\)](#) find that a CEO in the 90th percentile (in terms of hours worked) works 20 hours (and about 50 per cent) more than a CEO in the 10th percentile. They find a positive relationship between CEO hours and a firm's labor productivity. Similar results are found in [Bandiera et al. \(2013\)](#) looking at Indian manufacturing firms. Finally, [Bandiera et al. \(2014\)](#) looks at a smaller sample of recent CEOs (mostly in smaller firms) in the US. This study also documents a positive relationship between firm size and hours worked.

⁴⁵To the extent that the optimal tax system is regressive and the actual one linear or progressive over top incomes, it is likely to enhance variation in effort and hours.

elasticity that are broadly consistent with the data.

Empirical characterization of the tail of the talent distribution The calculated local Pareto coefficients for CEO talent under our benchmark parameterization are plotted in Figure 1. They show a sharp escalation consistent with a thin right tail to the talent distribution. They are drastically different from those for CEO incomes which were stable and consistent with a right Pareto tail to the income distribution with a Pareto coefficient of 2.1. These findings are in line with [Terviö \(2008\)](#) and [Gabaix and Landier \(2008\)](#) who provide corroborating evidence (in models without CEO effort or taxes) that the talent distribution is thin tailed.⁴⁶

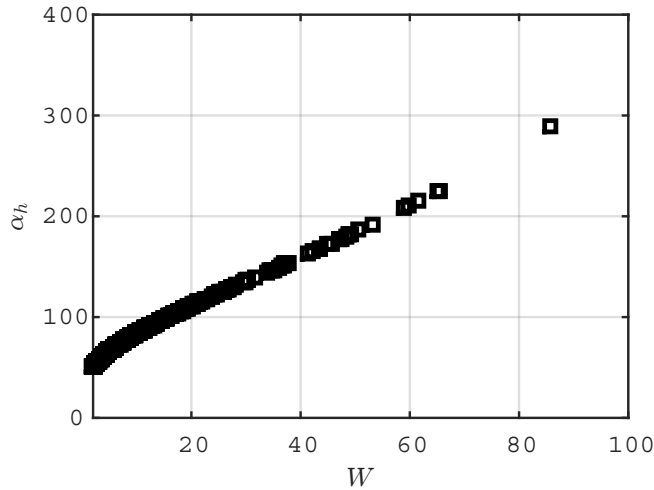


Figure 1: Estimates of the local Pareto coefficient for talent α_h as a function of corresponding CEO income.

5.3 Results

Substituting our empirical values for α_h and α_s into (43) along with values for \mathcal{E} and χ gives optimal marginal tax rates as function of CEO rank v .⁴⁷ Figure 2 shows these tax rates for the effort elasticity $\mathcal{E} = 1/15$ and for various values of χ . Those

⁴⁶Previous versions of the paper formally estimated the tail properties of the talent distribution. For all estimators considered the distribution was determined to be thin tailed and best characterized by a Weibull-like distribution.

⁴⁷Calculation of optimal marginal tax rates as a function of v using (43) does not require calculation of the entire optimal allocation. In Appendix H we compute the optimal allocation under a benchmark parameterization. We also undertake several counterfactual exercises suggested by a referee that indicate the respective roles of talent and effort in inducing equilibrium CEO income variation.

for $\chi = 0.6$ can be interpreted as optimal in the absence of a reform of current US profit taxes (assuming statutory rates are paid and with no additional concern for profits beyond their role as a direct source of revenue); those for $\chi = 1$ as optimal under a full reform of CEO income and firm profit taxation. The figure makes

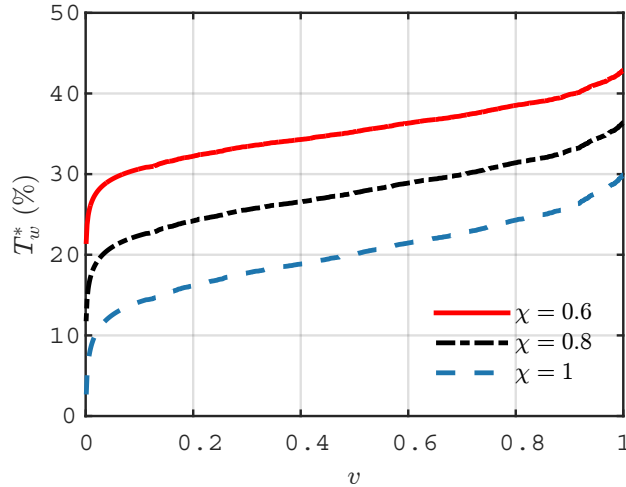


Figure 2: Marginal tax rates as function of v .

apparent that marginal tax rates are falling with χ and for $\chi = 1$ are low across all CEO ranks.

It is useful to relate these optimal marginal tax rates to equilibrium CEO incomes rather than rank. However, if $\chi = 1$, only the difference between a CEO's income and that of the least talented active CEO, $\Delta w^*(v) = w^*(v) - w^*(\tilde{v})$, is determined in the optimal tax equilibrium. The income $w^*(\tilde{v})$ is not determined, because the policymaker is indifferent between the realization of surplus as profit at the smallest active firm or as tax revenues taken from the least talented CEO. The policymaker is constrained only by the requirements that $w^*(\tilde{v}) - T[w^*(\tilde{v})]$ is above the consumption level needed to keep the CEO in the market and that firm profit is non-negative. For the $\chi < 1$ case, CEO income is only determined up to the inability to place a lump sum tax on the smallest firm's profit. Given this indeterminacy, Table 1 reports optimal marginal tax rates as functions of $\Delta w^*(v) = w^*(v) - w^*(\tilde{v})$.⁴⁸ The table highlights that over large ranges of high (but thinly populated) CEO incomes optimal marginal tax rates are declining in both income and χ . At χ equal to 0.6 marginal tax rates are slightly above 30 per cent on high incomes and declining

⁴⁸ $\Delta w^*(\tilde{v})$ is obtained by integrating out (17) using the values of the local Pareto coefficients previously obtained. If $w^*(\tilde{v})$ is normalized to the empirical value of $w(\tilde{v})$, then $w^*(v) = \Delta w^*(v) + \0.5 million.

Table 1: Marginal Tax Rates (per cent)

	Δw^* (in millions)		
	30	40	60
$\chi = 0.4$	40.2	39.3	38.4
$\chi = 0.6$	32.2	31.2	30.1
$\chi = 0.8$	24.2	23.0	21.8
$\chi = 1$	16.2	14.9	13.4

Notes: Optimal non-linear tax rates as function of societal weight χ and income level. Δw^* measured in millions of 2011 USD.

in income, while at $\chi = 1$ they average 15 per cent and are declining.

5.4 Robustness

Focusing on the case $\chi = 0.8$, we now undertake robustness analysis with respect to the parameters λ , θ and g . Recall that these parameters affect tax rates via their impact on the estimated values of α_h and α_s .⁴⁹ Table 2 reports results. Qualitatively, the responses of tax rates are as expected. For example, a higher value of $\mathcal{E} = 0.1$ (implying $\mathcal{E}_W = 0.153$) implies that optimal tax rates fall below 7%. On the other hand, a higher value of θ implies a lower share of CEO effective labor in the determination of firm market capitalization. This in turn lowers the distortionary effect of taxes on firm profits and leads to higher optimal marginal tax rates. Although variation in parameters affects optimal taxes, in all cases considered computed optimal marginal tax rates are much lower than the rates of 70 or 80 per cent proposed in the literature.

6 Conclusion

This paper considers the optimal taxation of top earning CEOs. To that end it extends optimal tax theory to an assignment setting in which firms buy CEO labor and some of the surplus generated in production accrues to firm claimants. The classic Diamond-Saez formula continues to prescribe very high marginal tax rates on top CEO incomes of over 70 per cent and sometimes 80 per cent, but it is only applicable if the policymaker has no social concern for profits. This in turn requires that the planner has no ability to tax profits, no direct social concern for

⁴⁹Specifically, in equations (53) and (54): $\mathcal{N} := \frac{1}{D\Xi} - \frac{\mathcal{E}}{1+\mathcal{E}}$, $\mathcal{P} := -\frac{\mathcal{E}}{1+\mathcal{E}}$ and $\mathcal{M} = 1 - \frac{1}{D\Xi}$.

Table 2: Marginal Tax Rates (per cent): Robustness Analysis

Parameter	Values	Δw^* (in millions)		
		30	40	60
θ	0.2	16.1	15.1	13.8
	0.4	24.2	23.0	21.8
	0.6	36.5	35.4	34
g	0.015	14.0	12.9	11.5
	0.025	24.2	23.0	21.8
	0.035	40.1	39.2	37.8
λ	0.1	32.0	31.0	29.7
	0.5	24.2	23.0	21.8
	2.0	22.5	21.3	20.1
\mathcal{E}	1/10	6.8	5.6	4.2
	1/15	24.2	23.0	21.8
	1/20	36.1	35.0	33.8

Notes: Δw^* measured in millions of 2011 USD. For each parameter value (θ, g, λ) the middle row represents our benchmark calculations as reported in Table 1.

firm claimants and that depressing profits has no adverse affects on firm entry. In the non-linear tax setting, a Mirrlees formula (expressed in terms of the CEO effort elasticity and a weighted tail coefficient of the CEO talent distribution) is valid if a comprehensive reform of CEO income and profit taxation is pursued. Our quantitative analysis suggests that the right tail of the CEO talent distribution is thin and that the optimal marginal tax in this case is around 16 per cent on incomes above \$30 million or so and lower on higher incomes. If profit taxation was left as is in the US and only CEO income taxation was reformed (and there was no direct social concern for firm claimants or impact of aggregate profits on firm entry), then the optimal marginal tax rate would be around 32 per cent and 40 per cent on incomes above \$30 million.

Our paper integrates an assignment model into a normative public economics framework. In this context it has emphasized the impact of CEO income taxes on firm profits and profit tax revenues. However, its broader message is that empirical assessment of the fiscal spillovers associated with top earners is critical for determining the top tax rate. More remains to be done in this direction. First, the limits to profit taxation (and, hence, the extent of the spillover from CEO effort to firm profit tax revenues) remain to be further explored and quantified. Potentially, profit taxes interact with financing frictions, encourage firms to realize profits out-

side of a tax jurisdiction and deter creation of the firm asset S in the first place. Further quantitative analysis of CEO income and profit taxation that takes explicit account of these various factors would be valuable. Second, our model abstracts from any impact of CEO effort on the demand for and, hence, wages of (socially) valued workers. Thus, it omits a potential motive for even lower marginal tax rates on top incomes than those reported here.⁵⁰ Analysis of this effect requires embedding the CEO assignment model into a general equilibrium framework that explicitly incorporates workers.⁵¹ Finally it remains to explore the quantitative implications of assignment for taxation in other top earner settings, such as entertainers, athletes and entrepreneurs. We leave analysis of these issues for future work.

References

- Ales, L., A. Bellofatto, and J. Wang (2015). Taxing Atlas: Using firm data to derive optimal income tax rates. Mimeo. Tepper School of Business, Carnegie Mellon University.
- Ales, L., M. Kurnaz, and C. Sleet (2015). Technical change, wage inequality and taxes. *American Economic Review* 105(10), 3061–3101.
- Bakija, J., A. Cole, B. T. Heim, et al. (2012). Jobs and income growth of top earners and the causes of changing income inequality: Evidence from us tax return data. Unpublished manuscript, Williams College.
- Bandiera, O., L. Guiso, A. Prat, and R. Sadun (2011). What do CEOs do? CEPR Discussion Paper No. DP8235.
- Bandiera, O., A. Prat, and R. Sadun (2013). Managerial capital at the top: Evidence from the time use of CEOs. Working paper.
- Bandiera, O., A. Prat, R. Sadun, J. Wulf, et al. (2014). Span of control or span of activity? Harvard Business School Strategy Unit Working Paper 12-053, 14–22.
- Diamond, P. and J. Mirrlees (1971). Optimal taxation and public production ii: Tax rules. *American Economic Review* 61(3), 261–278.
- Diamond, P. and E. Saez (2011). The case for a progressive tax: From basic research to policy recommendations. *The Journal of Economic Perspectives* 25(4), 165–190.

⁵⁰In the opposite direction, we have omitted pure rent-seeking on the part of CEOs of the sort emphasized by [Piketty et al. \(2014\)](#), a force for higher marginal tax rates.

⁵¹See [Scheuer and Werning \(2015\)](#) for qualitative work in this direction.

- Edmans, A. and X. Gabaix (2009). Is CEO pay really inefficient? A survey of new optimal contracting theories. *European Financial Management* 15(3), 486–496.
- Frydman, C. and D. Jenter (2010). CEO compensation. *Annual Review of Financial Economics* 2(1), 75–102.
- Frydman, C. and R. S. Molloy (2011). Does tax policy affect executive compensation? Evidence from postwar tax reforms. *Journal of Public Economics* 95(11), 1425–1437.
- Gabaix, X. and A. Landier (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics* 123, 49–100.
- Gayle, G.-L., L. Golan, and R. A. Miller (2015). Promotion, turnover, and compensation in the executive labor market. *Econometrica* 83(6), 2293–2369.
- Gayle, G.-L. and R. A. Miller (2009). Has moral hazard become a more important factor in managerial compensation? *American Economic Review* 99(5), 1740–1769.
- Goolsbee, A. (2000). What happens when you tax the rich? Evidence from executive compensation. *Journal of Political Economy* 108(2), 352–378.
- Gruber, J. and E. Saez (2002). The elasticity of taxable income: evidence and implications. *Journal of Public Economics* 84(1), 1–32.
- Lockwood, B., C. Nathanson, and E. G. Weyl (2014). Taxation and the allocation of talent. Available at SSRN 1324424.
- Munk, K. (1978). Optimal taxation and pure profit. *Scandinavian Journal of Economics* 80(1), 1–19.
- Munk, K. (1980). Optimal taxation with some non-taxable commodities. *Review of Economic Studies* 47(4), 755–765.
- Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal taxation of top incomes: A tale of three elasticities. *American Economic Journal: Economic Policy* 6(1), 230–271.
- Rosen, S. (1982). Authority control and the distribution of earnings. *Bell Journal of Economics* 13(2), 311–323.

- Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the Roy model. *Quarterly Journal of Economics* 128(2), 623–662.
- Rothschild, C. and F. Scheuer (2014). A theory of income taxation under multidimensional skill heterogeneity. Mimeo. Stanford University.
- Rothschild, C. and F. Scheuer (2015). Optimal taxation with rent-seeking. Working Paper, Stanford University.
- Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies* 68(1), 205–229.
- Saez, E., J. Slemrod, and S. H. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature* 50(1), 3–50.
- Scheuer, F. and I. Werning (2015). The taxation of superstars. Unpublished Working Paper.
- Stantcheva, S. (2014). Optimal income taxation with adverse selection in the labour market. *The Review of Economic Studies* 81(3), 1296–1329.
- Stiglitz, J. (1982). Self selection and Pareto efficient taxation. *Journal of Public Economics* 17(2), 213–240.
- Stiglitz, J. and P. Dasgupta (1971). Differential taxation, public goods and economic efficiency. *Review of Economic Studies* 38(2), 151–174.
- Terviö, M. (2008). The difference that CEOs make: An assignment model approach. *American Economic Review* 98(3), 642–668.
- Zucman, G. (2014). Taxing across borders: Tracking personal wealth and corporate profits. *The Journal of Economic Perspectives* 28(4), 121–148.

