TAXING TOP CEO INCOMES

Laurence Ales

Christopher Sleet*

Tepper School of Business, Carnegie Mellon University

This Version: February 21, 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, SUNY-Stony Brook and the Wharton School.
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% or 80%. 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. If the policymaker has neither the ability to tax profits nor any direct social concern for claimants to these profits, then the classic Diamond-Saez formula remains intact. Otherwise, the optimal marginal tax rate on CEO incomes is modified downwards. We show that under a full reform of CEO income and profit taxation, the optimal marginal tax on top CEO incomes is about 15%.

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 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

\(^1\)The 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% of the top 0.1% earners 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. We also believe that our theoretical and broad quantitative insights carry over to other “superstar” buyer-seller relationships that generate high incomes for sellers.
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 the firm at which this unit is the last hired. Since the marginal product of 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 or because tax receipts and firm claimants are equally valued), 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 income taxation and is silent on how agents generate income, it applies in our setting, but only if the policymaker cannot tax and does not value 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%) and, thus, maximizes surplus not paid to CEOs. Since in our model firm profits are pure rents, 100% taxation of profits is weakly optimal (absent further restrictions on profit taxation or economic frictions and assuming that firm profits are insufficient to pay for government spending). Thus, the Mirrleesian formula is relevant if a comprehensive reform of profit and CEO income taxation is implemented. If a reform of only CEO income taxation is contemplated (holding firm profit taxation fixed and assuming

---

It 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.
limited direct concern for firm owners) or if institutional constraints and other economic frictions render a profit tax rate above zero, but below 100% optimal, 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, any productive 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 both formulas hold.

The simpler linear tax setting just described provides intuition for the analysis of optimal nonlinear tax formulas. Specifically, and analogous to the linear setting, when a zero 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. 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 utilize an empirical strategy similar to Terviö (2008), but extended 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 18% at an income of $10 million to about 10% at an income of $100 million. If a partial reform of CEO income taxation occurs holding profit taxes close to their empirical values in the US of about 0.6%, then optimal tax rates decline from 34% to 27% over a similar income range. In either case they are very different from the rates of 70% to 80% recently advocated in the literature (and which only remain valid in our setting under the strong assumptions that the policymaker cannot tax profits and has no direct concern for firm claimants).

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

---

4 Analogous 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.
CEO top incomes. Section 4 considers optimal non-linear taxation. It provides 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 tax implications of fiscal spillovers. 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. The labor supply response of workers of a given skill type to a change in tax rates impacts skill prices and, hence, the incomes earned by other workers of the same and different skill types. Workers who are not directly affected by the tax change may be impacted indirectly through changes to the wage distribution. 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 latter 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. Individual workers are free to work as much or as little as they want in the occupation of their choice at a given wage and they receive all of the surplus that they produce. These features contrast with our model in which a firm matches with a single CEO, a CEO appropriates only part of the surplus she creates and a unilateral decision to work less by the CEO requires rematching with a different, less demanding firm. Rothschild and Scheuer (2014) extend the framework of the papers mentioned above 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. Lockwood et al. (2014) also contribute in this direction. 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

---

5 We 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.

6 Specifically, 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.
offered by firms and, hence, induce redistributions amongst different worker types (rather than between CEOs and firm owners). We abstract from private information between firms and CEOs, but enrich the model with heterogeneity on the side of firms and an assignment structure.

Previous work on the taxation of top earners has tended to stress 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 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.

An older literature from the 1970s and 1980s explored the optimal structure of commodity in the presence of profits and exogenous restrictions on profit taxation. 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 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 and important 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 are (optimally) taxed at 100% and, hence, 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 “Mirrrleesian” 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 for-
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. 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 thin tail to the talent distribution.

In Rosen (1981)’s model of superstars, sellers of differing talent produce services of differing quality and are assigned in equilibrium to populations of consumers. In contrast to us, Rosen (1981) does not include an intensive effort margin nor

8Terviö (2008) in particular elaborates on the nature of these assets.
does he include heterogeneity in buyers’ tastes for quality. On the other hand, his model allows sellers to use goods to replicate the service they provide. Superstars earn higher incomes because they charge more (for their higher quality service) and sell more (given the higher return to replication implied by a higher price). In our baseline model, highly talented CEOs sell more and at a higher price to larger firms. In Ales and Sleet (2015), we extend our model to permit sellers to both enhance the quality of service through effort and, through a replication technology that uses goods, sell to more customers. Thus, we incorporate the additional force for income inequality emphasized by Rosen (1981).

Matching games in which agents make investments prior to or after trade are considered by Cole et al. (2001) and many others. Our model has only one-sided ex post ‘investment’ (in effort) by sellers. Thus, in this dimension, the model is simpler than Cole et al. (2001). However, we augment the framework with taxation, which is absent in Cole et al. (2001).

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.\(^9\) 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:

\[
z^s = h(v)e. \tag{1}
\]

We assume that a CEO can sell labor to only one firm in a period.\(^{10}\)

The utility of a CEO over consumption \( c \) and effort \( e \) is given by \( U : \mathbb{R}_+ \times [0, \bar{e}] \rightarrow \)

---

\(^9\)To 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 \([0, 1]\).

\(^{10}\)See Ales and Sleet (2015) for the generalization of this assumption.
$\mathbb{R}$, with $U$ strictly concave, twice continuously differentiable on the interior of its domain, strictly increasing in $c$ and strictly decreasing in $e$. $U$ is also assumed to satisfy the Spence-Mirrlees single crossing property: $-\frac{U_e(c, z^s/h)}{hU_c(c, z^s/h)}$ is decreasing in $h$.

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:

$$U \left( w - T[w], \frac{z^s}{h(v)} \right).$$  \hspace{1cm} (2)$$

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

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.\textsuperscript{11} 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:

$$V(S(v), z^b) - w,$$  \hspace{1cm} (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 $p$:

$$V(S(v), z^b) := \sup_x W(S(v), z^b, x) - p \cdot x.$$  \hspace{1cm} (4)$$

The input vector $x$ could include adjustable capital and we explicitly extend the

\textsuperscript{11}See Terviö (2008) and the references therein for additional discussion.
model in such a direction in our later quantitative section.\textsuperscript{12,13}

**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.\textsuperscript{14}

Let $w : I \to \mathbb{R}_+$ give the CEO income paid by each firm, $\mu : I \to I \cup \{u\}$ the talent rank of each firm’s CEO (with $\mu(v) = u$ indicating that $v$ is unmatched) and $z : I \to \mathbb{R}_+$ the quantity of CEO effective labor purchased by each firm. The functions $w$, $\mu$ and $z$ are assumed to be Lebesgue measurable. In addition, the match function $\mu$ 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 $(\mu, z, w)$ 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 $v \in \mu^{-1}(I)$,

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

\textsuperscript{12}Despite the exclusion of adjustable capital, we refer to $S(v)$ as firm $v$’s assets adding the qualifiers “non-transferable” or “immovable” when needed.

\textsuperscript{13}The 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 perfect 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.

\textsuperscript{14}A tractable and important special case assumes quasilinear-constant elasticity CEO preferences $U(c, e) = c - \frac{e}{1+e} e^{\frac{1}{2}E}$ and multiplicative firm payoffs $V(S, z) = D S z$. In this special case, absent 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.
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:

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

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 \( (z(v), w(v)) \):

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

and

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

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

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 \( (z(\hat{v}), w(\hat{v})) \):

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

with at least one of the applicable inequalities above strict.

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

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

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.\(^{15}\) The “no rents to the least

\(^{15}\)CEOs 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
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 sim-
ple 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-
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 com-
plicated) 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 $(\mu, z, w)$ 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 $\hat{v} \in I$ such that (i) for
all $v \in (\hat{v}, 1]$, $\mu(v) = u$ and (ii) for all $v \in (0, \hat{v})$, $\mu(v) = v$. Moreover, $z$ and $w$ satisfy the
**participation conditions.** for all $v \in (0, \hat{v}]$,

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

(11)

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

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

(12)

and

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

(13)

On the other hand, given $\bar{U}$, if $T, \hat{v}, z$ and $w$ are such that (i) for all $v \in (0, \hat{v}]$, (11)
to (13) hold, (ii) $U \left( w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(\hat{v})} \right) = \bar{U}$ and $V(S(\hat{v}), z(\hat{v})) - w(\hat{v}) \geq 0$ and (iii) for
effective labor from a CEO.
all \( w' \notin w((0, \hat{v}]) \), \( T[w'] = w' \), then \((\mu, z, w)\) with \( \mu \) such that for all \( v \in (\hat{v}, 1] \), \( \mu(v) = u \), and for all \( v \in (0, \hat{v}] \), \( \mu(v) = v \), 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 \((0, \hat{v}]\) 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 \( T \) is differentiable at \( w(v) \) and \( w \) and \( z \) are differentiable at \( v \in (0, \hat{v}] \), then (12) implies the CEO’s first order condition:

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

Similarly, (13) implies the first order condition for firms:

\[
V_z(S(v), z(v)) z_v(v) - w_v(v) = 0.
\]

It follows from (15) that in equilibrium, the “price” of a given increment of effective labor equals the marginal product of the firm for whom this increment is the last hired. Since more talented CEOs match with and supply more effective labor to larger firms and since, by super modularity of \( V \), \( V_z \) is increasing in \( S \), a force for enhanced dispersion of CEO income across talent ranks is created. 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. On the other hand, 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), firm claimants capture some surplus as profit. By the envelope condition, profit increases with firm size rank according to:

\[
\pi_v(v) = V_S(S(v), z(v)) S_v(v).
\]

Combining (14) and (15), totally differentiating with respect to \( v \) and denoting
the compensated and uncompensated effort elasticities by $\mathcal{E}^c$ and $\mathcal{E}^u$ gives:

$$w_v/w = \left( \frac{V_z}{w} \right) \left( 1 + \mathcal{E}^u \right) \left( \frac{h_v}{h} \right) + \mathcal{E}^c \frac{V_s S}{V_z} \left( \frac{S_v}{S} \right) \frac{1}{1 - \mathcal{E}^c \left( \frac{T_{ww} V_z}{1 - T_{ww}} + \frac{V_z}{V_z} \right)}.$$ \hspace{1cm} (16)

Expression (16) relates equilibrium CEO income variation across rank $w_v/w$ to talent $h_v/h$ and firm size variation $S_v/S$ 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 is multiplicative $V(S, z) = D S z$, with $D$ a parameter. If, in addition, the tax function $T$ is locally linear, then equation (16) reduces to:

$$w_v/w = \left( \frac{D S z}{w} \right) \left( \frac{h_v}{h} + \mathcal{E}^u \frac{h_v}{h} + \mathcal{E}^c \frac{S_v}{S} \right).$$ \hspace{1cm} (17)

This equality explicitly decomposes CEO income variation into a part due to variation in the talent rent accruing to CEOs $\left( \frac{D S z}{w} \right) \frac{h_v}{h}$ and a part due to variation in CEO effort $\left( \frac{D S z}{w} \right) \frac{h_v}{e} = \left( \frac{D S z}{w} \right) \left( \mathcal{E}^u \frac{h_v}{h} + \mathcal{E}^c \frac{S_v}{S} \right).$ \hspace{1cm} (17) 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 (17) further reduces to $w_v/w = \left( \frac{D S z}{w} \right) \frac{h_v}{h}$. 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 (17) reduces to $w_v/w = \frac{h_v}{h} + \mathcal{E}^u \frac{h_v}{h}$ and CEO income variation is not enhanced by variation in firm size.\hspace{1cm} (18)

In addition to the previously defined talent and firm asset distributions $F$ and

\[16\] Here the compensated and uncompensated elasticities are given by $\mathcal{E}^c = -\frac{U_v/e}{2 U_v + \left( \frac{U_v}{e} \right)^2 U_v + U_v}$ and $\mathcal{E}^u = \frac{U_v/e - \left( \frac{U_v}{e} \right)^2 U_v + \left( \frac{U_v}{e} \right) U_v}{-2 U_v \left( \frac{U_v}{e} \right)^2 U_v + U_v}$ with each $U_x$ and $U_{xy}$, $x, y \in \{c, e\}$, giving the relevant partial first and second derivatives of $U$.

\[17\] Note 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 $S$.

\[18\] There is little systematic evidence on patterns of CEO hours worked. Bandiera et al. (2011) looks at CEOs from the largest Italian firms. The paper documents that CEOs vary considerably the amount of time worked: a CEO in the 90th percentile (in terms of hours worked) works 20 hours more than a CEO in the 10th percentile. The paper also confirms a positive relationship between CEO hours and 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. The study also documents a positive relationship between firm size and hours worked.
G and densities f and g, let M denote the distribution of CEO incomes and m its density. Equation (17) can be usefully re-expressed in terms of the local Pareto coefficients for these distributions: \( \alpha_h(v) := \frac{h(v)f(h(v))}{1-F(h(v))} \), \( \alpha_S(v) := \frac{S(v)g(S(v))}{1-G(S(v))} \) and \( \alpha_w(v) := \frac{w(v)m(w(v))}{1-M(w(v))} \). Using \( f(h(v)) = -\frac{1}{h(v)} \), \( g(S(v)) = -\frac{1}{S(v)} \), \( m(w(v)) = -\frac{1}{w(v)} \) and multiplying by the counter-cumulative distributions (17) becomes:

\[
\frac{1}{\alpha_w} = DS_z \frac{1}{w} \left\{ (1 + E^u) \frac{1}{\alpha_h} + E^c \frac{1}{\alpha_S} \right\}. \tag{18}
\]