Online Appendix

A Assignment Economy Proofs

We now give three results that characterize a tax equilibrium and re-express it in a form suitable for optimal tax analysis. The first result confirms that assortative matching occurs in equilibrium.

Lemma A.1. *If (μ, z, w) is an equilibrium at (T, \bar{U}) , then either (a) $\mu = u$, no firm produces and all candidate CEOs take their outside option or (b) there is a $\tilde{v} \in I$ such that (i) for all $v \in (\tilde{v}, 1]$, $\mu(v) = u$ and (ii) for all $v \in (0, \tilde{v})$, $\mu(v) = v$.*

Proof of Lemma A.1. If $v' > v$, $\mu(v) = u$ and $\mu(v') \in I$, then $V(S(v), z(v')) - w(v') > V(S(v'), z(v')) - w(v') \geq 0$ and firm v is made strictly better off matching with CEO $\mu(v')$ at income $w(v')$ and effective labor supply $z(v')$. Since CEO $\mu(v')$ is obviously no worse off, this cannot be an equilibrium outcome and, hence, if $\mu(v') \in I$, then $\mu(v) \in I$. It follows that either (i) $\mu = u$ or (ii) there is some $\tilde{v} \in I$ such that for $v \in (0, \tilde{v})$, $\mu(v) \in I$ and for $v \in (\tilde{v}, 1]$, $\mu(v) = u$.

We next argue that μ is increasing on $(0, \tilde{v})$. Suppose not, then there exists a pair $v < v' \leq \tilde{v}$ such that $\mu(v') < \mu(v)$. We argue that by exchanging partners v and $\mu(v')$ can both improve their payoffs contradicting the fact that (μ, z, w) is an equilibrium. If (μ, z, w) is an equilibrium then, letting $c(v) := w(v) - T[w(v)]$ and $c(v') := w(v') - T[w(v')]$,

$$U\left(c(v), \frac{z(v)}{h(\mu(v))}\right) \geq U\left(c(v'), \frac{z(v')}{h(\mu(v))}\right)$$

and:

$$U\left(c(v'), \frac{z(v')}{h(\mu(v'))}\right) \geq U\left(c(v), \frac{z(v)}{h(\mu(v'))}\right).$$

If the first of these conditions did not hold, then CEO $\mu(v)$ could make herself and firm v' slightly better off by offering to supply $z(v')$ to firm v' for an income very slightly below $w(v')$. Similarly, if the second did not hold, then CEO $\mu(v')$ could make herself and firm v slightly better off by offering to supply $z(v)$ to firm v for an income very slightly below $w(v)$. The Spence-Mirrlees property of U then implies that $z(v') \geq z(v)$ and $c(v') \geq c(v)$ (and, hence, since no firm would pay a higher income to obtain a lower after-tax income and consumption for its CEO, $w(v') \geq w(v)$). Thus, $\mu(v')$ supplies more effective labor to v' than $\mu(v)$ supplies to v . If $\mu(v')$ instead works for v , supplies the same effective labor $z(v')$ and accepts the same income as before, then the payoff of firm v is changed by: $V(S(v), z(v')) - V(S(v), z(v)) - [w(v') - w(v)]$. If this change is positive, then a contradiction is obtained since firm v is made better off by the partner swap, while $\mu(v')$ is no worse off and so (μ, z, w) cannot be an equilibrium. If this is change is

non-positive and $z(v') > z(v)$, then:

$$\begin{aligned} 0 &\geq V(S(v), z(v')) - V(S(v), z(v)) - [w(v') - w(v)] \\ &> V(S(v'), z(v')) - V(S(v'), z(v)) - [w(v') - w(v)], \end{aligned}$$

where the second equality uses the strict super-modularity of V and so:

$$V(S(v'), z(v)) - w(v) > V(S(v'), z(v')) - w(v').$$

Thus, firm v' is made strictly better off by swapping partners with firm v , which again contradicts the requirement that (μ, z, w) is an equilibrium. Finally, consider the case in which $V(S(v), z(v')) - V(S(v), z(v)) - [w(v') - w(v)] = 0$, $z(v') = z(v)$ and $w(v) = w(v')$. If firms v and v' swap partners and continue to pay the same incomes and require the same effective labors from their CEOs, then no firm or CEO is made worse off. Denote the common effective labor amount by \hat{z} and the common income by \hat{w} and, to simplify the exposition suppose that the tax function T is differentiable at \hat{w} with derivative $T_w[\hat{w}]$. It cannot be that: $V_z(S(v), \hat{z})h(\mu(v')) = -\frac{U_e(\hat{w}-T[\hat{w}], \hat{z}/h(\mu(v')))}{U_c(\hat{w}-T[\hat{w}], \hat{z}/h(\mu(v')))}(1 - T_w[\hat{w}])$ and $V_z(S(v'), \hat{z})h(\mu(v)) = -\frac{U_e(\hat{w}-T[\hat{w}], \hat{z}/h(\mu(v)))}{U_c(\hat{w}-T[\hat{w}], \hat{z}/h(\mu(v)))}(1 - T_w[\hat{w}])$ since if so the following contradiction emerges:

$$\begin{aligned} V_z(S(v), \hat{z})h(\mu(v')) &= -\frac{U_e(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v')))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v')))}(1 - T_w[\hat{w}]) \\ &< -\frac{U_e(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v)))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v)))}(1 - T_w[\hat{w}]) \\ &= V_z(S(v'), \hat{z})h(\mu(v)) < V_z(S(v), \hat{z})h(\mu(v')), \end{aligned}$$

where the first inequality follows from the fact that $h(\mu(v')) > h(\mu(v))$ and the Spence-Mirrlees property of U and the second inequality follows from $S(v) > S(v')$, the strict super-modularity of V and $h(\mu(v')) > h(\mu(v))$. Thus, after rematchng at least one pair $(v, \mu(v'))$ or $(v', \mu(v))$ is not at a Pareto optimum. It is then possible for this pair to adjust CEO effort and income to make both firm and CEO better off. Again, this contradicts the assumption that (μ, z, w) is an equilibrium. We conclude that μ is increasing on $(0, \tilde{v}]$.

Finally, we show that for $\tilde{v} > 0$, μ is the identity map on $(0, \tilde{v})$. Since μ is measure-preserving and increasing, it is sufficient to show that there are no discontinuities in μ and that $\lim_{v \downarrow 0} \mu(v) = 0$. Suppose that μ has a discontinuity at some \hat{v} , but (without loss of generality) is continuous from the right. Then $\mu(\hat{v}) > \mu(\hat{v}_-) := \lim_{v \uparrow \hat{v}} \mu(v)$ and CEOs between $(\mu(\hat{v}_-), \mu(\hat{v}))$ are unmatched. But for $\tilde{m} \in (\mu(\hat{v}_-), \mu(\hat{v}))$, $U(w(\hat{v}) - T[w(\hat{v})], z(\hat{v})/h(\tilde{m})) > U(w(\hat{v}) - T[w(\hat{v})], z(\hat{v})/h(\mu(\hat{v}))) \geq \bar{U}$, contradicting the definition of equilibrium. Thus, μ is continuous. By a very similar argument if $\lim_{v \downarrow 0} \mu(v) > 0$, then there are unmatched CEOs in $(0, \lim_{v \downarrow 0} \mu(v)]$. These CEOs would be made strictly better off by matching with a firm and accepting the terms the firm is giving to her current CEO. Again this is inconsistent with an equilibrium.

Finally, we characterize μ at \tilde{v} . Suppose $\tilde{v} \in (0, 1)$ and let $\underline{v}_n \uparrow \tilde{v}$ and $\bar{v}_n \downarrow \tilde{v}$ (with

each $0 < \underline{v}_n < \tilde{v} < \bar{v}_n < 1$). We have:

$$W_n := U(w(\underline{v}_n) - T[w(\underline{v}_n)], z(\underline{v}_n)/h(\underline{v}_n)) \geq \bar{U} \geq U(w(\underline{v}_n) - T[w(\underline{v}_n)], z(\underline{v}_n)/h(\bar{v}_n)).$$

As observed previously $w_n = w(\underline{v}_n)$, $c_n = w(\underline{v}_n) - T[w(\underline{v}_n)]$ and $z_n = z(\underline{v}_n)$ are bounded, decreasing sequences. Denote the limits of these sequences $(w_\infty, c_\infty, z_\infty)$. Since $\lim h(\underline{v}_n) - h(\bar{v}_n) \downarrow 0$ and U is continuous, it follows that $U(c_n, z_n/h(\underline{v}_n)) - U(c_n, z_n/h(\bar{v}_n))$ converges to 0. Hence, $W_n \downarrow \bar{U}$. By a similar argument, $V(S(\underline{v}_n), z(\underline{v}_n)) - w(\underline{v}_n) \downarrow 0$. It follows that if T is continuous, then the \tilde{v} firm and CEO are indifferent about matching at the effective labor-income (z_∞, w_∞) . \square

Without loss of generality we select equilibria in which if $\tilde{v} > 0$, then $\mu(\tilde{v}) = \tilde{v}$. The next proposition simplifies the equilibrium conditions (5) to (9) in Definition 1 in a way that is convenient for tax analysis. It shows that given (T, \bar{U}) , if a pair of effective labor and income functions (z, w) on a domain $(0, \tilde{v}]$ are such that no CEO $v \in I \cap [0, \tilde{v}]$ is made strictly better off exchanging places with CEO $v'(0, \tilde{v}]$ and accepting the terms v' receives and similarly for firms, then no firm-CEO pair $(v, v') \in (0, \tilde{v}]^2$ can benefit (both weakly and at least one side strictly) from re-matching and selecting an arbitrary effective labor and income in the codomain of w . Furthermore, if the \tilde{v} -ranked CEO and firm receive the outside options \bar{U} and 0 respectively and if T is such that $T(w) = w$ at all w outside of the co-domain of w , then (\tilde{v}, w, z) is an equilibrium at (T, \bar{U}) . Thus, the stability conditions on firms and CEOs in Definition 1 are decoupled and re-expressed as separate firm and CEO incentive conditions. In addition, Proposition A.1 supplies a converse result: associated with any (non-trivial) equilibrium is a (\tilde{v}, z, w) satisfying the conditions described above.

Proposition A.1. *Let $T : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a tax function, \tilde{v} be a number in I and $z : (0, \tilde{v}] \rightarrow \mathbb{R}_+$ and $w : (0, \tilde{v}] \rightarrow \mathbb{R}_+$ be a pair of effective labor and income functions satisfying the **participation conditions**, for all $v \in (0, \tilde{v}]$,*

$$U\left(w(v) - T[w(v)], \frac{z(v)}{h(v)}\right) \geq \bar{U} \quad \text{and} \quad V(S(v), z(v)) - w(v) \geq 0, \quad (\text{A.1})$$

and the **incentive conditions**, for all $v, v' \in (0, \tilde{v}]$,

$$U\left(w(v) - T[w(v)], \frac{z(v)}{h(v)}\right) \geq U\left(w(v') - T[w(v')], \frac{z(v')}{h(v)}\right), \quad (\text{A.2})$$

and

$$V(S(v), z(v)) - w(v) \geq V(S(v), z(v')) - w(v'). \quad (\text{A.3})$$

Then there is no tuple (v, v', z', w') with $(v, v') \in (0, \tilde{v}]$ and $w' \in w((0, \tilde{v}])$ such that:

$$U\left(w' - T[w'], \frac{z'}{h(v')}\right) \geq U\left(w(v') - T[w(v')], \frac{z(v')}{h(v')}\right),$$

and

$$V(S(v), z') - w' \geq V(S(v), z(v)) - w(v),$$

with at least one of these inequalities strict. In addition, if (i) $U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) = \bar{U}$ and $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) \geq 0$ and (ii) for all $w' \notin w((0, \tilde{v}])$, $T(w') = w'$, then (w, z) defines an equilibrium at (T, \bar{U}) . Conversely, if (\tilde{v}, w, z) is an equilibrium at (T, \bar{U}) , then (\tilde{v}, w, z) satisfies (A.1) to (A.3), $U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) \geq \bar{U}$ and $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) \geq 0$ with equality if $\tilde{v} \in (0, 1)$.

Proof of Proposition A.1. Suppose the first claim in the proposition is false and that there is a tuple (v, v', z', w') with $v, v' \in (0, \tilde{v}]$ and $w' = w(\hat{v}) \in w((0, \tilde{v}])$ such that:

$$U\left(w' - T[w'], \frac{z'}{h(v')}\right) \geq U\left(w(v') - T[w(v')], \frac{z(v')}{h(v')}\right),$$

and

$$V(S(v), z') - w' \geq V(S(v), z(v)) - w(v),$$

with at least one of the previous inequalities strict. If $z' \geq z(\hat{v})$, then:

$$\begin{aligned} U\left(w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(v')}\right) &\geq U\left(w' - T[w'], \frac{z'}{h(v')}\right) \\ &\geq U\left(w(v') - T[w(v')], \frac{z(v')}{h(v')}\right) \geq U\left(w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(v')}\right), \end{aligned}$$

where the first inequality follows from $z' \geq z(\hat{v})$, the strict monotonicity of U in e and $w' = w(\hat{v})$, the second inequality is by assumption and the third follows from (A.2). If $z' > z(\hat{v})$, the first of the preceding inequalities is strict implying a contradiction. Thus, $z' \leq z(\hat{v})$ and if $z' = z(\hat{v})$, the v' -ranked CEO is no better off working for the v -ranked firm at (z', w') . If $z' < z(\hat{v})$, then

$$V(S(v), z(v)) - w(v) \geq V(S(v), z(\hat{v})) - w(\hat{v}) > V(S(v), z') - w(\hat{v}),$$

where the first inequality is by (A.3) and the second is from the strict monotonicity of V in z , and the v -ranked firm is worse off matching with the v' -ranked CEO at (z', w') . For this firm to be strictly better off, $z' > z(\hat{v})$. We conclude that v -ranked firm cannot be made strictly better off without making the v' -ranked CEO strictly worse off and vice versa. A contradiction is attained.

It follows that if (z, w) satisfies the conditions in the proposition over the domain $(0, \tilde{v}]$, then each firm (resp. CEO) $v \in (0, \tilde{v}]$ is better off matched with CEO (resp. firm) v , than matched with an alternative partner $v' \in (0, \tilde{v}]$ at an income $w' \in w((0, \tilde{v}])$ and an effective labor supply that improves their alternative partner's payoff relative to $(z(v'), w(v'))$. Moreover, if, $T(w') = w'$ for all $w' \in \mathbb{R}_+ \setminus w((0, \tilde{v}])$, then there is 100 per cent taxation of any price outside of the range of w on $(0, \tilde{v}]$. Clearly, no firm or CEO would wish to choose such an income and, hence, no firm or CEO in $(0, \tilde{v}]$ can benefit from rematching with another partner v' in this set and choosing any effective labor supply and income that improves their alternative partner's payoff relative to $(z(v'), w(v'))$. In addition, if $U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) = \bar{U}$ and $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) = 0$, then no firm or CEO $v \in (0, \tilde{v}]$ is better off unmatched

than matched at $(w(v), z(v))$, i.e.

$$U\left(w(v) - T[w(v)], \frac{z(v)}{h(v)}\right) \geq U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) \geq U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) \geq \bar{U},$$

and similarly for firms. Finally, if $\tilde{v} \in (0, 1]$ and $U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) = \bar{U}$ and $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) = 0$, then by similar logic to that given above, it is readily verified that all firms and CEOs $v \in (\tilde{v}, 1]$ are better off unmatched than matched with a partner in I at a (z', w') that gives the partner as much as it could obtain from remaining its current match or remaining unmatched.