Similarly, letting \( \alpha_\pi \) 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:

\[
\frac{1}{\alpha_\pi} = \frac{DS_z \frac{1}{\pi}}{\alpha_S}. \tag{19}
\]

Together (18) and (19) permit recovery of the local Pareto coefficients for (unobservable) CEO talent and firm asset from the corresponding coefficients for CEO incomes and firm profits. In particular, (18) and (19) imply:

\[
(1 + E^u) \frac{1}{\alpha_h} = w \frac{1}{DS_z} \frac{1}{\alpha_w} - E^c \frac{\pi}{DS_z} \frac{1}{\alpha_\pi}. \tag{20}
\]

Larger values for the reciprocal of the Pareto coefficient in the right tail of a distribution indicate a thicker or fatter tail.\(^{19}\) Thus, (20) 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 (20), \( \frac{1}{\alpha_w} \) may be large (a fat CEO income tail) and \( \frac{1}{\alpha_\pi} \) small (a thin CEO talent tail) if \( \frac{w}{DS_z} \) 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

\(^{19}\)A 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.
linear above a threshold income $w_0$:

$$T[w] := T[w_0] + \tau (w - w_0) \quad w \in [w_0, \infty).$$  \hfill (21)

We consider the consequences of variation in $\tau$ (keeping $w_0$, $T[w_0]$ and $U$ 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 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 $1 - \tau$. Similarly, we make the dependence of equilibrium effective labor on the tax rate and define $z(v; 1 - \tau)$ to be the equilibrium effective labor of CEO $v$ given retention rate $1 - \tau$.\footnote{Function $\omega$ is well defined, see Appendix A.}

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

$$\omega(z(v; 1 - \tau); 1 - \tau) = \omega(z_0; 1 - \tau) + \int_{z_0}^{z(v; 1 - \tau)} DS(v(z'; 1 - \tau))dz',$$  \hfill (22)
The overall impact of a rise in $1 - \tau$ on the income of CEO $v$ is thus:\footnote{Equation (23) uses $\frac{\partial \omega}{\partial (1 - \tau)}(z_0; 1 - \tau) = 0$, see Appendix B for the derivation.}

\[
\frac{d \omega}{d (1 - \tau)}(v; 1 - \tau) = \frac{\partial \omega}{\partial z}(z(v; 1 - \tau); 1 - \tau) \frac{\partial z}{\partial (1 - \tau)}(v; 1 - \tau)
\]
\[
+ \frac{\partial \omega}{\partial (1 - \tau)}(z(v; 1 - \tau); 1 - \tau)
\]
\[
= DS(v) \frac{\partial z}{\partial (1 - \tau)}(v; 1 - \tau)
\]
\[
+ \int_{z_0}^{z(v; 1 - \tau)} \left\{ DS_v(v(z'; 1 - \tau)) \frac{\partial v}{\partial (1 - \tau)}(z'; 1 - \tau) \right\} dz'.
\]

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.\footnote{More precisely, if CEO preferences are quasilinear and there are no income effects on effort or if income effects on effective labor are negative, but the CEO’s uncompensated behavioral elasticity of effort is positive, then a rise in $1 - \tau$ induces CEOs to work harder, see Appendix B.} The second term gives the impact of the tax change on the pricing schedule holding the CEO’s effective labor fixed at $z(v; 1 - \tau)$. 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 $z(v; 1 - \tau)$.

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:

\[
\frac{d \pi}{d (1 - \tau)}(v; 1 - \tau) = - \frac{\partial \omega}{\partial (1 - \tau)}(z(v; 1 - \tau); 1 - \tau)
\]
\[
= - \int_{z_0}^{z(v; 1 - \tau)} \left\{ DS_v(v(z'; 1 - \tau)) \frac{\partial v}{\partial (1 - \tau)}(z'; 1 - \tau) \right\} dz' > 0. \tag{24}
\]

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 $E_w(v; 1 - \tau)$ and $\pi_v(w; 1 - \tau)$ 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 (23) and (24) to derive expressions for $E_w(v; 1 - \tau)$ and $\pi_v(w; 1 - \tau)$. 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 $U(c, w) = c - E_1 + E E_1$, with $E > 0$ the elasticity of CEO effort, then $E_w(v; 1 - \tau)$ and $\pi_v(w; 1 - \tau)$ are given by:

$$E_w(v; 1 - \tau) = \frac{DS(v)z(v; 1 - \tau)}{w(v; 1 - \tau)} E - \left(\frac{\pi(v; 1 - \tau) - \pi_0}{w(v; 1 - \tau)}\right) E.$$  (25a)

$$\pi_v(v; 1 - \tau) = \left(\frac{\pi(v; 1 - \tau) - \pi_0}{\pi(v; 1 - \tau)}\right) E,$$  (25b)

where $\pi_0$ is the profit of the smallest firm paying its CEO at least $w_0$ and $w(v; 1 - \tau) = \omega(z(v; 1 - \tau); 1 - \tau)$.

### 3 Optimal Linear Taxes

As a precursor to analysis of optimal nonlinear taxes, we consider the problem of selecting an optimal linear tax function over a range of top incomes. 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, the ability to tax these profits or direct concern for firm claimants 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 (21) and that she selects a marginal tax rate $\tau$ to maximize a weighted sum of income tax revenues and firm profits:

$$\sup_{\tau \in [0, 1]} \int_0^{v_0} \{w(v; 1 - \tau) - w_0\}dv + \chi \int_0^{v_0} \pi(v; 1 - \tau)dv,$$  (26)

keeping $\bar{U}$ and $w_0$ fixed.\(^\text{23}\)

\(^{23}\)Here, $v_0$ denotes the rank of the least talented CEO earning at least $w_0$. Note that the choice of $\tau$ does not the incomes of CEOs earning less than $w_0$ and, hence, does not affect $v_0$. 

18
Interpreting The Weight χ  The weight χ is an important parameter in our analysis. A first interpretation is that χ is the marginal tax on firm profits. In this case, the policymaker’s objective (26) is composed entirely of tax revenues (raised from CEO income and profit taxation) and the policymaker’s problem corresponds to the maximization of tax revenues via choice of the CEO income tax rate τ holding the tax rate χ on firm profits fixed. In our 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 χ = 1. Analysis of optimal CEO income taxation with χ fixed at a value of less than one then corresponds to a partial reform of the tax system (with the profit tax rate fixed at a sub-optimal level) or reflects a fully optimal outcome under some (unmodelled) economic friction or institutional constraint that restricts profit taxes. For example, if the asset S could relocate outside of the tax jurisdiction\textsuperscript{24} or if some non-publicly observable and, hence, non-expensable cost had to be incurred to accumulate S in the first place, then a tax rate on profits below 1 may be desirable.\textsuperscript{25}

The weight χ may also incorporate direct social concern for firm claimants. For example, if χ\textsuperscript{F} is the welfare weight placed on firm claimant incomes (inclusive of any lump sum transfers), τ\textsuperscript{F} is the tax rate imposed on profit and, as before, a unit welfare weight is placed on tax revenues, then the effective weight placed on profits is: χ = τ\textsuperscript{F} + χ\textsuperscript{F}(1 − τ\textsuperscript{F}). If τ\textsuperscript{F} is (constrained to be) less than one, then the effective weight on profits is enhanced by direct concern for firm claimants and is entirely due to such concern if τ\textsuperscript{F} = 0.

Further Assumptions  The objective in (26) places no direct weight on CEOs. In our quantitative analysis in Section 5, we focus upon CEOs earning salaries above $10 million. The zero welfare weight placed on CEOs thus 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

\textsuperscript{24}In this case, even if profits were repatriated it might be difficult for the local tax authority to distinguish between profit and returns to adjustable capital. Moreover, some profit would be captured by the foreign tax authority.

\textsuperscript{25}A 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), Munk (1980) and Huzinga and Nielsen (1997). 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%, this concern is less important and a higher commodity tax may be warranted. This parallels our result that if χ < 1, then marginal taxes on CEO incomes are optimally higher.
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 \( \tau \) to the extent that they benefit from tax revenues or have a claim on firm profits.\(^\text{26}\)

**Aggregate Elasticities and Tail Coefficients** To state the optimal tax formulas implied by (26), 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 \( \Pi \) the corresponding total firm profit. For notational ease we suppress the dependence of \( W, \Pi \) and other variables on \( 1 - \tau \) for the remainder of this section. Define the “aggregate” elasticities at the optimum by:

\[
\varepsilon_W = \frac{1 - \tau}{W} \frac{\partial W}{\partial (1 - \tau)} \quad \text{and} \quad \varepsilon_{\Pi} = \frac{1 - \tau}{\Pi} \frac{\partial \Pi}{\partial (1 - \tau)}.
\]

(27)

To the extent that high income earners are identified with CEOs, \( \varepsilon_W \) is the model counterpart of the elasticity of taxable income (ETI) for high earners emphasized in the empirical public finance literature.\(^\text{27}\) These elasticities are aggregates of the individual level elasticities given in the preceding section:

\[
\varepsilon_W = \int_0^{\tau_0} \frac{W}{W} \varepsilon_w (v) dv \quad \text{and} \quad \varepsilon_{\Pi} = \int_0^{\tau_0} \frac{\Pi}{\Pi} \varepsilon_{\pi} (v) dv.
\]

Both incorporate adjustment in the equilibrium CEO income and effective labor schedules. By our previous discussion, \( \varepsilon_{\Pi} \) is positive under reasonable economic restrictions. In particular, in the quasilinear/constant elasticity setting we have:

\[
\varepsilon_W = \left\{ 1 + \frac{\pi_0}{W} \right\} \varepsilon \quad \text{and} \quad \varepsilon_{\Pi} = \frac{1}{A_{\Pi}} \varepsilon,
\]

(28)

where \( A_{\Pi} = \frac{\Pi}{\Delta \Pi} \), with \( \Delta \Pi = \Pi - \pi_0 \), and as before \( \varepsilon \) 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 \( A_{\Pi} \) as the tail coefficients of CEO income and profit. It is easy to verify that:

\[
\frac{1}{A_W} = \int_0^{\tau_0} \frac{w(v)}{W} \frac{1}{w_w (v)} dv.
\]

\(^{26}\)If (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.

\(^{27}\)See Saez et al. (2012) for an extensive discussion of the role of ETI in public finance. In the language of Scheuer and Werning (2015), \( \varepsilon_W \) and \( \varepsilon_{\Pi} \) are “macro-elasticities.”
In particular, if CEO income has a Paretian right tail above \( w_0 \), then \( A_W = \alpha_w \), where \( \alpha_w \) is the constant Pareto coefficient.

**Optimal tax formulas** With the preceding definitions in place, a concise formula for the optimal marginal tax rate \( \tau^* \) is available. Rearranging the first order condition from (26) and using the definitions given above yields:

\[
\tau^* = \frac{1}{1 + A^*_W \frac{\varepsilon^*_W + \chi \frac{\Pi^*_{II}}{w^*_E} \varepsilon^*_E}{1 - \frac{\chi A^*_W \frac{\Pi^*_{II}}{w^*_E} \varepsilon^*_E}{\Pi^*_W} - A^*_W \frac{\varepsilon^*_E}{\Pi^*_W}}},
\tag{29}
\]

where *’s denote optimal values. Formula (29) contrasts with the standard expression derived by Saez (2001) and emphasized by Diamond and Saez (2011):

\[
\tau^{Saez} = \frac{1}{1 + A^*_W \varepsilon^*_W}. \tag{30}
\]

The logic behind (29) extends that behind (30) to include concern for the spillover from CEO income taxes to profits. In particular, the term \( \frac{\varepsilon^*_W + \chi \frac{\Pi^*_{II}}{w^*_E} \varepsilon^*_E}{1 - \frac{\chi A^*_W \frac{\Pi^*_{II}}{w^*_E} \varepsilon^*_E}{\Pi^*_W} - A^*_W \frac{\varepsilon^*_E}{\Pi^*_W}} \) in (29) replaces the standard elasticity \( \varepsilon^*_W \) in the classic Diamond-Saez formula (30). Given \( \varepsilon^*_W > 0 \) these terms are equal only if profits receive no weight in the policymaker objective (\( \chi = 0 \)). More generally, if \( \chi > 0 \), then the depressing effect of \( \tau \) on profits creates a motive for lower marginal taxes.