For the converse, if (\tilde{v}, z, w) is an equilibrium at (T, \bar{U}) , then it is immediate that it satisfies (A.1) to (A.3) and $U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) \geq \bar{U}$ and $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) \geq 0$. If $\tilde{v} \in (0, 1)$, then $U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) = \bar{U}$ and $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) = 0$ since, otherwise, there is an interval (\tilde{v}, v') such that for each $v \in (\tilde{v}, v')$ either $U\left(w(\tilde{v}) - T[w(\tilde{v})], \frac{z(\tilde{v})}{h(\tilde{v})}\right) > \bar{U}$ or $V(S(v), z(\tilde{v})) - w(\tilde{v}) > 0$. This contradicts the equilibrium definition. \square

Proof of Proposition 1 The proof of Proposition 1 is a now a direct consequence of Lemma A.1 and Proposition A.1. \square

Lemma A.2 provides monotonicity results for various equilibrium functions. It also proves the existence of a function ω relating equilibrium income to effective labor supply.

Lemma A.2. *If (\tilde{v}, z, w) is an equilibrium threshold and a pair of equilibrium effective labor and income functions at (T, \bar{U}) , then (z, w, c) , with $c(v) = w(v) - T[w(v)]$, $v \in (0, \tilde{v}]$ are non-increasing and there exists a function $\omega : z((0, \tilde{v}]) \rightarrow \mathbb{R}_+$ satisfying for each $v \in (0, \tilde{v}]$, $\omega(z(v)) = w(v)$. In addition, equilibrium CEO utility Φ , $\Phi(v) = U(c(v), z(v)/h(v))$ for $v \in (0, \tilde{v}]$, and firm profits π , $\pi(v) = V(S(v), z(v)) - w(v)$ for $v \in (0, \tilde{v}]$, are decreasing.*

Proof of Lemma A.2. Monotonicity of z and c follow from (A.2), the Spence-Mirrlees property of U and standard arguments. Hence, if $v > v'$, then $c(v) \leq c(v')$ and so if $w(v) > w(v')$, then $T[w(v)] - T[w(v')] > w(v) - w(v')$. But clearly no firm would choose to buy from a CEO at income $w(v)$ (they could strictly reduce the income they pay and weakly raise their CEO's consumption). Hence, w must be non-decreasing as well. Moreover, w is $\sigma(z)$ -measurable (where $\sigma(z)$ denotes the sigma-algebra induced by z). Hence, there exists a function ω , with $\omega(z(v)) = w(v)$ (see, for example, Klenke (2008), Corollary 1.97, p. 41). Finally, if $v < v'$, then $\Phi(v) = U(c(v), z(v)/h(v)) \geq U(c(v'), z(v)/h(v)) > U(c(v'), z(v')/h(v')) = \Phi(v')$ and similarly for π . \square

B Derivation of Elasticities for Section 3

In this appendix, we derive expressions for elasticities used in Section 3 under the assumption of a multiplicative firm surplus function DSz . We first assume general preferences and then specialize to the quasilinear/constant effort elasticity case. The elasticities given in the main text are “aggregate elasticities” that summarize the CEO income and firm profit responses of populations of CEOs and firms to a tax rate change. Below we build these elasticities up from individual level responses and equilibrium conditions.

Individual income elasticities for CEOs As in the main text let $\omega(z, 1 - \tau)$ give CEO income as a function of effective labor supply and the tax rate. In this appendix, we also make the dependence of a CEO’s effective labor on the retention rate explicit in the notation and let $z(v, 1 - \tau)$ denote the equilibrium effective labor supply of CEO v given retention rate $1 - \tau$. Suppressing dependence of functions on their arguments in the notation, the CEOs first order conditions are:

$$(1 - \tau) \frac{\partial \omega}{\partial z} h U_c \left(c, \frac{z}{h} \right) + U_e \left(c, \frac{z}{h} \right) = 0. \quad (\text{B.1})$$

where: $c(v, 1 - \tau) = \omega(z(v, 1 - \tau), 1 - \tau) - \tau(\omega(z(v, 1 - \tau), 1 - \tau) - w_0) - T[w_0]$. The corresponding firm’s first order condition is:

$$\frac{\partial \omega}{\partial z} = DS. \quad (\text{B.2})$$

Since $z(v, 1 - \tau) = h(v)e(v, 1 - \tau)$, we have $z_v = z \left\{ \frac{h_v}{h} + \frac{e_v}{e} \right\}$. Substituting (B.2) into (B.1), totally differentiating and using the definitions of the uncompensated and compensated CEO effort elasticities \mathcal{E}^u and \mathcal{E}^c gives:

$$\frac{e_v}{e} = \mathcal{E}^u \frac{h_v}{h} + \mathcal{E}^c \frac{e_v}{e}.$$

And so:

$$\frac{z_v}{z} = \left\{ (1 + \mathcal{E}^u) \frac{h_v}{h} + \mathcal{E}^c \frac{S_v}{S} \right\}. \quad (\text{B.3})$$

Differentiating (B.2) with respect to v and combining with (B.3) gives:

$$\frac{\partial^2 \omega}{\partial z^2} = \frac{DS_v}{z \left\{ (1 + \mathcal{E}^u) \frac{h_v}{h} + \mathcal{E}^c \frac{S_v}{S} \right\}}.$$

Totally differentiating the CEO's first order condition (B.1) with respect to $1 - \tau$ gives:

$$\frac{\partial z}{\partial(1-\tau)} = - \frac{\frac{\partial \omega}{\partial z} h U_c + \left\{ w - w_0 + (1-\tau) \frac{\partial \omega}{\partial(1-\tau)} \right\} \left\{ (1-\tau) \frac{\partial \omega}{\partial z} h U_{cc} + U_{ec} \right\} + (1-\tau) \frac{\partial^2 \omega}{\partial z \partial(1-\tau)} h U_c}{(1-\tau) \frac{\partial^2 \omega}{\partial z^2} h U_c + \left\{ (1-\tau) \frac{\partial \omega}{\partial z} \right\}^2 h U_{cc} + 2(1-\tau) \frac{\partial \omega}{\partial z} U_{ce} + U_{ee}/h}.$$

Totally differentiating the firm's first order condition (B.2) gives:

$$\frac{\partial^2 \omega}{\partial z^2} \frac{\partial z}{\partial(1-\tau)} + \frac{\partial^2 \omega}{\partial z \partial(1-\tau)} = 0.$$

Combining the preceding conditions:

$$\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)} = - \left(\frac{1-\tau}{z} \right) \frac{DS h U_c + \left[\frac{w-w_0}{w} + \frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)} \right] w \{ (1-\tau) DS h U_{cc} + U_{ec} \}}{\{ (1-\tau) DS \}^2 h U_{cc} + 2(1-\tau) DS U_{ce} + U_{ee}/h}. \quad (\text{B.4})$$

Note if CEO utility is quasilinear in consumption and there are no income effects, then $U_{cc} = U_{ce} = 0$ and the elasticity in (B.4) reduces to the usual behavioral elasticity $\mathcal{E} = \frac{U_e}{U_{ce}e}$. If income effects on effort supply are negative and $\frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)} \leq 0$, then

$$\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)} \geq \left(\frac{U_e}{U_{ce}e} \right) \frac{U_c + \frac{w}{DShe} \left\{ -\frac{U_e}{U_c} U_{cc} + U_{ec} \right\} e}{\left\{ \frac{U_e}{U_c} \right\}^2 U_{cc} - 2 \frac{U_e}{U_c} U_{ce} + U_{ee}} > \left(\frac{U_e}{U_{ce}e} \right) \frac{U_c + \left\{ -\frac{U_e}{U_c} U_{cc} + U_{ec} \right\} e}{\left\{ \frac{U_e}{U_c} \right\}^2 U_{cc} - 2 \frac{U_e}{U_c} U_{ce} + U_{ee}},$$

where the final right hand side is the usual behavioral uncompensated effort elasticity. In particular, in this case, if the latter elasticity is non-negative, then so too is $\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)}$. In addition, at v_0 , $w = w_0$ and if $\frac{\partial \omega(z(v_0), 1-\tau)}{\partial(1-\tau)} = 0$ (as will be shown below), then $\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)}$ equals the usual compensated effort elasticity.

We now verify that at v_0 neither the firm's nor the CEO's payoff changes in response to a small retention rate change. Let Φ_0 denote the utility of the v_0 -ranked CEO prior to the tax change. Let z_0 denote this CEO's effective labor, w_0 her income, h_0 her talent, π_0 firm v_0 's profit and S_0 its asset size (all prior to the tax change). Then:

$$U \left(w_0 - T[w_0], \frac{z_0}{h_0} \right) = \Phi_0. \quad (\text{B.5})$$

A small change to $1 - \tau$ perturbs CEO v_0 's utility by: $U_c(1-\tau) \frac{\partial \omega}{\partial(1-\tau)}(z_0, 1-\tau)$. If this CEO continues to receive Φ_0 after the change, then $\frac{\partial \omega}{\partial(1-\tau)}(z_0, 1-\tau) = 0$. Furthermore,

$$\frac{\partial \pi}{\partial(1-\tau)}(v_0, 1-\tau) = \left\{ DS_0 - \frac{\partial \omega}{\partial z}(z_0, 1-\tau) \right\} \frac{\partial z}{\partial(1-\tau)}(v_0, 1-\tau) - \frac{\partial \omega}{\partial(1-\tau)}(z_0, 1-\tau) = 0.$$

Consider three cases. First $v_0 = \bar{v} = 1$ and all firms are matched prior to the tax change with $\Phi_0 = \bar{U}$. The v_0 -firm cannot reduce the utility of its CEO (which must

remain above \bar{U}), has no desire to raise the utility of its CEO and by our equilibrium assumption no need to. Thus, the CEO's payoff remains at \bar{U} and the firm's profit remains at π_0 . Second, $v_0 = \bar{v} < 1$. In this case not all firms are matched prior to the tax change, but $\Phi_0 = \bar{U}$. Again, the v_0 -firm cannot reduce the utility of its CEO and has no desire to raise this utility. If it continues to offer a utility of \bar{U} to its CEO, then its payoff is unchanged (and equal to 0). Firms $v \in (v_0, 1)$ (continue to) make strictly smaller and, hence, negative profits if they enter and match with the v_0 ranked CEO or with their correspondingly ranked (candidate) CEO. Hence, these firms do not enter the assignment market and firm v_0 does not need to offer CEO v_0 a utility above \bar{U} . Similar logic ensures that in the third case $v_0 < \bar{v} \leq 1$, firm v_0 continues to offer the CEO the same utility as was offered prior to the small retention rate variation and continues to receive the same payoff.

Since $\frac{\partial \omega}{\partial(1-\tau)}(z_0, 1-\tau) = 0$ there is no adjustment to the equilibrium CEO income schedule at w_0 . Hence, at $v = v_0$ using (B.2) and (B.4), the elasticity of the CEO's income with respect to taxes is:

$$\begin{aligned} \mathcal{E}_w(v_0) &:= \frac{1-\tau}{w} \frac{dw}{d(1-\tau)} \Big|_{v_0} = \frac{1-\tau}{w_0} \frac{\partial \omega}{\partial z} \frac{\partial z}{d(1-\tau)} + \frac{1-\tau}{w_0} \frac{\partial \omega}{\partial(1-\tau)} \Big|_{v_0} \\ &= \left(\frac{z}{w_0} \frac{\partial \omega}{\partial z} \right) \left(\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)} \right) \Big|_{v_0} = \frac{DS_0 z_0}{w_0} \mathcal{E}_0^c, \end{aligned}$$

where \mathcal{E}_0^c is the compensated effort elasticity of CEO v_0 in equilibrium. Define $v(z', 1-\tau)$, $z' \in z([0, 1], 1-\tau)$, to be the rank of the CEO exerting effective labor z' when the retention rate is $1-\tau$, i.e.

$$v(z(v, 1-\tau), 1-\tau) = v. \quad (\text{B.6})$$

At all points of differentiability, (B.6) implies:

$$\frac{\partial v}{\partial z}(z, 1-\tau) \frac{\partial z}{\partial(1-\tau)}(v(z, 1-\tau), 1-\tau) + \frac{\partial v}{\partial(1-\tau)}(z, 1-\tau) = 0. \quad (\text{B.7})$$

It follows from (B.7) that:

$$\frac{\partial v}{\partial(1-\tau)}(z, 1-\tau) > 0 \quad (\text{B.8})$$

if the corresponding effective labor elasticity $\frac{1-\tau}{z} \frac{dz}{d(1-\tau)}$ (at $(v(z, 1-\tau), 1-\tau)$) is positive (and if there is no bunching at $v(z, 1-\tau)$ so that $z_v(v(z, 1-\tau)) > 0$). Using (B.2), we have:

$$\omega(z, 1-\tau) = \omega(z_0, 1-\tau) + \int_{z_0}^z DS(v(z', 1-\tau)) dz'. \quad (\text{B.9})$$

Differentiating (B.9) with respect to $1-\tau$ (and using $\frac{\partial \omega}{\partial(1-\tau)}(v_0) = 0$) gives:

$$\frac{\partial \omega}{\partial(1-\tau)}(z, 1-\tau) = \int_{z_0}^z \left\{ DS_v(v(z', 1-\tau)) \frac{\partial v}{\partial(1-\tau)}(z', 1-\tau) \right\} dz'. \quad (\text{B.10})$$

This equation combined with (B.7) and the elasticity formula (B.4) gives an explicit (but complicated) formula for $\frac{\partial\omega}{\partial(1-\tau)}$. It follows from (B.10) and our earlier discussion of $\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)}$ that if the uncompensated behavioral elasticity for effort is positive over an interval of effective labors and there is no bunching, then $\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)}$ and $\frac{\partial v}{\partial(1-\tau)}$ are both positive over the interval and $\frac{\partial\omega}{\partial(1-\tau)}(z, 1-\tau)$ is negative. The logic is straightforward. A rise in $1-\tau$ induces all CEO's to work harder. Consequently, increments in effective labor between z_0 and z are now associated with lower ranked (i.e. higher v) firms and CEOs. Lower ranked firms pay less for the last unit of effective labor they hire. (Specifically, firm v 's optimality condition may be expressed as: $DS(v) = \frac{\partial\omega}{\partial z}(z(v, 1-\tau), 1-\tau)$. Thus, firm v pays $DS(v)$ (only) for the last unit of effective labor it buys. It must do so to secure this marginal unit in the face of competition from slightly less productive firms.) Hence, the additional income paid for each given incremental unit of effective labor hired (by now lower ranked firms) is reduced after a rise in $1-\tau$ and the income paid to a CEO supplying a given z is reduced.