**Optimal taxes when \( \chi = 0 \).** Diamond and Saez (2011) and Saez et al. (2012) use formula (30) 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 \( \chi = 0 \), i.e. if firm profits are untaxed and the marginal welfare weight on recipients of firm profits is zero. It is also approximately valid if \( \varepsilon^*_W \) is small. Even if neither of these conditions hold, evaluations of (30) provide a useful benchmark for subsequent optimal tax calculations. Various authors proceed as if \( A_W \varepsilon_W \) is relatively stable in the face of marginal tax rate changes and use empirical evaluations of \( A_W \varepsilon_W \) 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 quasilinear/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 \( \varepsilon_W \) to 0.25 implying a
top tax rate of 72.7%. It is possible that the population of top earning CEOs is different from the general population of top earners. In Appendix F, we estimate $A_W$ for the CEO population using Standard and Poor’s ExecuComp database and find that it is stable above an income of about $12 million (2014 USD) and equal to 2.1. There is limited direct evidence on $\mathcal{E}_W$ for top CEOs. Direct time series evidence shows a strong negative correlation between top marginal tax rates and CEO incomes in the US, see Frydman and Molloy (2011). Bakija et al. (2012) estimate a fairly large elasticity of taxable income of 0.7 for the top 0.1% of US income earners using tax return data. Their econometric specification attempts to control for shifting between the corporate and personal income tax bases. In addition, while their estimates are for the general population of top earners, they find that executives, managers, supervisors and financial professionals account for 60% of this population. Frydman and Molloy (2011) estimate a small contemporaneous response of CEO incomes to tax reforms in the cross section. 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. 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 (30) implies a top tax rate of 65.6%. 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%.

**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 (30) 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% of income is shifted and that this shifted income is taxed at 30%. Their modified version of (30) then implies that the optimal tax rate rises from 72.7% to 76.8%. 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%.

The assignment model has provided a useful framework within which to under-

\footnote{Stantcheva (2014) considers a model in which neither the policymaker nor firms observe workers 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 $-\lambda_{\Pi W}^\Pi \mathcal{E}_\Pi^*$, that captures the social desirability of such redistribution into her optimal tax formulas.}
stand variation in CEO pay both within the cross section and over time. Implicit in this model is the idea that CEO’s do not capture all of the value they create. As discussed above when augmented with an intensive effort margin and taxes this model further implies a negative spillover from CEO income taxes to profits. We focus upon this.\textsuperscript{29} The literature gives little guidance on the size of this spillover, i.e. on the magnitude of $\varepsilon_{1t}$. However, Figure 1 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 over the period 1920 to 2014.\textsuperscript{30} A negative

![Figure 1: The relationship between net corporate dividend payments and marginal top income tax rates.](image)

correlation between these variables is apparent. Along the lines of Piketty et al. (2014) (who focus on the relationship between top incomes and the retention rate), we estimate the following log linear relationship:

$$
\log\left(\text{Dividends}/\text{GDP}_t\right) = \beta_0 + \beta_1 \log(1 - \tau_t) + \varepsilon_t,
$$

where $\tau_t$ is the top marginal tax rate on income at time $t$. Estimates from 1929 to 2015 provide a positive and statistically significant elasticity of 0.202 (0.066).\textsuperscript{31} The inclusion of a linear trend returns an even larger elasticity estimate of 0.33 (0.07). 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. Moreover,

\textsuperscript{29}In 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.

\textsuperscript{30}Top marginal rates are taken from the tax foundations. Data on dividends is from the BEA Table 6.20A (series: A3302C0) data on GDP is from Table 1.1.5.

\textsuperscript{31}The standard errors are Newey-West standard errors with 8 lags.
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. Given this,
in Section 5, we extend an alternative approach of Terviö (2008) to isolate the
structural parameters needed to calculate optimal (nonlinear) tax functions. We
then compute optimal tax rates for positive values of $\chi$. 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 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. 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 $\tau^*$ in (26) may be organized as:

$$
\frac{\tau^*}{1 - \tau^*} (W^*E_{W}^* + \Pi^*E_{\Pi}^*) + \left(\frac{\chi - \tau^*}{1 - \tau^*}\right) \Pi^*E_{\Pi}^* - \Delta W^* = 0.
$$

(31)

The first term gives the marginal behavioral impact on tax revenues of a change in
$1 - \tau$ if the entire firm surplus $R^* = \int_0^{10} r^*(v) dv$, $r^*(v) = DS(v)z(v;1 - \tau^*)$, is taxed at the
rate $\tau^*$. The second term adjusts the first to take into account the fact that profits
are only taxed at (or valued at) the rate $\chi$. 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 $\Delta W^* - \left(\frac{\chi - \tau^*}{1 - \tau^*}\right) \Pi^*E_{\Pi}^* = \Delta W^* - \Pi^*E_{\Pi}^* = (1 + \epsilon) \int_0^{10} r^* \frac{1}{\alpha_h} dv$, where, as before,
$\alpha_h$ is the local Pareto coefficient of the talent distribution. In addition, the first term
in (31) is simply $\frac{\tau^*}{1 - \tau^*} R^* \epsilon$. Consequently, when $\chi = 1$ and CEO preferences are of
the quasilinear and constant elasticity form, equation (31) implies:

$$
\tau^* = \frac{1}{1 + \frac{\epsilon}{1 + \frac{1}{\epsilon} \int_0^{10} \frac{1}{\alpha_h} dv}}.
$$

(32)

If the talent distribution has a Paretian right tail, then (32) further reduces to:

$$
\tau^* = \frac{1}{1 + \frac{\epsilon}{1 + \frac{1}{\epsilon} R^* \epsilon \alpha_h}}.
$$

(33)

Formulas (32) and (33) 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 (32) holds, but the standard formula expressed in terms of incomes (30) does not and when $\chi = 0$ the situation is reversed. For alternative $\chi$ values neither standard formula holds. The logic behind these results is straightforward. The standard formula expressed in terms of incomes (30) is valid when the policymaker seeks to maximize income tax revenues; the standard formula expressed in terms of talents (33) 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 tax revenue equals total surplus not captured by high income earners and so both formulas hold. But in the assignment setting either the policymaker is maximizing tax revenues or she is maximizing total surplus not captured by CEOs (or she is maximizing a weighted sum of income tax revenues and profits) and so at most one of the formulas holds.