This does not mean that the income earned by CEO v falls after a rise in $1-\tau$ since that CEO will supply more effective labor in equilibrium. Specifically, the overall income elasticity of CEO v with respect to the retention rate $1-\tau$ is:

$$\begin{aligned} \mathcal{E}_w(v) &= \left(\frac{z}{w} \frac{\partial\omega}{\partial z} \right) \left(\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)} \right) + \left(\frac{1-\tau}{w} \frac{\partial\omega}{\partial(1-\tau)} \right) \\ &= \left(\frac{DS(v)z(v)}{w} \right) \left(\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)} \right) + \left(\frac{1-\tau}{w} \frac{\partial\omega}{\partial(1-\tau)} \right), \end{aligned} \quad (\text{B.11})$$

where $\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)}$ is defined as in (B.4) and the first term is positive if $\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)}$ is and, using (B.10),

$$\frac{1-\tau}{w} \frac{\partial\omega}{\partial(1-\tau)}(z, 1-\tau) = \frac{1-\tau}{w} \int_{z_0}^z \left\{ DS_v(v(z', 1-\tau)) \frac{\partial v}{\partial(1-\tau)}(z', 1-\tau) \right\} dz'. \quad (\text{B.12})$$

The preceding formulas are simplified in the absence of income effects. Then $\frac{1-\tau}{z} \frac{\partial z}{\partial(1-\tau)} = \mathcal{E}^c = \mathcal{E}^u = \mathcal{E} := \frac{U_e}{U_{ee}}$ and, after a change of variables and an integration by parts,

$$-\frac{1-\tau}{w} \frac{\partial\omega}{\partial(1-\tau)}(z(v)) = \frac{DSz\mathcal{E}(v) - DS_0z_0\mathcal{E}(v_0)}{w} + \frac{1}{w} \int_v^{v_0} DS(v')z(v')\mathcal{E}(v')\mathcal{K}(v')dv', \quad (\text{B.13})$$

with:

$$\mathcal{K}(v') := \left(1 + \mathcal{E}(v') + \frac{1-\tau}{\mathcal{E}} \frac{\partial\mathcal{E}}{\partial(1-\tau)}(v') \right) \frac{h_v}{h}(v') + \left(\mathcal{E}(v') + \frac{1-\tau}{\mathcal{E}} \frac{\partial\mathcal{E}}{\partial(1-\tau)}(v') \right) \frac{S_v}{S}(v')$$

In the constant elasticity case, (B.13) is further reduced to:

$$\frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)} = - \left(\frac{\pi - \pi_0}{w} \right) \mathcal{E} < 0.$$

Absent income effects, the total elasticity of CEO income with respect to the retention rate $1 - \tau$ is:

$$\mathcal{E}_w(v) = \frac{DSz}{w} \mathcal{E}(v) + \frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)}(v),$$

with $\frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)}$ defined as in (B.13). In the constant effort elasticity case, this is further reduced to:

$$\mathcal{E}_w(v) = \left\{ \frac{DS_0 z_0 - w_0}{w(v)} + 1 \right\} \mathcal{E} \geq \mathcal{E} \geq 0. \quad (\text{B.14})$$

Note that in this case as $v \downarrow 0$, $\mathcal{E}_w(v)$ converges to \mathcal{E} .

Aggregate income elasticities for CEOs Defining aggregate CEO income above w_0 to be $W := \int_0^{v_0} w(v)dv$, the elasticity of W with respect to $1 - \tau$ is thus:

$$\mathcal{E}_W = \frac{1}{W} \int_0^{v_0} w(v) \mathcal{E}_w(v) dv,$$

with $\mathcal{E}_w(v)$ defined as in (B.11). In the quasi-linear/constant effort elasticity case, from (B.14), this is simply:

$$\mathcal{E}_W = \mathcal{E} \int_0^{v_0} \frac{w(v)}{W} \left\{ \frac{DS_0 z_0 - w_0}{w(v)} + 1 \right\} dv = \left(\frac{R}{W} - \frac{\Delta \Pi}{W} \right) \mathcal{E} \geq 0, \quad (\text{B.15})$$

where $R := \int_0^{v_0} DS(v)z(v, 1 - \tau)dv$ and $\Delta \Pi := \int_0^{v_0} \{\pi(v) - \pi_0\}dv$.

Individual and aggregate firm profit elasticities Next consider the elasticity of firm profit $\pi = Sz - w$ with respect to $1 - \tau$. Using the firm's first order condition (B.2), this is:

$$\frac{1-\tau}{\pi} \frac{d\pi}{d(1-\tau)} = - \frac{w}{\pi} \frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)}, \quad (\text{B.16})$$

with $\frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)}$ defined as in (B.10) or (B.13). Note that since $\frac{1-\tau}{w} \frac{\partial \omega}{\partial(1-\tau)} < 0$ if the uncompensated behavioral elasticity is positive, it flows that $\frac{1-\tau}{\pi} \frac{d\pi}{d(1-\tau)} > 0$. In the quasilinear/constant elasticity case, (B.16) reduces to:

$$\frac{1-\tau}{\pi} \frac{d\pi}{d(1-\tau)} = \left(\frac{\pi - \pi_0}{\pi} \right) \mathcal{E} \geq 0,$$

where the inequality uses the monotonicity of π in v . Defining total firm profit above $\pi_0 = \pi(v_0)$ to be $\Pi := \int_0^{v_0} \pi(v)dv$, the elasticity of Π with respect to $1 - \tau$ is:

$$\mathcal{E}_\Pi = \int_0^{v_0} \frac{\pi}{\Pi} \frac{1 - \tau}{\pi} \frac{d\pi}{d(1 - \tau)} dv.$$

Again this is positive if the underlying uncompensated behavioral effort elasticity. In the quasilinear/constant elasticity case, this formula reduces to:

$$\mathcal{E}_\Pi = \frac{\Delta\Pi}{\Pi} \mathcal{E} \geq 0.$$

Optimal tax formulas The preceding ‘‘aggregate’’ elasticities can be substituted into the tax formula (30) given in the main text to give optimal taxes in terms of individual elasticities, the equilibrium CEO income schedule and the various distributions. In the remainder of this appendix we extend (30) to allow for societal concern for top CEOs. We then consider the case $\chi = 1$ in detail and derive (34) in the main text.

Societal concern for top CEOs Let ψ denote the welfare weight on top CEOs. Then the policymaker’s first order condition is modified as:

$$\begin{aligned} & - \int_0^{v_0} \{ \omega(z^*(v); 1 - \tau^*) - w_0 \} dv \\ & + \int_0^{v_0} \psi(v) U_c(v) \left\{ \omega(z^*(v); 1 - \tau^*) - w_0 + (1 - \tau^*) \frac{\partial \omega}{\partial (1 - \tau)}(z^*(v); 1 - \tau^*) \right\} dv \\ & + \tau^* \int_0^{v_0} \left\{ \frac{\partial \omega}{\partial z}(z^*(v); 1 - \tau^*) \frac{\partial z}{\partial (1 - \tau)}(v, 1 - \tau^*) + \frac{\partial \omega}{\partial (1 - \tau)}(z^*(v); 1 - \tau^*) \right\} dv \\ & - \chi \int_0^{v_0} \frac{\partial \omega}{\partial (1 - \tau)}(z^*(v); 1 - \tau^*) dv = 0, \end{aligned} \tag{B.17}$$

where $z^*(v) := z(v, 1 - \tau^*)$. Specializing to the quasilinear/constant elasticity case, rearranging and expressing in terms of aggregate elasticities and attributes of the CEO income and firm profit distributions evaluated at the optimum (and denoted by stars) gives:

$$\tau^* = \frac{1}{1 + A_W^* \frac{\mathcal{E}_W^* + \chi \frac{\Pi^*}{W^*} \mathcal{E}_\Pi^*}{1 - \phi - (\chi - \phi) A_W^* \frac{\Pi^*}{W^*} \mathcal{E}_\Pi^*}}. \tag{B.18}$$

The $\chi = 1$ case In the main body of the paper we show that the optimal linear tax rate when $\chi = 1$ is:

$$\tau^* = \frac{1}{1 + A_W^* \frac{\mathcal{E}_W^* + \frac{\Pi^*}{W^*} \mathcal{E}_\Pi^*}{1 - A_W^* \frac{\Pi^*}{W^*} \mathcal{E}_\Pi^*}}, \tag{B.19}$$

where the aggregate elasticities in (B.19) are evaluated at the optimum. We now show that under the assumptions of quasi-linear/constant elasticity preferences and a constant talent Pareto coefficient (B.19) reduces to (34) in the main text. After simple calculations:

$$\frac{\Pi^*}{W^*} \mathcal{E}_{\Pi}^* = \frac{R^*}{W^*} \mathcal{E}_R^* - \mathcal{E}_W^*, \quad (\text{B.20})$$

where $R^* = \int_0^{v_0} DS(v)z(v, 1 - \tau^*)dv$ and:

$$\mathcal{E}_R^* := \frac{1 - \tau}{R} \frac{dR}{d(1 - \tau)} \Big|_{1 - \tau^*} = \int_0^{v_0} \frac{r^*(v)}{R^*} \frac{1 - \tau^*}{z(v, 1 - \tau^*)} \frac{dz(v, 1 - \tau^*)}{d(1 - \tau)} dv, \quad (\text{B.21})$$

with $r^*(v) = DS(v)z(v, 1 - \tau^*)$. It then follows from (B.19) and (B.20) that:

$$\tau^* = \frac{1}{1 + \frac{\mathcal{E}_R^*}{\frac{W^*}{R^*} \frac{1}{A_W^*} - \frac{\Pi^*}{R^*} \mathcal{E}_{\Pi}^*}}. \quad (\text{B.22})$$

From the definition of A_W^* ,

$$\frac{W^*}{A_W^*} = \Delta W^* = \int_0^{v_0} \{w^*(v) - w_0\} dv.$$

Next using the definition of the local Pareto coefficient $\alpha_w^*(v) := \frac{m(w^*(v))w^*(v)}{1 - M(w^*(v))}$ and the fact that $v = 1 - M(w^*(v))$,

$$\int_0^{v_0} \frac{w^*(v)}{\alpha_w^*(v)} dv = \int_0^{v_0} \frac{1 - M(w^*(v))}{m(w^*(v))} dv = \int_0^{v_0} w_v^*(v) v dv = \int_0^{v_0} \{w^*(v) - w_0\} dv,$$

where the last equality is via an integration by parts. Hence,

$$\frac{W^*}{A_W^*} = \int_0^{v_0} \frac{w^*(v)}{\alpha_w^*(v)} dv. \quad (\text{B.23})$$

Denoting the local Pareto coefficients for CEO talent and firm profit (at the optimum) by α_h and α_{π}^* respectively and using (21) in the main body of the paper gives:

$$\frac{w}{\alpha_w^*} = DSz(1 + \mathcal{E}^u) \frac{1}{\alpha_h} + \pi \mathcal{E}^c \frac{1}{\alpha_{\pi}^*}. \quad (\text{B.24})$$

So, combining (B.23) and (B.24) and evaluating at the optimum:

$$\frac{\Delta W^*}{R^*} = \frac{1}{R^*} \frac{W^*}{A_W^*} = \int_0^{v_0} \frac{r^*}{R^*} (1 + \mathcal{E}^{u*}) \frac{1}{\alpha_h} dv + \int_0^{v_0} \frac{\pi^*}{R^*} \mathcal{E}^{c*} \frac{1}{\alpha_{\pi}^*} dv.$$

Consequently,

$$\begin{aligned} \frac{1}{R^*} \frac{W^*}{A_W^*} - \frac{\Pi^*}{R^*} \mathcal{E}_{\Pi}^* &= \int_0^{v_0} \frac{r^*}{R^*} (1 + \mathcal{E}^{u*}) \frac{1}{\alpha_h} dv + \frac{1}{R^*} \int_0^{v_0} \pi^* \mathcal{E}^{c*} \frac{1}{\alpha_{\pi}^*} dv \\ &+ \frac{1 - \tau^*}{R^*} \int_0^{v_0} \int_{z_0}^{z^*(v)} \left\{ DS_v(v(z', 1 - \tau)) \frac{\partial v}{\partial(1 - \tau)}(z', 1 - \tau^*) \right\} dz' dv. \end{aligned} \quad (\text{B.25})$$

Together (B.21), (B.22) and (B.25) imply:

$$\tau^* = \frac{1}{1 + \frac{\int_0^{v_0} \frac{r^*}{R^*} \frac{1 - \tau}{z} \frac{dz}{d(1 - \tau)} dv}{\int_0^{v_0} \frac{r^*}{R^*} (1 + \mathcal{E}^{u*}) \frac{1}{\alpha_h} dv + \frac{1}{R^*} \int_0^{v_0} \mathcal{E}^{c*} \frac{\pi^*}{\alpha_{\pi}^*} dv + \frac{1 - \tau^*}{R^*} \int_0^{v_0} \int_{z_0}^{z^*(v)} \left\{ DS_v(v(z', 1 - \tau)) \frac{\partial v}{\partial(1 - \tau)}(z', 1 - \tau^*) \right\} dz' dv}}.$$

This formula is greatly simplified in the quasilinear/constant effort elasticity case. Then $\frac{1 - \tau}{z} \frac{dz}{d(1 - \tau)}$, \mathcal{E}^c and \mathcal{E}^u take the common value \mathcal{E} which can be pulled through integrals. In this case, using (B.13) gives:

$$\begin{aligned} \frac{1}{R^*} \frac{W^*}{A_W^*} - \frac{\Pi^*}{R^*} \mathcal{E}_{\Pi}^* &= (1 + \mathcal{E}) \int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv + \frac{\mathcal{E}}{R^*} \int_0^{v_0} \frac{\pi^*}{\alpha_{\pi}^*} dv \\ &- \frac{\mathcal{E}}{R^*} \int_0^{v_0} \left\{ DS(v) z^*(v) - DS_0 z_0 + \int_v^{v_0} w_v^*(v') dv' \right\} dv \\ &= (1 + \mathcal{E}) \int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv + \frac{\mathcal{E}}{R^*} \int_0^{v_0} \frac{\pi^*}{\alpha_{\pi}^*} dv \\ &- \frac{\mathcal{E}}{R^*} \int_0^{v_0} \left\{ DS(v) z^*(v) - w^*(v) - \{DS_0 z_0 - w_0\} \right\} dv. \end{aligned}$$

Then using $\pi^*(v) = DS(v) z^*(v) - w^*(v)$, $\pi_0 = DS_0 z_0 - w_0$ and $\int_0^{v_0} \frac{\pi^*}{\alpha_{\pi}^*} dv = \int_0^{v_0} \{ \pi^*(v) - \pi_0 \} dv$, the previous equation reduces to:

$$\frac{1}{R^*} \frac{W^*}{A_W^*} - \frac{\Pi^*}{R^*} \mathcal{E}_{\Pi}^* = (1 + \mathcal{E}) \int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv.$$

The right hand side of this expression is a weighted integral of (reciprocals of) talent local Pareto coefficients. Furthermore, $\mathcal{E}_R^* = \mathcal{E}$. Hence,

$$\tau^* = \frac{1}{1 + \frac{\mathcal{E}}{1 + \mathcal{E}} \int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv}}.$$

If the local talent Pareto coefficient is constant, then the previous formula reduces to:

$$\tau^* = \frac{1}{1 + \frac{\mathcal{E}}{1 + \mathcal{E}} \frac{1}{\alpha_h}}.$$

Independence of $\mathcal{E}_W A_W$ from the marginal tax rate We conclude this appendix by showing that the product $\mathcal{E}_W A_W$ is independent of the marginal tax rate under

the assumptions of quasilinear/constant elasticity CEO preferences, a multiplicative firm objective and a tax system that is linear above a threshold. This assumption allows us to relate estimates of the product $\mathcal{E}_W A_W$ in US data to the optimal value and, hence, the optimal tax rate in the Diamond-Saez formula. Note that from (B.15) under the assumptions just described:

$$A_W \mathcal{E}_W = \frac{W}{\Delta W} \frac{\mathcal{E}}{W} \int_0^{v_0} \{DS_0 z_0 - w_0 + w(v)\} dv = \mathcal{E} \left\{ 1 + \frac{DS_0 z_0}{\Delta W} (1 - v_0) \right\}$$

and

$$\Delta W = \int_0^{v_0} \{w(v) - w_0\} dv = \int_0^{v_0} w_v(v) v dv = \int_0^{v_0} \mathcal{H}(v) z(v) dv,$$

with $\mathcal{H}(v)$ independent of marginal taxes. Hence,

$$\frac{\partial \Delta W}{\partial (1 - \tau)} = \frac{\mathcal{E}}{1 - \tau} \Delta W.$$

Similarly,

$$\frac{\partial DS_0 z_0}{\partial (1 - \tau)} = \frac{\mathcal{E}}{1 - \tau} DS_0 z_0.$$

Thus, the ratio $\frac{DS_0 z_0}{\Delta W}$ is unaffected by the marginal tax rate and neither is the product $A_W \mathcal{E}_W$.

C Elasticity of Firm Profits to Top Income Tax Rate

In this section we take a closer look at the empirical evidence on the elasticity of firm profits to the top income tax rate, i.e. on the magnitude of \mathcal{E}_Π . The literature gives little guidance on the size of this elasticity. However, Figure 3 provides some suggestive evidence. It displays the time series for net corporate dividends of US domestic industries and the top marginal tax rate on income and corporate profits over the period 1919 to 2014.⁵² We begin looking at the relationship between dividends and top marginal income tax rate. A negative correlation between these variables is apparent. Along the lines of [Piketty, Saez, and Stantcheva \(2014\)](#) (who focus on the relationship between top incomes and the retention rate), we estimate the following log linear relationship:

$$\log(\text{Dividends}/\text{GDP}_t) = \beta_0 + \beta_1 \log(1 - \bar{\tau}_t) + \varepsilon_t, \quad (\text{C.1})$$

⁵²Top marginal rates are taken from the *Tax Foundation*. Data on dividends from 1940 to 2014 is from the BEA Table 6.20A (series: A3302C0) data on GDP is from Table 1.1.5. Data on dividends from 1919 to 1939 is from the NBER Macrohistory Database (series: a08185). The NBER and BEA dividend series overlap between 1929 to 1939, differences are small but systematic. We use the average difference observed between 1929 to 1939 to adjust the NBER data.

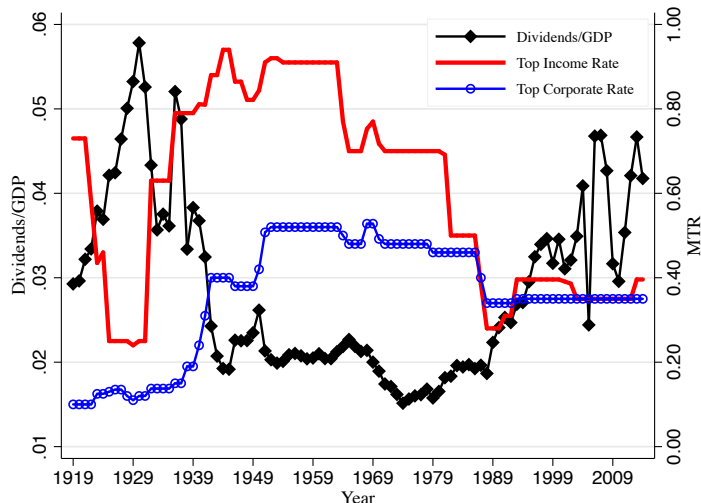


Figure 3: The relationship between net corporate dividend payments (left axis) and marginal top income and corporate tax rates (right axis).

where $\bar{\tau}_t$ is the top marginal tax rate on income at time t . Estimates for β_1 from 1919 to 2014 provide a positive and statistically significant elasticity of 0.232 (0.042). Of course, this does not establish a casual relationship between the time series as an omitted third factor might be responsible for both time profiles. A candidate for this third factor is the top marginal rate on corporate profits, also displayed in Figure 3. The two top marginal tax rates (after 1940) display very similar profiles. We complement the relationship (C.1) with the factor $\beta_2 \log(1 - \bar{\tau}_t^c)$ where $\bar{\tau}_t^c$ is the top marginal corporate tax rate at time t . The estimate of β_1 in this case is equal to 0.0574 (0.0333). In this case we can reject a value of $\beta_1 = 0$ at the 10 per cent level. Overall the estimate for the elasticity is close to the value implied for \mathcal{E}_Π by the structural model in the body of the paper. A formal connection between the two, however, cannot be fully established as in the assignment model profit is defined to be the rents accruing to owners of the asset S . These rents exclude payments to adjustable capital and may not be realized contemporaneously with the application of CEO effective labor.

D Derivation of optimal nonlinear tax formulas

When firm claimants have welfare weight χ , CEOs have welfare weights ψ , the policymaker's (relaxed) mechanism design problem reduces to:

$$\begin{aligned} \sup_{\Phi, z, w, \bar{v}} \int_0^{\bar{v}} \{ \chi V(S(v), z(v)) + (1 - \chi)w(v) - C[\Phi(v), z(v)/h(v)] + \psi(v)\Phi(v) \} dv \\ + \int_{\bar{v}}^1 \{ \psi(v)\bar{U} + T^0 \} dv, \end{aligned} \quad (\text{D.1})$$

subject to $\tilde{v} \in I = (0, 1]$, $\Phi(\tilde{v}) = \bar{U}$, $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) \geq 0$, and for almost all $v \in I$,

$$\begin{aligned}\Phi_v(v) &= -U_e \left(\mathcal{C} \left[\Phi(v), \frac{z(v)}{h(v)} \right], \frac{z(v)}{h(v)} \right) \frac{z(v)}{h(v)} \frac{h_v(v)}{h(v)} \\ w_v(v) &= V_z(S, z) z_v(v).\end{aligned}$$

In the relaxed problem (D.1) the incentive constraints on firms and CEO's are replaced with the firms' first order and the CEOs' envelope conditions respectively. Monotonicity of optimal z and w (and, hence, π) for the relaxed problem are checked ex post in all of our numerical calculations. Problem (D.1) may be formulated as an optimal control problem in which Φ , z and w are the state variables, z_v is the control variable and \tilde{v} is a choice variable. The Hamiltonian for this optimal control problem is:

$$\begin{aligned}\mathcal{H}(v) &= -p^\Phi(v) U_e \left(\mathcal{C} \left[\Phi(v), \frac{z(v)}{h(v)} \right], \frac{z(v)}{h(v)} \right) \frac{z(v)}{h(v)} \frac{h_v(v)}{h(v)} \\ &\quad + p^z(v) z_v(v) + p^w(v) V_z(S(v), z(v)) z_v(v) \\ &\quad + \chi V(S(v), z(v)) + (1 - \chi) w(v) - \mathcal{C} \left[\Phi(v), \frac{z(v)}{h(v)} \right] + \psi(v) \Phi(v),\end{aligned}$$

with co-states p^Φ , p^z and p^w . Let q^V denote the multiplier on $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) \geq 0$ and q^U the multiplier on $\Phi(\tilde{v}) - \bar{U} = 0$. The first order condition for z_v implies that:

$$p^z + p^w V_z(S, z) = 0. \quad (\text{D.2})$$

Differentiating (D.2) with respect to v gives:

$$p_v^z + p_v^w V_z(S, z) + p^w [V_{zS}(S, z) S_v + V_{zz}(S, z) z_v] = 0. \quad (\text{D.3})$$

The optimal co-state equations are:

$$p_v^\Phi = \frac{1}{U_c} + p^\Phi \frac{U_{ec}}{U_c} \frac{z}{h} \frac{h_v}{h} - \psi \quad (\text{D.4})$$

$$p_v^w = -(1 - \chi) \quad (\text{D.5})$$

$$p_v^z = p^\Phi \left\{ \left[U_{ec} \left(-\frac{U_e}{U_c} \right) + U_{ee} \right] \frac{z}{h} + U_e \right\} \frac{h_v}{h} \frac{1}{h} - p^w V_{zz} z_v - \frac{U_e}{U_c} \frac{1}{h} - \chi V_z. \quad (\text{D.6})$$

The first order condition for \tilde{v} is:

$$\mathcal{H}(\tilde{v}) - \psi(\tilde{v}) \bar{U} - T^0 + q^V V_S(S(\tilde{v}), z(\tilde{v})) S_v(\tilde{v}) \geq 0, \quad (\text{D.7})$$

with equality if $\tilde{v} \in (0, 1)$. The transversalities at $v = \tilde{v}$ are:

$$p^w(\tilde{v}) = -q^V \quad p^z(\tilde{v}) = q^V V_z(S(\tilde{v}), z(\tilde{v})) \quad p^\Phi(\tilde{v}) = q^U. \quad (\text{D.8})$$

There are also transversalities at $v = 0$. This is because (unlike typical optimal control problems), there are no initial conditions for w and Φ . We have:

$$\lim_{v \downarrow 0} p^w(v) = \lim_{v \downarrow 0} p^z(v) = \lim_{v \downarrow 0} p^\Phi(v) = 0.$$

Integrating (D.4) gives:

$$p^\Phi(v) = \int_0^v \left[\left(\frac{1}{U_c(u)} - \psi(u) \right) \exp \left\{ - \int_v^u \frac{U_{ec}(u')}{U_c(u')} \left(\frac{e(u')}{h(u')} \right) h_v(u') du' \right\} \right] du, \quad (\text{D.9})$$

while integrating (D.5) gives

$$p^w(v) = -(1 - \chi)v. \quad (\text{D.10})$$

Combining conditions (D.3) to (D.6) and (D.10) implies:

$$V_z + \frac{U_e/h}{U_c} = -\frac{p^\Phi}{h} \left\{ \left[U_{ec} \left(-\frac{U_e}{U_c} \right) + U_{ee} \right] \frac{z}{h} + U_e \right\} \frac{-h_v}{h} + (1 - \chi)v V_{zS}(-S_v). \quad (\text{D.11})$$

As described in the main text, this optimality condition captures the marginal benefits and costs associated with a small change in CEO v 's effective labor (holding her utility fixed). Together the CEOs' and the firms' first order conditions (14) and (16) imply:

$$(1 - T_w[w])V_z h U_c = -U_e. \quad (\text{D.12})$$

The optimal marginal tax rate must be set to align a CEO's private return to effort with the social return. Since the CEO's effective wage coincides with the firm's marginal product, this reduces to ensuring that her pre-tax return on effort equates to the right hand side of (D.11). If $\chi = 1$, then the right hand side of (D.11) equals the usual marginal informational rents term from the Mirrlees model and combining (D.11) and (D.12) and the definition of α_h the standard formula for optimal marginal tax rates obtains:

$$T_w[w] = \frac{1}{1 + \frac{1}{\tilde{p}^\Phi} \frac{\mathcal{E}^c}{1 + \mathcal{E}^u} \alpha_h}, \quad (\text{D.13})$$

where $\tilde{p}^\Phi = \frac{U_c p^\Phi}{1 - F(h)}$ is the normalized co-state. More generally, using (D.11) and (D.12) and the definitions of α_h and α_S , the optimal marginal tax rate is:

$$T_w[w] = \frac{1 + (1 - \chi) \frac{1}{\tilde{p}^\Phi} \frac{\mathcal{E}^c}{1 + \mathcal{E}^u} \frac{V_{zS} S}{V_z} \frac{\alpha_h}{\alpha_S}}{1 + \frac{1}{\tilde{p}^\Phi} \frac{\mathcal{E}^c}{1 + \mathcal{E}^u} \alpha_h}. \quad (\text{D.14})$$

Special utility cases In the case of quasilinear/constant elasticity $U(c, e) = c - \frac{\mathcal{E}}{1+\mathcal{E}} e^{\frac{1+\mathcal{E}}{\mathcal{E}}}$ and a multiplicative firm surplus DSz , (D.14) reduces to:

$$T_w[w] = \frac{1 + (1 - \chi) \frac{\mathcal{E}}{1+\mathcal{E}} \frac{\alpha_h}{\alpha_S}}{1 + \frac{\mathcal{E}}{1+\mathcal{E}} \alpha_h}. \quad (\text{D.15})$$

If instead CEO utility has the log-constant compensated elasticity form $U(c, e) = \log c - \log \left(1 + \frac{\mathcal{E}}{1+\mathcal{E}} e^{\frac{1+\mathcal{E}^c}{\mathcal{E}^c}} \right)$, then (D.14) becomes:

$$T_w[w] = \frac{1 + (1 - \chi) \frac{A_c - 1}{A_c} \frac{\mathcal{E}^c}{1+\mathcal{E}^u} \frac{\alpha_h}{\alpha_S}}{1 + \frac{A_c - 1}{A_c} \frac{\mathcal{E}^c}{1+\mathcal{E}^u} \alpha_h}, \quad (\text{D.16})$$

where $A_c = \int_0^v c(v') dv' / \int_0^v \{c(v') - c(v)\} dv'$ is the tail coefficient of consumption and c the consumption allocation. In Section 5, the quantitative section of this paper, we focus on the quasilinear/constant elasticity CEO preference case. If preferences are modified to be log-constant compensated elasticity, the formula (53) for recovering α_h from the data is unchanged up to the modification of the coefficients \mathcal{N} and \mathcal{P} to

$$\mathcal{N} := \frac{1}{D\mathcal{E}} - \frac{\mathcal{E}^c}{1 + \mathcal{E}^u} \quad \text{and} \quad \mathcal{P} := -\frac{\mathcal{E}^c}{1 + \mathcal{E}^u}.$$

Saez (2001) considers utility functions of the log-constant compensated elasticity form and selects compensated elasticity values of \mathcal{E}^c equal to 0.25 and 0.5. At high incomes \mathcal{E}^u is close to zero (and less than \mathcal{E}^c). These values thus imply values for $\frac{\mathcal{E}^c}{1+\mathcal{E}^u}$ that are greater than the values implied by our calibration for $\frac{\mathcal{E}}{1+\mathcal{E}}$. Equation (53) then implies even larger values of α_h than our calibration. Turning to (D.16) evaluated at $\chi = 1$, it follows that tax rates are reduced by the increased values for $\frac{\mathcal{E}^c}{1+\mathcal{E}^u} \alpha_h$, but increased by $\frac{A_c - 1}{A_c} < 1$.