Note that in the conventional labor supply setting (32) is consistent with a high optimal marginal tax rate. In particular, in this setting $r^* = w^*$ and $(1 + \E) \frac{1}{\alpha_h} = \frac{1}{\alpha_w}$ implying $\int_0^{w^*} \frac{1}{r^* \frac{1}{\alpha_h}} dv = \int_0^{w^*} \frac{1}{w^* \frac{1}{\alpha_w}} dv = A^* W$. In addition, $\E = \E^*_W$ and so $\int_0^{w^*} \frac{1}{r^* \frac{1}{\alpha_h}} dv = A^*_W \E^*_W$. As noted previously, empirical evaluations of the latter product 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 (20)) and so $\alpha_h$ is larger than $\alpha_w$ (and larger than would be implied by attempts to infer $\alpha_h$ from the income data using the standard labor market model). Thus, in the assignment model, $\int_0^{w^*} \frac{1}{r^* \frac{1}{\alpha_h}} dv > \int_0^{w^*} \frac{1}{w^* \frac{1}{\alpha_w}} dv = A^*_W$ and (32) is consistent with a lower optimal tax rate than (30).

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 one and then recover the 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 then 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 \((\tilde{v}, z, w, \Phi)\), where \(\Phi\) gives the CEO’s utility with \(\Phi(v) := U(c(v), z(v)/h(v))\). Let \(C(\phi, z/h)\) be the consumption of a CEO when her utility is \(\phi\) 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.\(^{32}\) 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.\(^{33}\)

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:

\[
\sup_{\tilde{v}, \Phi, z, w} \int_0^{\tilde{v}} \left\{ \chi V(S(v), z(v)) + (1 - \chi)w(v) - C[\Phi(v), z(v)/h(v)] \right\} dv + T_0 \int_0^1 dv \tag{34}
\]

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 \(v \in (0, \tilde{v}]\):

\[
\Phi_v(v) = -U_e \left( C \left( \frac{\Phi(v), z(v)}{h(v)} \right) \frac{z(v)}{h(v)} \frac{h_v(v)}{h(v)} \right) \tag{35}
\]

\[
w_v(v) = V_z(S(v), z(v))z_v(v). \tag{36}
\]

The optimal tax formula in terms of primitive distributions After manipulating the first order and co-state equations from the optimal control problem (34) (see Appendix C for details), the following optimality condition emerges for all \(v \in (0, \tilde{v}]\):\(^{35}\)

\[
V_z + \frac{U_c}{U_c} = - \frac{p^\Phi}{h} \left\{ U_{ee} \left( - \frac{U_c}{U_c} \right) + U_{ev} \right\} \frac{z}{h} + U_e \left\{ \frac{-h_v}{h} + (1 - \chi)vV_{zS}(-S_v) \right\} \tag{37}
\]

where \(p^\Phi\) is the co-state associated with CEO utility \(\Phi\). This condition captures the marginal benefits and costs associated with a small change in CEO \(v\)’s effective

\(^{32}\)In 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.

\(^{33}\)The general case in which CEOs receive non-negative welfare weights is analyzed in Appendix C.

\(^{34}\)The case \(\chi = 1\) is somewhat easier to solve since \(w\) no longer appears in (G.2). It can be formulated as an optimal control problem with two rather than three state variables.

\(^{35}\)All 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.
labor (holding her utility fixed). It has a very natural interpretation. The left hand side of (37) gives the marginal benefit of the 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 $U_c h / U_c$. 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 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 whom 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 (modulo adjustments in CEO consumption), and, if $\chi < 1$, welfare falls. If, on the other hand, $\chi = 1$ either because the policymaker places equal welfare weight on income tax revenues and firm profits or because the policymaker can tax firm profits at 100%, then this redistributional effect has no impact on welfare, the policy objective is to maximize surplus extracted from CEOs, and (37) 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 (15) imply:

$$ (1 - T_w[w])V_z h U_c = - U_c. $$

(38)

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 (37). If $\chi = 1$, then the right hand side of (37) equals the usual marginal informational rents term from the Mirrlees model and combining
and (38) and the definition of $\alpha_h$ the standard formula for optimal marginal tax rates obtains:

$$T_w[w] = \frac{1}{1 + \frac{1}{\tilde{p}^{\Phi}} \frac{\varepsilon^c}{1 + \varepsilon^u} \alpha_h},$$  \hspace{1cm} (39)$$

where $\tilde{p}^{\Phi}$ is the normalized co-state $\frac{U_c}{1 - F(h)} p^{\Phi}$. More generally, using (37) and (38) and the definitions of $\alpha_h$ and $\alpha_S$, the optimal marginal tax rate is:

$$T_w[w] = \frac{1 + (1 - \chi) \frac{1}{\tilde{p}^{\Phi}} \frac{\varepsilon^c V_{\varepsilon^S}}{1 + \varepsilon^u} \frac{\alpha_h}{\alpha_S}}{1 + \frac{1}{\tilde{p}^{\Phi}} \frac{\varepsilon^c}{1 + \varepsilon^u} \alpha_h}. \hspace{1cm} (40)$$

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 (37) 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 it achieves its 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 much more complicated in the nonlinear setting. We pursue such a perturbation now and, hence, relate equations (39) and (40) 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 E.

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:

\[- \frac{T_w[w_0]}{1 - T_w[w_0]} m(w_0) w_0 \hat{E}_w(w_0)\]

where \(\hat{E}_w(w_0)\) 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 E. 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 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:

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

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 \(\chi\) 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:

\[-\chi \frac{1}{1 - T_w[w_0]} \frac{\hat{E} S_0 z_0}{w_0} \frac{\alpha_w(v_0)}{1 + \hat{E}\mathcal{T}(w_0) S_0 z_0 \alpha S(v_0)} \times \int_0^{v_0} \exp \left\{ - \int_v^{v_0} \frac{S(v') \hat{E}\mathcal{T}(w^*(v'))}{w'(v')} \frac{S(v') z^*(v')}{w^*(v')} dv' \right\} dv.\]

The term (41) is the analogue for the nonlinear setting of the spillover term \(\frac{\Pi}{1 - \tau} \hat{E}_\Pi\) in the derivation of (29). Adding it to the other behavioral and mechanical terms leads to an optimal tax equation for the non-linear setting analogous to (29).
5 Quantitative Evaluation of Optimal Non-Linear Taxes

Equation (40) 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 multiplicative firm production functions. In this case (40) may be rewritten as:

\[ T^*_w[w^*(v)] = \frac{1 + (1 - \chi) E \frac{\alpha_h(v)}{1 + \alpha_s(v)}}{1 + \frac{E}{1 + \alpha_h(v)}}. \tag{42} \]

To quantitatively evaluate the implications of this formula for optimal taxes values for the weight \(\chi\), the Pareto coefficients \(\alpha_h\) and \(\alpha_s\) and the elasticity \(E\) are required. We discuss the choice of \(\chi\) first.