Determining \tilde{v} Substituting the definition of the Hamiltonian and (D.8) into (D.7) gives:

$$\begin{aligned} & -p^\Phi(\tilde{v}) U_e(\tilde{v}) \frac{z(\tilde{v})}{h(\tilde{v})} \frac{h_v(\tilde{v})}{h(\tilde{v})} + \chi V(S(\tilde{v}), z(\tilde{v})) - T^0 \\ & + (1 - \chi) w(\tilde{v}) - \mathcal{C} \left[\bar{U}, \frac{z(\tilde{v})}{h(\tilde{v})} \right] + q^V V_S(S(\tilde{v}), z(\tilde{v})) S_v(\tilde{v}) \geq 0. \end{aligned} \quad (\text{D.17})$$

If $\chi = 1$, then the preceding expression reduces to:

$$V(S(\tilde{v}), z(\tilde{v})) - \mathcal{C} \left[\bar{U}, \frac{z(\tilde{v})}{h(\tilde{v})} \right] - T^0 - p^\Phi(\tilde{v}) U_e(\tilde{v}) \frac{z(\tilde{v})}{h(\tilde{v})} \frac{h_v(\tilde{v})}{h(\tilde{v})} + q^V V_S(S(\tilde{v}), z(\tilde{v})) S_v(\tilde{v}) \geq 0. \quad (\text{D.18})$$

The term $V(S(\tilde{v}), z(\tilde{v})) - \mathcal{C} [\bar{U}, z(\tilde{v})/h(\tilde{v})] - T^0$ captures the direct loss of social surplus from shutting down the \tilde{v} firm. The remaining terms capture the benefits

to closing the firm down in terms of reduced rents to firms and CEOs ranked above \tilde{v} . Note that (D.18) does not pin down $w(\tilde{v})$, which may take any value in $[0, V(S(\tilde{v}), z(\tilde{v}))]$. When $\chi = 1$, the policymaker does not care whether the social surplus at \tilde{v} is realized as firm profit $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v})$ (perhaps to be captured by taxation on claimants) or realized as CEO income tax revenues $T[w(\tilde{v})] = w(\tilde{v}) - C[\bar{U}, z(\tilde{v})/h(\tilde{v})]$. From (D.8) and (D.10), $q^V = -p^w(\tilde{v}) = (1 - \chi)\tilde{v} \geq 0$. Thus, if $\chi < 1$, then $q^V > 0$, $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v}) = 0$ and (D.17) reduces to:

$$V(S(\tilde{v}), z(\tilde{v})) - C\left[\bar{U}, \frac{z(\tilde{v})}{h(\tilde{v})}\right] - T^0 + (1 - \chi)\tilde{v}V_S(S(\tilde{v}), z(\tilde{v}))S_v(\tilde{v}) - p^\Phi(\tilde{v})U_e(\tilde{v})\frac{z(\tilde{v})}{h(\tilde{v})}\frac{h_v(\tilde{v})}{h(\tilde{v})} \geq 0. \quad (\text{D.19})$$

Inequality (D.19) is interpreted similarly to (D.18). Now, however, firm \tilde{v} generates no profit; everything above the consumption amount necessary to give CEO \tilde{v} utility \bar{U} is captured by the policymaker in the form of taxation on CEO \tilde{v} .

D.1 Firm Entry and the Marginal Social Value of Profit

The preceding analysis treats the marginal social value of aggregate profit χ as a fixed parameter. In Section 3 we discuss the interpretation and determination of χ . If the policymaker is solely concerned with maximizing tax revenues, then, in our baseline model, χ equals the tax on profit τ^F . Since profit is pure rent, it is optimal to select τ^F equal to 100% and, hence, χ equal to 1. In Section 3 we describe how the firm's ability to privately transfer resources to firm owners may restrain profit taxation and the marginal social value of profit below 1. It is natural to conjecture that in a richer model with a firm entry margin, the disincentive to create firms also deters high profit taxation and, hence, leads to a lower marginal social value of profit. In fact, while a firm entry margin can create a rationale for lower rates of profit taxation, it can also introduce an additional motive for valuing profit: after-tax profits retained by firms relax firm entry conditions. Consequently, the marginal social value of aggregate profit can exceed the profit tax rate. If firm entry conditions involve aggregate profit and there are no restrictions on the level of CEO incomes and taxes, then the marginal social value of profit can equal one independent of the profit tax.

We develop two simple extensions of our model that highlight the issues.⁵³ Through out the body of the paper and previously in this section we assumed a fixed population of candidate firms (because, for example, a fixed population of candidate entrepreneurs draw a finite number of business ideas from a given quality distribution) and a firm entry condition:

$$V(S(v), z(v)) - w(v) \geq 0.$$

This is equivalent to assuming that firm entry costs are fully deductible and that a firm knows its potential size $S(v)$ before entering. Since firm profit is pure rent, then, as noted, the optimal profit tax and the corresponding marginal social value

⁵³We thank a referee for encouraging us to explore this issue.

of profit are both one. If instead firms pay a common and non-deductible entry cost, then the relevant entry condition becomes:

$$(1 - \tau^F)\{V(S(v), z(v)) - w(v)\} \geq k > 0.$$

It remains the case that the marginal social value of profit χ equals the profit tax τ^F . Thus, restrictions on profit taxes, say from profit diversion, continue to restrain the marginal social value of profit at the optimum. Absent such restrictions (and absent restrictions on the income received $w(\tilde{v})$ and taxes paid $T[w(\tilde{v})]$ by the least talented active CEO and, hence, the *level* of the CEO income and tax schedules), the optimal CEO consumption and effective labor allocation equals that from a model with a unit shadow social weight on profit. The tax/wage implementation requires setting the profit tax (arbitrarily close) to 100% and the pre-tax income and taxes of the least talented CEO to arbitrarily small (negative) numbers. Any additional restriction on the level of CEO incomes or taxes prevents this optimum from being achieved. Further details are available on request.

Consider next the following alternative model of entry.⁵⁴ Before knowing its potential asset size S , a (candidate) firm draws a nondeductible entry cost k from a distribution L . If the firm chooses to enter, then it pays its cost k and draws an asset size S from a distribution G . If a firm draws $S(v)$, then it makes after-tax profit $(1 - \tau^F)\{V(S(v), y(v)) - w(v)\}$ if it chooses to produce and zero otherwise. A firm chooses to enter and pay the cost k if:

$$(1 - \tau^F) \int_0^{\tilde{v}} \{V(S(v), y(v)) - w(v)\} dv - k \geq 0,$$

where as before $V(S(\tilde{v}), y(\tilde{v})) - w(\tilde{v}) = 0$. Given τ^F , y and w , there will be a cost threshold \hat{k} such that the $L(\hat{k})$ firms with costs below this threshold choose to enter the market and the rest do not and:

$$(1 - \tau^F) \int_0^{\tilde{v}} \{V(S(v), y(v)) - w(v)\} dv - \tilde{k} = 0.$$

The central difference between this model and that in the last paragraph is that now “aggregate” profit, i.e. ex ante expected profit $\int_0^{\tilde{v}} \{V(S(v), y(v)) - w(v)\} dv$, rather than the profit of the smallest producing firm $V(S(\tilde{v}), z(\tilde{v})) - w(\tilde{v})$ relaxes the entry constraint. Total tax revenues are:

$$L(\hat{k}) \left\{ \tau^F \int_0^{\tilde{v}} \{V(S(v), y(v)) - w(v)\} dv + \int_0^{\tilde{v}} T[w(v)] dv \right\},$$

where to simplify the analysis tax revenues collected from unassigned CEOs are abstracted from. The objective may be rewritten as:

$$L(\hat{k}) \left\{ \tau^F \int_0^{\tilde{v}} \{V(S(v), y(v)) - w(v)\} dv + \int_0^{\tilde{v}} [w(v) - c(v)] dv \right\}.$$

⁵⁴For a related model, see [Scheuer and Werning \(2015\)](#).

Substituting the binding entry constraint into the objective, the latter may be re-expressed (on the space of feasible allocations) as:

$$L(\hat{k}) \left\{ \int_0^{\tilde{v}} \{V(S(v), y(v)) - w(v)\} dv + \int_0^{\tilde{v}} [w(v) - c(v)] dv - \hat{k} \right\}.$$

Note that this re-specified objective places equal weight on firm profits and CEO tax revenues. Given an optimal choice of \hat{k} , y and c and \tilde{v} may be chosen to maximize:

$$\int_0^{\tilde{v}} \{V(S(v), y(v)) - c(v)\} dv,$$

subject to the CEO incentive and participation constraints. Thus, firm profits and CEO income tax revenues are weighted equally, with the (normalized) marginal social value of profit set to 1. The (c, y) allocation pins down w_v via the firms' first order conditions, while τ^F and $w(\tilde{v})$ are chosen to satisfy:

$$(1 - \tau^F) \int_0^{\tilde{v}} \left\{ V(S(v), y(v)) - w(\tilde{v}) - \int_v^{\tilde{v}} w_v(v') dv' \right\} dv - \hat{k} = 0$$

and $V(S(\tilde{v}), y(\tilde{v})) - w(\tilde{v}) = 0$. Thus, if the firm entry constraint binds at $\hat{k} > 0$, the optimal value of $\tau^F < 1$.

Proceeding slightly differently, if $L(\hat{k})\psi$ is the multiplier on the firm entry condition, then, after normalizing by $L(\hat{k})$, the planner's Lagrangian inclusive of the firm entry condition implies that the shadow social weight on firm profit is:

$$\chi = \tau^F + \psi(1 - \tau^F).$$

Thus, the marginal social value of profit is enhanced by its impact on the firm entry constraint. The first order condition for $w(\tilde{v})$ is:

$$(1 - \tau^F) - \psi(1 - \tau^F) = 0,$$

which implies that $\psi = 1$. Hence, up to the determination of \hat{k} , the model is equivalent to one with equal social weighting of profit and CEO income tax revenues ($\chi = 1$) and no explicit firm entry constraint. Note that this is true regardless of the choice of τ^F . That $\chi = 1$ stems from the fact that the planner can transfer resources from tax revenues to firm profits one for one by adjusting the *level* of taxes and firm wages.

In summary, relaxation of the firm entry constraint by aggregate firm profits can enhance their social value beyond their direct contribution to tax revenues. Hence, the marginal social value of profit can exceed the profit tax rate. If aggregate profits enter the firm entry constraint (and there are no binding restrictions on $w(\tilde{v})$ and $T[w(\tilde{v})]$ and, hence, the *level* of CEO incomes and income taxes), then the marginal social value of profit can equal that of taxes.

E Connection with Gabaix and Landier (2008)

Our estimates of the talent distribution of CEOs are consistent with the evidence presented in Gabaix and Landier (2008). In this paper, the authors calibrate the tail index of the distribution of talent using evidence on the distribution of firm size and the pay to firm-size elasticity. Their evidence is consistent with prior work that characterizes the distribution of firm size as a Pareto distribution with coefficient approximately equal to 1 and with “Roberts’ Law” (the empirical regularity relating CEO compensation and firm size). These two facts imply a negative tail index of CEO talent consistent with a Weibull distribution. A previous version of our paper estimated the shape parameter of a GEV distribution fitted on CEO talent. For that case estimates of the shape parameter also implied a Weibull distribution in our setting. Our current approach is complementary to the one in Gabaix and Landier (2008). First, Gabaix and Landier (2008) use the data directly to guide them in selecting market capitalization as an empirical counterpart to S . We follow Terviö (2008) in using economic theory to connect firm market capitalization to firm surplus. This theory explicitly allows for long lasting effects of CEO labor and for adjustable capital. Second, Gabaix and Landier (2008) focus on asymptotic tail properties of the CEO talent distribution. They appeal to extreme value theory to justify functional form restrictions on this distribution. We instead completely characterize the right tail of the CEO talent and firm asset distributions. Knowledge of both these distributions is needed for the computation of the non-linear tax schedule for values of $\chi \neq 1$. Third, with respect to the model of Gabaix and Landier (2008), we allow for elastic labor. Finally, there are differences in the data used. Specifically, we use measures of CEO compensation calculated using the value of options when exercised; Gabaix and Landier (2008) use the value of options when granted.

F Mechanical and Behavioral Impacts of Local Marginal Tax Rate Changes

Here we compute the mechanical and behavioral impacts of marginal tax rate changes in our assignment setting. We focus on the quasilinear/constant elasticity CEO utility-multiplicative firm payoff setting. Let T denote a twice continuously differentiable tax function and consider the following perturbed function:

$$\tilde{T}(w) = \begin{cases} T[w] & w \in [0, w_0) \\ T[w] + \sqrt{\delta}[w - w_0] & w \in [w_0, w_0 + \sqrt{\delta}) \\ T[w] + \delta & w \in [w_0 + \sqrt{\delta}, \infty) \end{cases}.$$

Let \hat{w} and \hat{z} denote the initial equilibrium schedules for income and effective labor and let \tilde{w} and \tilde{z} denote the equilibrium schedules occurring after the tax perturbation. Let v_0 and z_0 be given by $w(v_0) = w_0$ and $z_0 = z(v_0)$.

CEOs (and firms) can be partitioned into four groups which we label groups

0,1,2 and 3 respectively. Group 0 consists of the least talented CEOs with types in $[v_0, 1]$. Their behavior is unaffected by the tax perturbation. Group 1 CEOs with types in $[v_1(\delta), v_0]$ bunch at the kink point in the tax schedule. They earn w_0 and supply effective labor z_0 . The threshold $v_1(\delta)$ satisfies:

$$(1 - T_w[w_0] - \sqrt{\delta})S(v_1(\delta))h(v_1(\delta)) = \left(\frac{z_0}{h(v_1(\delta))} \right)^{\frac{1}{\varepsilon}}.$$

Group 2 CEOs have types $(v_2(\delta), v_1(\delta))$ earn incomes between w_0 and $w_0 + \sqrt{\delta}$ and pay the higher marginal tax $T_w[w] + \sqrt{\delta}$. Their first order condition for effective labor supply is given by:

$$(1 - T_w[\tilde{w}(v; \delta)] - \sqrt{\delta})S(v)h(v) = \left(\frac{\tilde{z}(v; \delta)}{h(v)} \right)^{\frac{1}{\varepsilon}},$$

where the notation makes the dependence of the functions \tilde{w} and \tilde{z} on δ explicit. The threshold $v_2(\delta)$ is given by:

$$v_2(\delta) = \sup\{v : \tilde{w}(v; \delta) \geq w_0 + \sqrt{\delta}\}.$$

CEOs in group 3 have ranks $(0, v_2(\delta)]$. Their marginal taxes are determined by the original tax schedule and their first order conditions are given by:

$$(1 - T_w[\tilde{w}(v; \delta)])S(v)h(v) = \left(\frac{\tilde{z}(v; \delta)}{h(v)} \right)^{\frac{1}{\varepsilon}}.$$

At $v_2(\delta)$ there is a discontinuity in \tilde{w} and \tilde{z} . Firms and CEOs at and arbitrarily close to $v_2(\delta)$ must be optimizing. In particular, this implies that:

$$S(v_2)\tilde{z}(v_2(\delta); \delta) - \tilde{w}(v_2(\delta); \delta) = S(v_2(\delta); \delta)\tilde{z}_+(v_2(\delta); \delta) - \tilde{w}_+(v_2(\delta); \delta),$$

where $\tilde{z}_+(v_2(\delta))$ and $\tilde{w}_+(v_2(\delta))$ are, respectively, the right limits of \tilde{z} and \tilde{w} at $v_2(\delta)$ (i.e. the limits of effective labor supplied and incomes earned by group 2 CEOs).

The government's revenues after the perturbation are given by:

$$R(\delta) := \int_0^{v_2(\delta)} \{T[\tilde{w}(v; \delta)] + \delta\}dv + \int_{v_2(\delta)}^{v_0} \{T[\tilde{w}(v; \delta)] + \sqrt{\delta}[\tilde{w}(v; \delta) - w_0]\}dv. \quad (\text{F.1})$$

Mechanical effect The mechanical effect is obtained from the terms:

$$\int_0^{v_2(\delta)} \delta dv + \int_{v_2(\delta)}^{v_0} \sqrt{\delta}[\tilde{w}(v; \delta) - w_0]dv.$$

Totally differentiating this with respect to δ and setting δ to zero gives:

$$1 - M(w_0).$$

Behavioral effect Next we turn to the behavioral effect.⁵⁵ It is obtained from the remaining terms in (F.1):

$$\int_0^{v_2(\delta)} T[\tilde{w}(v; \delta)] dv + \int_{v_2(\delta)}^{v_0} T[\tilde{w}(v; \delta)] dv. \quad (\text{F.2})$$

Totally differentiating this with respect to δ and evaluating the limit as δ converges to 0 yields:

$$\begin{aligned} & - \frac{T_w[w_0]}{1 - T_w[w_0]} \frac{\mathcal{E} \frac{S_0 z_0}{w_0}}{1 + \mathcal{E} \mathcal{T}[w_0] \frac{S_0 z_0}{w_0}} m(w_0) w_0 \\ & + \frac{T_w[w_0]}{1 - T_w[w_0]} \frac{\mathcal{E} \frac{S_0 z_0}{w_0}}{1 + \mathcal{E} \mathcal{T}[w_0] \frac{S_0 z_0}{w_0}} m(w_0) w_0 \\ & \times \frac{1}{g(S_0) S_0 T_w[w_0]} \int_0^{v_0} \frac{T_w[\hat{w}(v)]}{1 + \mathcal{E} \mathcal{T}[\hat{w}(v)] \frac{S(v) \hat{z}(v)}{\hat{w}(v)}} \exp \left\{ - \int_v^{v_0} \frac{\frac{S_v(v')}{S(v')} \mathcal{E} \mathcal{T}[\hat{w}(v')] \frac{S(v') \hat{z}(v')}{\hat{w}(v')}}{1 + \mathcal{E} \mathcal{T}[\hat{w}(v')] \frac{S(v') \hat{z}(v')}{\hat{w}(v')}} dv' \right\} dv. \end{aligned} \quad (\text{F.3})$$

Despite its complexity this expression has a straightforward interpretation. A higher marginal tax “at” w_0 induces CEOs at this income to work less hard causing a reduction in revenues. This effect is captured by the term on the first line of (F.3). But it also raises the incomes of more talented CEOs with ranks $v \in (0, v_0)$ and, hence, the revenues collected from them. This is captured by the term spread across the second and third lines of (F.3).

Adding the mechanical term as well, setting the sum to zero and rearranging gives the optimal marginal tax rate for $\chi = 0$,

$$T_w[w] = \frac{1}{1 + \frac{m(w_0) w_0}{1 - M(w_0)} \tilde{\mathcal{E}}_w(w_0)},$$

where $\tilde{\mathcal{E}}_w(w_0)$ is defined as:

$$\tilde{\mathcal{E}}_w(w_0) := \frac{\mathcal{E} \frac{S_0 z_0}{w_0}}{1 + \mathcal{E} \mathcal{T}[w_0] \frac{S_0 z_0}{w_0}} \left\{ 1 - \frac{1}{g(S_0) S_0} \frac{1}{T_w[w_0]} \int_0^{v_0} \frac{T_w[\hat{w}(v)]}{1 + \mathcal{E} \mathcal{T}[\hat{w}(v)] \frac{S(v) \hat{z}(v)}{\hat{w}(v)}} \exp \left\{ - \int_v^{v_0} \frac{\frac{S_v(v')}{S(v')} \mathcal{E} \mathcal{T}[\hat{w}(v')] \frac{S(v') \hat{z}(v')}{\hat{w}(v')}}{1 + \mathcal{E} \mathcal{T}[\hat{w}(v')] \frac{S(v') \hat{z}(v')}{\hat{w}(v')}} dv' \right\} dv \right\}. \quad (\text{F.4})$$

Further manipulation establishes that at the $\chi = 0$ optimum,

$$\tilde{\mathcal{E}}_w(w_0) := \frac{\mathcal{E} \frac{S_0 z_0}{w_0}}{1 + \mathcal{E} \mathcal{T}(v_0) \frac{S_0 z_0}{w_0}} \left\{ 1 - \frac{1 - G(S_0)}{g(S_0) S_0} \right\}.$$

When $\chi > 0$, the impact of the marginal tax rate change on firm profits is also

⁵⁵In this appendix, we use the term “behavioral effect” to describe the overall effect of the tax rate change on CEO incomes and, hence, tax revenues. It consists of individual effective labor responses and an equilibrium response of the CEO income schedule.

relevant. The impact of such a change is obtained from:

$$\lim_{\delta \downarrow 0} \frac{\partial}{\partial \delta} \int_0^{v_0} \tilde{\pi}(v; \delta) dv, \quad (\text{F.5})$$

where $\tilde{\pi}(v; \delta) := S(v)\tilde{z}(v; \delta) - \tilde{w}(v; \delta)$. Calculating the limit in (F.5) yields:

$$\begin{aligned} m(w_0) & \frac{1}{g(S_0)S_0} \frac{\mathcal{E}S_0z_0}{1 - T_w[w_0]} \frac{1}{1 + \mathcal{E}\mathcal{T}[w_0] \frac{S_0z_0}{w_0}} \\ & \times \int_0^{v_0} \exp \left\{ - \int_v^{v_0} \frac{\frac{S_v(v')}{S(v')} \mathcal{E}\mathcal{T}[\hat{w}(v')] \frac{S(v')\hat{z}(v')}{\hat{w}(v')}}{1 + \mathcal{E}\mathcal{T}[\hat{w}(v')] \frac{S(v')\hat{z}(v')}{\hat{w}(v')}} dv' \right\} dv. \end{aligned} \quad (\text{F.6})$$

The term in (F.6) is the analogue of $\chi \frac{\Pi}{1-\tau} \mathcal{E}_\Pi$ in the derivation of (30). Adding it to the other behavioral and mechanical terms leads to an optimal tax equation for the non-linear setting analogous to (30).

G Optimal Affine Tax Rates When $\chi = 0$

In this section we calculate the top affine tax rate for CEOs when $\chi = 0$. Recall that if the policymaker attaches no social weight to profits ($\chi = 0$) or CEOs, then in the affine setting the optimal tax rate on top earning CEOs is given by the Diamond-Saez formula:

$$\tau^* = \frac{1}{1 + \mathcal{E}_W^* A_W^*}. \quad (\text{G.1})$$

The values of the elasticity \mathcal{E}_W^* and the tail coefficient A_W^* are those arising in the optimal equilibrium. In general, they are endogenous and jointly determined with the optimal policy. However, under the assumption of quasilinear/constant elasticity CEO preferences, a multiplicative firm objective and linearity of taxes in incomes above a threshold, the product $\mathcal{E}_W A_W$ is independent of the marginal tax rate (see Appendix B). Below we combine existing evidence on elasticities of taxable income and our own estimates of A_W in US data to obtain an estimate of the product $\mathcal{E}_W A_W$. We then recover the tax rate implied by formula (G.1).

Selecting a value for \mathcal{E}_W There is limited direct evidence on \mathcal{E}_W for CEO's. [Bakija et al. \(2012\)](#) estimate a fairly large elasticity of taxable income with respect to the retention rate of 0.7 for the top 0.1 per cent of US income earners using tax return data. In addition, they find that executives, managers, supervisors and financial professionals account for 60 per cent of the top 0.1 per cent income earners. Time series evidence shows a strong negative correlation between top marginal tax rates and CEO incomes in the US. However, regressions provided by [Frydman and Malloy \(2011\)](#) indicate a small contemporaneous response of CEO incomes to tax reforms. They reject a value of \mathcal{E}_W above 0.2. [Goolsbee \(2000\)](#) studies data from 1991 to 1995 and rejects an elasticity above 0.4. In the context of top income earners

(but not necessarily CEOs), [Diamond and Saez \(2011\)](#) select a value for \mathcal{E}_W of 0.25. Given this range of values we use multiple values for \mathcal{E}_W . We proceed cautiously and use the Diamond-Saez value of 0.25 as an upper bound. We also use a more conservative value of $\mathcal{E}_W = 0.1$.

Recovering A_W from CEO income data The model assumes a continuum of CEOs and firms. In the data there are, of course, a finite number of each. To connect the model to the data, we will treat CEO income (and later firm market size) data as if it is a noisy and incomplete realization of a continuum economy. We will then fit a distribution to the tail of this data and use this to derive tax policy implications for the corresponding (continuum) economy. In doing this we are implicitly assuming that the resulting policy implications are approximately optimal for the (repeated) draws of large finite firm and CEO populations occurring in the US.

In this appendix, we compute A_W using CEO compensation data from the Standard and Poor's ExecuComp database for the year 2011.⁵⁶ The measure of compensation considered includes the amounts received by a CEO (within a fiscal year) from salary, bonus, restricted stock grants and an evaluation of long term incentive pay. This last item is mostly comprised of options. The value of options received as compensation can be calculated either by evaluating at the time they are granted (using the Black-Scholes formula) or by determining the profit obtained at the time the options are exercised. This last approach is used by the IRS to determine the taxable amount and our benchmark results follow suite.⁵⁷

Recall that $A_W = \frac{W}{\Delta W}$ and if the right tail of the CEO income distribution is Pareto above a threshold income \bar{w} , then A_W is constant and equal to the (constant) Pareto coefficient α_w of the distribution above this threshold. Non-parametric calculations of A_W indicate that it is indeed quite stable above an income of \$10 million or so in our data (see Figure 4). Thus, we fit a Pareto distribution to the right tail of the CEO income distribution. We use a two step procedure of [Clauset, Shalizi, and Newman \(2009\)](#). This entails first estimating α_w by maximum likelihood at each fixed \bar{w} and then selecting the \bar{w} value (and corresponding α_w estimate) that maximizes a Kolmogorov-Smirnov goodness-of-fit statistic. An estimate for α_w of 2.1 with a 95 per cent confidence interval equal to (1.13 3.06) (in 2011 data) is obtained. The threshold \bar{w} is estimated to be 13.8 million dollars. We thus compute a top optimal marginal tax rate for CEOs in a continuum economy in which $A_W = 2.1$.

The estimated value of 2.1 for α_w is inline with the numbers typically used to describe the right tail of the income distribution in the taxation literature. [Saez \(2001\)](#) reports an estimated value of α_w equal to 2, while [Diamond and Saez \(2011\)](#) and [Piketty et al. \(2014\)](#) assume a slightly lower value of 1.5. This suggests that

⁵⁶We have computed estimates of A_W for all years from 1947 to 2011 using data from [Frydman and Sacks \(2010\)](#) and from ExecuComp. The estimates display fluctuations over time however the (time series average) of the Pareto coefficient is 2.23 well within the 95 per cent confidence intervals for the 2011 value. Details are available on request.

⁵⁷We have also computed estimates of A_W using the former measures of CEO income. They lead to slightly higher estimates of A_W . Details are available on request.

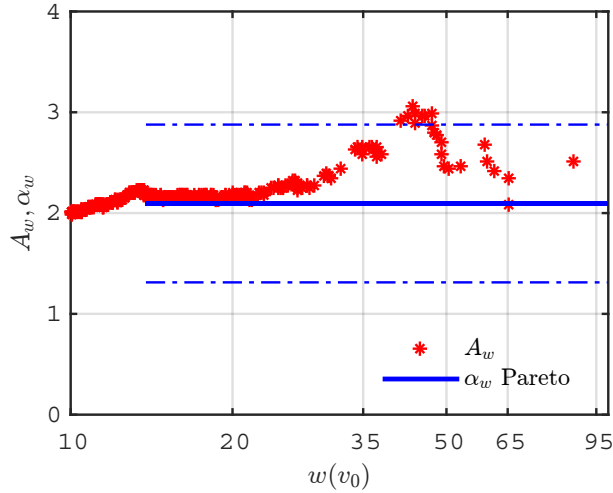


Figure 4: A_W denotes the tail coefficient for the income distribution. Figure 4 displays the level of the Pareto coefficient for income α_w above the threshold value of 13.8 millions. 95 per cent confidence intervals for α_w are also plotted.

the right tail of the US CEO income distribution is not very different from the tail of the general population's income distribution.⁵⁸

Computed optimal tax Together a value of A_W equal to 2.1 and an elasticity of taxable income \mathcal{E}_W equal to 0.25 imply an optimal tax rate on top CEO incomes (above the threshold \bar{w} of about \$14 million) of approximately 56 per cent. The more conservative taxable income elasticity $\mathcal{E}_W = 0.1$ together with $A_W = 2.1$ implies a marginal income tax on incomes above the threshold of close to 76 per cent. The latter combination of low elasticity and somewhat higher Pareto coefficient generates a top tax rate in line with those reported by [Diamond and Saez \(2011\)](#). Table 3 reports optimal marginal tax rates for different income thresholds w_0 based on these estimated coefficients, the coefficients defining the confidence interval and the empirical A_W values displayed in Figure 4 for \mathcal{E}_W equal to 0.1 and 0.25.

G.1 Comparisons of the CEO Tail Coefficient to that in the General Population

In this subsection we compare our estimate of A_W from the population of CEOs to estimates obtained from the entire population. [Saez \(2001\)](#) plots $\int_{w_0}^{\infty} wm(w)dw / w_0$ (p. 211) and observes that its value between incomes of \$100 thousand and about \$30 million in 1993 in current dollars is about 2. Hence, he infers that A_W is about 2. [Alvaredo, Atkinson, Piketty and Saez \(2013\)](#), using the World Top Incomes

⁵⁸In Appendix G.1 we estimate the Pareto coefficient for earned (labor) income amongst the general population. Our estimates for the period since 1990 are in the neighborhood of 2.

Table 3: Marginal Tax Rates (per cent): Affine Tax Rates

Tail Properties	Elasticity	w_0 (in millions)		
		30	40	60
(I) Based on Pareto Distribution	$\frac{1}{4}$	65.6 [57.3 76.8]		
	$\frac{1}{10}$	82.7 [77.0 89.2]		
(II) Based on A_W	$\frac{1}{4}$	62.6	65.0	61.9
	$\frac{1}{10}$	80.7	82.6	80.2

Notes: w_0 measured in millions of 2011 USD. Row (I) in the table displays rates based on a Paretian right tail of the income distribution. The rate is constant above a threshold value of 13.8 million dollars. Row (II) displays results based on the empirical tail coefficients of the income distribution.

Database (WTID), report a value of 1.6 for A_W in the US in 2013. [Diamond and Saez \(2011\)](#) and [Piketty, Saez, and Stantcheva \(2014\)](#) both use values of 1.5 in their analyses. These values are a little below the value of 2.1 for A_W that we find in our sample of CEOs. However, it should be noted that estimates of A_W for the general population refer to the distribution of all income irrespective of source. For example the definition of income in the WTID includes not only *wages, salaries and pensions* (which is the quantity of interest in the optimal tax analysis of this paper) but also: *entrepreneurial income, dividends, interest income and rents*. In addition these additional categories are of progressively more important for high income quantiles.