5.1 Selecting values for \(\chi\)

As described in Section 3 the weight \(\chi\) reflects the tax rate placed on profit, perhaps enhanced by direct social concern for firm claimants. If, as modeled here, firm claimant income is pure rent and a comprehensive reform of both top CEO income taxation and firm profit taxation is envisaged, then the optimal CEO income taxes obtained under the assumption \(\chi = 1\) are the relevant ones. In the US, corporate profits over much of the tax schedule are taxed at 35%; in addition, taxes are levied at the state level and on the disbursement of profits via dividend taxes on individuals. The OECD reports that the overall tax rate on distributed profits in the US is 57.6%. If top CEO income tax policy was to be reformed holding these other taxation policies fixed, the appropriate value for \(\chi\) would be (at least) 0.58. Direct social concern for firm claimants would further enhance this value. The extent of this social concern 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 refer to 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

\[ \text{Data taken from OECD Tax Database, Corporate and Capital Income Taxes: Table II.4. Historical values for the US range from 51.7 percent in 2003 to 88.7 in 1981. The average value for OECD countries in 2015 is 43.1 percent.} \]
Overall, 65 percent of these are in the hands of households with incomes of less than $500,000. The median household in the SCF reports an income of about $52,000 and equity (held directly and indirectly) valued at $33,000 and constituting 40\% of its financial wealth. Collectively, this provides some motivation for considering weights on firm claimant incomes above 0.58 even if the reform of tax policy is restricted to the CEO income tax.

5.2 Recovering measures of $\alpha_h$ and $\alpha_S$

In our firm-CEO assignment model, equations (18) and (19) related 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 $\alpha_h$ and $\alpha_S$. There are two complications in doing so. First use of (18) and (19) 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 a CEO effective labor may be realized over time. Specifically, a given set of CEO decisions may have a long lasting impact on the stream of firm 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 Tervio 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 (18) and (19) 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.

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 B.1). Total direct and indirect ownership of the household sector amounts to 19,495 billions (Table L223). Additional amounts are held by state and local pension funds (4,888 billions) and insurance and life insurance companies (2,053 billions) (Table L.223).
Connecting $\alpha_h$ and $\alpha_S$ 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 utility 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 $\lambda$. 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:

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

(43)

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 $\{CSZ_t(1+g)^t\}_{t=0}^{\infty}$ 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:

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

(44)

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:

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

(45)
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:

$$V_t(S, z_t) := (1 + g)^t DSz_t,$$

with $D := \frac{\lambda}{1 + \lambda - B}$. (46)

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:

$$w_v \frac{w}{w} = \left(\frac{DSz}{w}\right) \frac{z_v}{z}. \quad (47)$$

Equation (47) 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. 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:

$$P(v) := \frac{S(v)z(v) - w(v)}{1 - B}. \quad (48)$$

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 (44). Substituting this optimal choice and the definition of $P(v)$ in (48) into the definition for the market capitalization $Q(v)$ implies:

$$Q(v) = \left(\frac{1}{\Xi}\right) \frac{S(v)z(v)}{1 - B} - \frac{w(v)}{1 - B'}, \quad (49)$$

where $\Xi := \frac{1 - \theta}{1 - \theta + \frac{\theta}{\Xi}(1 - B)}$ inflates the firm surplus to account for adjustable capital. Hence, the firm’s first order condition (47) can be re-expressed as:

$$w_v \frac{w}{w} = \left(\frac{DSz}{w}\right) \frac{z_v}{z} = D\Xi \left(\frac{w + (1 - B)Q}{w}\right) \frac{z_v}{z}, \quad (50)$$

where (49) 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:

\[
\frac{z_v}{z} = \frac{h_v}{h} + \frac{e_v}{e} = (1 + \varepsilon) \frac{h_v}{h} + \varepsilon \frac{S_v}{S}.
\] (51)

Together (50) and (51) 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 that part of effort variation attributable to (unobserved) firm asset variation. This is done by totally differentiating (49) to obtain an expression for \( S_v / S \) and combining the result with (50) and (51) 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 \( \alpha_Q \) the local Pareto coefficient for firm market capitalization), yields:

\[
\frac{1}{\alpha_h} = \mathcal{N} \left( \frac{w}{w + (1 - B)Q} \right) \frac{1}{\alpha_w} + \mathcal{P} \left( \frac{(1 - B)Q}{w + (1 - B)Q} \right) \frac{1}{\alpha_Q},
\] (52)

where \( \mathcal{N} > 0 > \mathcal{P} \) are constants depending on the parameters \( \lambda, r, g, \theta \) and \( \varepsilon \). Equation (52) is the empirically operational version of (20). 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 (20)). It is distinct from analogous expressions in Terviö (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 (20), the weight \( \mathcal{P} = -\frac{\varepsilon}{1 + \varepsilon} \) 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 \( \alpha_S \) are also needed. A derivation very similar to that underpinning (52) yields:

\[
\frac{1}{\alpha_S} = \mathcal{M} \left( \frac{w}{w + (1 - B)Q} \right) \frac{1}{\alpha_w} + \left( \frac{(1 - B)Q}{w + (1 - B)Q} \right) \frac{1}{\alpha_Q},
\] (53)

where \( \mathcal{M} < 0 \) depends upon parameters.
Connecting CEO income and firm market capitalization data  We use (52) to recover estimates of the local Pareto coefficients for talent. This requires prior estimation of the local Pareto coefficients for CEO compensation and firm market capitalization along with calculation of the parameters $B$, $N$ and $P$. We use CEO compensation data for the year 2011 taken from the ExecuComp dataset. We compute firm market capitalizations using data on the number of shares outstanding and average monthly share price in 2011 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 is robust in the data, but is obviously not perfect. To bring the model to the data, we first sort by CEO income. We then impute corresponding market capitalizations by estimating the following log-linear relationship:

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

where $w_i$ is the income of the $i$-th CEO and $Q_i$ is the associated market capitalization of the firm he or she manages. Using the estimated coefficients $(\hat{\beta}_0, \hat{\beta}_1)$, market capitalization values are set equal to:\(^38\)

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

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 Pareto coefficients for both.

Selecting parameter values  We select a benchmark conservative value for the effort elasticity $E$ of $1/15$. However, we also consider a range of (conservative) effort elasticities from $1/30$ to $1/12$. For the remaining parameters, we follow Terviö (2008) and select values for $g = 0.025$, $r = 0.05$, $\lambda = 0.5$ and $\theta = 0.4$. Additional robustness tests of our results with respect to these parameters are performed in Subsection 5.4. Collectively, these choices pin down the parameters $B$, $N$ and $P$ in (52).\(^39\) Note that if, in addition to our CEO preference assumptions, the tax rate is

---

\(^38\)We experimented with a wide variety of smoothing techniques, in all cases the magnitude and behavior of the estimated $\alpha_h$ and the computed optimal tax rates are little changed.