In the direction of correcting the WTID estimates for source, we use the data available in the WTID to obtain the tail parameter for earned income (*wages, salaries and pensions*). This data is, however, aggregated. We use the following strategy to purge non-labor income. Suppose total income y is the sum of earned income w and other sources of income z . We assume that w is distributed at the top according to a Pareto distribution with unknown tail parameter α_w . In addition we assume that there exist a strictly monotone relationship between w and y , so that ordering individuals by w or y will yield the same ranking. The WTID reports by percentile threshold (the thresholds are: 90th, 95th, 99th, 99.5th, 99.9th and 99.99th) both the fraction of total income due to earned income and the conditional average total income for that income group. We assume that for all individuals in a given income group i , earned income is related to total income according to: $w = \rho_i \cdot y$. Given information on the average y within an income group i we can then recover the conditional average for earned compensation \tilde{w}_i within the group. If the income distribution has a Pareto tail, it follows that the threshold earned income value for group i , \underline{w}_i , is related to \tilde{w}_i according to:

$$\tilde{w}_i = \frac{\alpha_w}{\alpha_w - 1} \underline{w}_i.$$

The Pareto assumption further implies:

$$\underline{w}_i = \frac{\bar{w}}{(1 - P_i)^{\frac{1}{\alpha_w}}},$$

where P_i is the fraction of agents below \underline{w}_i . Then considering two percentile categories and simplifying we obtain an estimate for α_w equal to:

$$\alpha_w = \frac{\log\left(\frac{1-P_i}{1-P_j}\right)}{\log\left(\frac{\tilde{w}_i}{\tilde{w}_j}\right)}. \quad (\text{G.2})$$

In Figure 5 we plot our estimates of α_w from the WTID. We also plot our estimates of the tail parameter using CEO compensation data (in blue) and the Pareto tail parameter reported for the entire WTID (in red). As noted earlier the coefficient

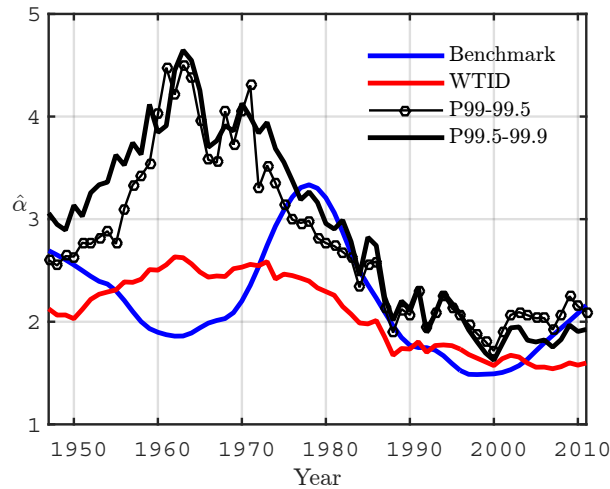


Figure 5: Estimates of α_w . Benchmark refers to the estimates from CEO data using options exercised. WTID from [Alvaredo, Atkinson, Piketty and Saez \(2013\)](#). P99-99.5 and P99.5-99.9 computed using equation (G.2). Compensation values from 1947 to 1991 are from [Frydman and Sacks \(2010\)](#). From 1992 onwards from ExecuComp.

using CEO data displays a more compact distribution than that obtained using the entire income distribution as reported in the WTID. However as we control for the sources of income focusing on earned income the difference becomes smaller. This is particularly evident in the earlier part of the sample. In terms of historical patterns all three approaches display a stretching out of the distribution starting from the 1970s through to 2000.

H Optimal Tax Allocations

This appendix describes the allocation associated with the optimal tax function at a benchmark parametrization. It provides additional information on the incidence of taxation at the optimum and undertakes counterfactual exercises that explore the role of effort and talent in generating dispersion in CEO incomes.

H.1 Preliminaries

We specialize the optimal control problem described in the body of the paper to the case of quasi-linear/constant elasticity CEO preferences and a multiplicative firm objective. Specifically, assume:

$$V(S(v), z(v)) := DS(v)z(v)$$

and

$$\Phi(v) := U\left(c(v), \frac{z(v)}{h(v)}\right) = c(v) - \frac{1}{1 + \frac{1}{\varepsilon}} \left(\frac{z(v)}{h(v)}\right)^{1 + \frac{1}{\varepsilon}},$$

so that

$$C[\Phi(v), z(v)/h(v)] = \Phi(v) + \frac{1}{1 + \frac{1}{\varepsilon}} \left(\frac{z(v)}{h(v)}\right)^{1 + \frac{1}{\varepsilon}}. \quad (\text{H.1})$$

The optimal control problem is then:⁵⁹

$$\sup_z \int_0^1 \left\{ \chi DS(v)z(v) + (1 - \chi)w(v) - \Phi(v) - \frac{1}{1 + \frac{1}{\varepsilon}} \left(\frac{z(v)}{h(v)}\right)^{1 + \frac{1}{\varepsilon}} \right\} dv \quad (\text{H.2})$$

subject to:

$$\Phi_v(v) = \left(\frac{z(v)}{h(v)}\right)^{1 + \frac{1}{\varepsilon}} \frac{h_v(v)}{h(v)} \quad (\text{H.3})$$

$$w_v(v) = DS(v)z_v \quad (\text{H.4})$$

$$\Phi(1) \geq \bar{U} \quad \text{and} \quad DS(1)z(1) \geq w(1). \quad (\text{H.5})$$

This is a standard optimal control problem with three state variables (Φ, w, z) and one control z_v .

Solving the optimal control problem We solve a scaled version of the optimal control problem in (H.2) using the numerical solver DIDO version 7.3.7.⁶⁰ The bulk of the parameter values utilized are those presented in the paper. In addition, we set $\chi = 0.8$ and $\bar{U} = 2$. This value of \bar{U} is low enough that all firms and CEOs are active. Alternative values of \bar{U} that ensure the full set of firms and CEOs are

⁵⁹Under the parameter values we consider $\bar{v} = 1$.

⁶⁰For details on the solution algorithm, refer to [Ross and Fahroo \(2003\)](#).

active affect the level of CEO consumption, but not the optimal effort allocation or optimal marginal taxes.

H.2 The Optimal CEO Effort Allocation

Variation in effort at the optimum The optimal effort allocation is displayed in Figure 6. Effort is decreasing and convex in v up until the median CEO at which

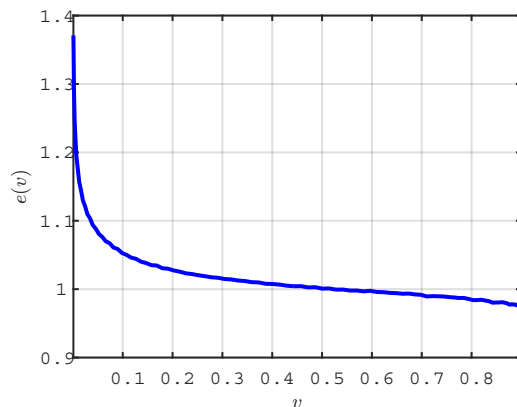


Figure 6: Effort.

point the profile becomes concave. Overall the top CEO’s effort is about 40 per cent more than that of the bottom. Care must be taken in interpreting these numbers since they are impacted by our functional form choices. However, the effort variation we find is of a similar order of magnitude to the hours variation amongst top CEOs reported by [Bandiera, Prat, and Sadun \(2011\)](#).⁶¹ Next we document the impact of optimal effort and effective labor variation on firm surplus and on CEO compensating differentials measured in dollars.

Impact of CEO effort and talent on firm surplus We next perform two counterfactuals that highlight the roles of CEO effort and talent in generating firm output at the optimum (under our parameterization). First, we fix the effort of all CEOs above the median at the level of the median CEO’s effort and recompute firm surplus. Second, we fix the talent of all CEOs above the median at that of the median, but keep their efforts fixed at their levels under the optimal allocation. Results are shown in Figure 7 as a fraction of optimal surplus. For the top firm, having a CEO exerting the same amount of effort as the median CEO (a CEO working in a company roughly 100 times smaller) causes a reduction in surplus of approximately 27 per cent. For the top firm, having a CEO with the same talent as the median (but exerting the higher effort of a highly talented CEO at the optimum), leads to a reduction in surplus of 17 per cent. Overall holding effort fixed at the median level

⁶¹These authors report that when CEOs are ranked by hours worked, then those at the 90th percentile work on average 20 hours (or roughly 50 per cent more) than those at the 10th percentile. They also document a positive relationship between recorded CEO effort and firm labor productivity.

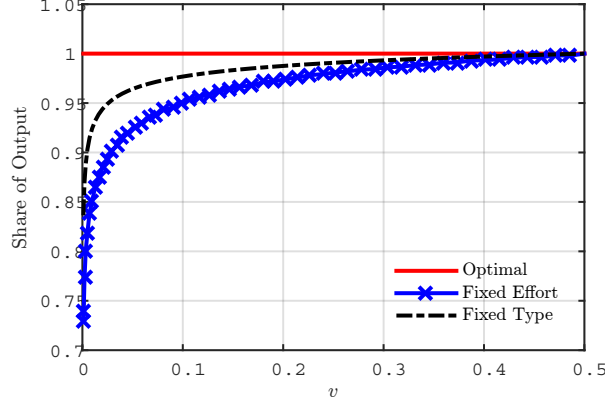


Figure 7: The Impact of effort and talent on firm surplus.

has slightly larger effect than holding talent fixed at this level. Reducing the effort elasticity \mathcal{E} further below $1/15$ brings the effect of holding talent fixed closer to that from holding effort fixed and eventually reverses the relative size of these effects.

Compensating differentials Define the compensating differential:

$$\Delta V(v) := -v \left(\frac{z(0.5)}{h(0.5)} \right) + v \left(\frac{z(v)}{h(v)} \right).$$

$\Delta V(v)$ gives the extra consumption needed to make the median CEO indifferent between her equilibrium allocation and an alternative allocation in which she supplies the effective labor of the v -th CEO. Since $\Delta V(v)$ is net-of-tax, it understates the increase in gross pay firm v would need to pay to induce the median CEO to supply CEO v 's effective labor.⁶² Figure 8 plots $\frac{\Delta V(v)}{w(v)}$, i.e. the compensating differential normalized by the equilibrium income of the v -th CEO $w(v)$. For the median CEO to be indifferent between her equilibrium allocation and that of an allocation in which supplies the effective labor of a top ranked CEO, her (after-tax) income would have to increase by an amount equal to 7 times that of the v -th CEO's gross pay. Thus, it is extremely costly to motivate the median CEO sufficiently that she replicates the top CEO's performance.

Tax counterfactual and burden We next analyze the incidence of taxation upon CEOs and firms. The optimal tax system is of the form:

$$T[w] = T[w_0] + \int_{w_0}^w T_w[w'] dw'$$

⁶²The CEO's gross pay would need to increase by $\Delta w(v) = C^{-1}[\Delta V(v) + c(0.5)] - w(0.5)$, where $C(w) := w - T[w]$ and $c(0.5)$ is the equilibrium consumption of the median CEO.

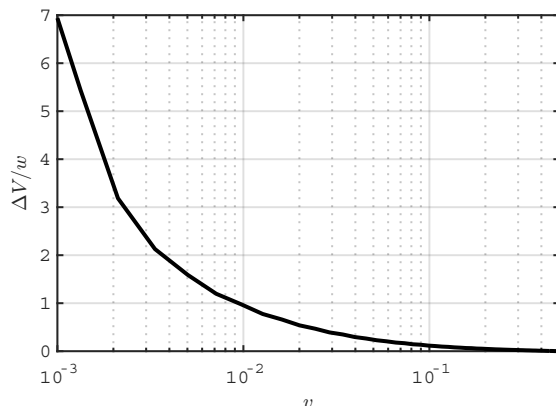


Figure 8: Compensating Differentials.

We consider a counterfactual in which T_w is set to zero at all w' .⁶³ We compute the corresponding equilibrium allocation and, hence, calculate the consumption gains for CEOs and profit gains for firms relative to the optimal allocation in which positive marginal taxes are used to collect additional tax revenue. For CEOs we then compute the proportional consumption gain from switching to the zero marginal tax equilibrium. We display these changes for different CEOs in the distribution in Figure 9(a). By definition the welfare impact of such a switch is zero for the CEO at the bottom since competition maintains her utility at \bar{U} . More talented CEOs work harder in response to the reduction in marginal tax rates and capture some of the increase in surplus in the form of additional utility (see the envelope condition (H.3)). These gains amount to a 3 per cent of consumption gain for the most talented CEOs. As noted throughout the paper, lower marginal tax rates on CEOs raise firm profits. Figure 9(b) shows that such increases are proportionally greater in smaller, lower profit firms, whose lower income CEOs face higher marginal taxes at the optimum.

APPENDIX REFERENCES

Alvaredo, F., A. Atkinson, T. Piketty, and E. Saez (2013). The Top 1 per cent in International and Historical Perspective. *Journal of Economic Perspectives* 27(3), 3–20.

Bandiera, O., A. Prat, and R. Sadun. (2013). Managing firms in an emerging economy: Evidence from the time use of Indian CEOs. Working Paper.

⁶³The lowest ranked active firm must pay its CEO enough that her after-tax income and, hence, consumption (and effective labor) is sufficient to attain her reservation utility \bar{U} . The lump sum tax $T[w_0]$ imposed on the lowest ranked active CEO is used by the policymaker to extract all profit from the smallest active firm. The ability to impose this tax has little effect on the CEO's utility, since competition always forces the lowest ranked CEO's utility to \bar{U} . However, it has a large proportional effect on the lowest ranked firm's profit. In this counterfactual we focus on the tax burden associated with non-negative marginal taxes.

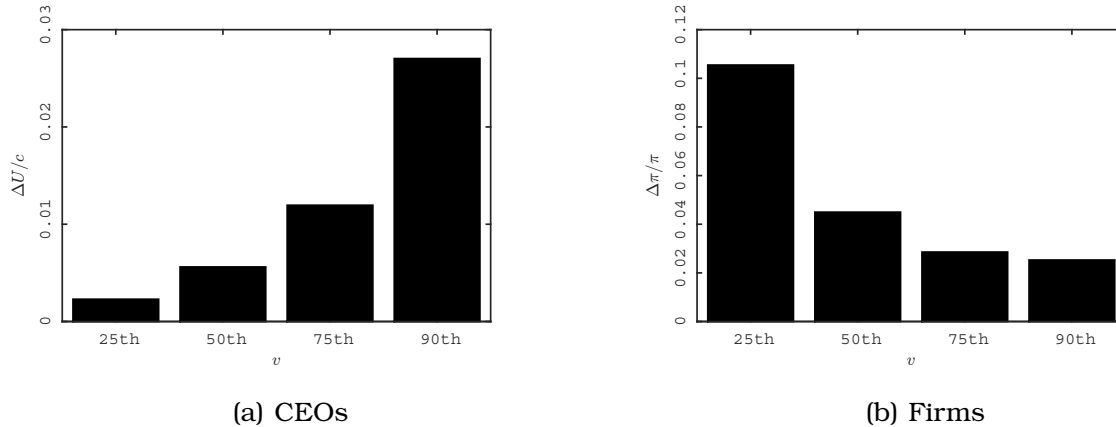


Figure 9: Tax Burden Calculations.

Clauset, A., C. Shalizi, and M. Newman (2009). Power law distributions in empirical data. *SIAM Review* 51(4), 661–703.

Diamond, P. and E. Saez (2011). The Case for a Progressive Tax: From Basic Research to Policy Recommendations. *Journal of Economic Perspectives* 25(4), 165–190.

Frydman, C. and R. Malloy. (2011). Does tax policy affect executive compensation? Evidence from post war tax reforms. *Journal of Public Economics* 95(11-12), 1425–1437.

Frydman, C. and R. Sacks. (2010). Executive compensation: A new view from a long term perspective, 1936-2005. *Review of Financial Studies* 23(5), 2099–2138.

Gabaix, X. and A. Landier. (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics* 123(5), 49–100.

Goolsbee, A. (2000). What happens when you tax the rich? Evidence from executive compensation. *Journal of Political Economy* 108(2), 352–378.

Klenke, A. (2008). *Probability theory: A comprehensive course*. Springer-Verlag.

Piketty, T., E. Saez and S. Stantcheva. (2014). Optimal Taxation of Top Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy* 6(1), 230–271.

Ross, I. and F. Fahroo. (2003). Legendre pseudispectral approximations of optimal control problems. *New Trends in Nonlinear Dynamics and Control and their Applications*, p.p. 327–342. Springer.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies* 68(1), 205–229.

Terviö, M. (2008). The difference that CEOs make: An assignment approach. *American Economic Review* 98(3), 642–668.