\(^39\)Previous versions of the paper formally estimated the tail properties of the talent distribution. This was achieved by normalizing the value of talent at the bottom: an assumption that does not affect the tail properties of the distribution. For all estimators considered the distribution was determined to be thin tailed and best characterized with a Weibull-like distribution consistent with the calibration of Gabaix and Landier (2008).
linear above $w_0$, then $\mathcal{E}_W = \Gamma \mathcal{E}$, where $\Gamma := 1 + \frac{DS(0)z(0) - w(0)}{W}$. Using (49), our model further implies that $\Gamma = 1 + \frac{D\Xi(1-B)Q(0) + (D\Xi-1)w(0)}{W}$. Our data and parameter choices for $g$, $r$, $\lambda$ and $\theta$ (and, hence, $B$, $D$ and $\Xi$), imply a value of $\Gamma = 1.26$. Our choices of $\mathcal{E}$ then imply that $\mathcal{E}_W$ is in the range 0.042 to 0.11, with our benchmark $\mathcal{E} = 1/15$ implying $\mathcal{E}_W = 0.084$. These are conservative values for $\mathcal{E}_W$ relative to the literature.

**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 2. 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 parameter 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.

![Figure 2: Estimates of the local Pareto coefficient for talent $\alpha_h$ as a function of corresponding CEO income.](image)

5.3 Results

Substituting our empirical values for $\alpha_h$ and $\alpha_S$ into (42) along with values for $\mathcal{E}$ and $\chi$ gives optimal marginal tax rates as function of CEO rank $v$.\textsuperscript{40} Figure 3 shows

\textsuperscript{40}Calculation of optimal marginal tax rates as a function of $v$ using (42) does not require calculation of the entire optimal allocation. In Appendix G we compute the optimal allocation under a benchmark parameterization. We also undertake several counterfactual exercises suggested by a
these tax rates for the effort elasticity $\mathcal{E} = 1/15$ and for various values of $\chi$. Those for $\chi = 0.6$ and can be interpreted as optimal in the absence of a reform of current US profit taxes (and with very little additional concern for firm claimants); those for $\chi = 1$ as optimal under a full reform of CEO income and firm profit taxation. The

![Figure 3: Marginal tax rates as function of $v$.](image)

figure makes apparent that marginal tax rates are falling with $\chi$ 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 $\Delta w^*(v) = w^*(v) - w^*(\tilde{v})$.\footnote{The table highlights that over large ranges of high (but thinly populated) CEO incomes optimal marginal tax rates are declining in both income and $\chi$. At $\chi$ equal

referee that indicate the respective roles of talent and effort in inducing equilibrium CEO income variation.}

\footnote{$\Delta w^*(\tilde{v})$ is obtained by integrating out (16) using the values of the local Pareto coefficients previously obtained. For perspective, $w(\tilde{v})$ in 2011 US data was approximately $250,000.$}
Table 1: Tax Calculations

<table>
<thead>
<tr>
<th>$\Delta w^*$ (in millions)</th>
<th>30</th>
<th>40</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi = 0.6$</td>
<td>30.0</td>
<td>29.3</td>
<td>28.4</td>
</tr>
<tr>
<td>$\chi = 0.8$</td>
<td>22.0</td>
<td>21.3</td>
<td>20.2</td>
</tr>
<tr>
<td>$\chi = 1$</td>
<td>14.1</td>
<td>13.2</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Notes: Optimal non-linear tax rates as function of societal weight $\chi$ and income level.

to 0.6 marginal tax rates are below 30% on high incomes and declining in income, while at $\chi = 1$ they are below 15% and declining.

5.4 Robustness

Focusing on the case $\chi = 0.8$, we now undertake robustness analysis with respect to the parameters $\lambda$, $\theta$ and $g$. Recall that these parameters affect tax rates via their impact on the estimated values of $\alpha_h$ and $\alpha_S$.\(^{42}\) Table 2 reports results. Qualitatively, the responses of tax rates are as expected. For example, a high value of $\theta$ 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. 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 percent 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% and sometimes 80%, but it is only applicable if the policymaker has no ability to tax and no direct social concern for firm claimants.

\(^{42}\)Specifically, in equation (52):

\[
B := \frac{1 + g}{1 + r}; \quad D = \frac{\lambda}{1 + \lambda - B}; \quad N' := \frac{1}{D\Xi} - \frac{\xi}{1 + \xi}; \quad \Xi := \frac{1 - \theta}{1 - \theta + \frac{\theta}{\tilde{r}(1 - B)};}; \quad \mathcal{P} := -\frac{\xi}{1 + \xi}.
\]
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 proposed and if profits are pure rent to the owners of indivisible firm assets. 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 under 15% 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), then the optimal marginal tax rate would be below 30% 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 formally modeled and quantified. Potentially, profit taxes interact with financing frictions, encourage firms to realize profits outside of a tax jurisdiction and deter creation of the firm asset $S$ in the first place. An integrated analysis of CEO income and profit taxation that takes explicit account of these 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.\textsuperscript{43} Analysis of this effect requires embedding the CEO assignment model into a general equilibrium framework that explicitly incorporates workers. 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


\textsuperscript{43}In 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.


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 \((T, \tilde{U})\), then either (a) \(μ = u\), 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 \(v \in (0, \tilde{v})\), \(u(v) = u\) and (ii) for all \(v \in (\tilde{v}, 1]\), \(μ(v) = v\).

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

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 \(μ(v') < μ(v)\). We argue that by exchanging partners \(v\) and \(μ(v')\) 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 \(c(v') := w(v') - T[w(v')]\),

\[
U \left( c(v), \frac{z(v)}{h(μ(v))} \right) \geq U \left( c(v'), \frac{z(v')}{h(μ(v))} \right)
\]

and:

\[
U \left( c(v'), \frac{z(v')}{h(μ(v))} \right) \geq U \left( c(v), \frac{z(v)}{h(μ(v'))} \right).
\]

If the first of these conditions did not hold, then CEO \(μ(v)\) 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 \(μ(v')\) 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, \(μ(v')\) supplies more effective labor to \(v'\) than \(μ(v)\) supplies to \(v\). If \(μ(v')\) 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 \(μ(v')\) 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:

\[
0 \geq V(S(v), z(v')) - V(S(v), z(v)) - [w(v') - w(v)] \\
> V(S(v'), z(v')) - V(S(v'), z(v)) - [w(v') - w(v)],
\]

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 \( (\mu, z, w) \) 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: \( V_{\hat{z}}(S(v), \hat{z})h(\mu(v')) = -\frac{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v'))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v'))} (1 - T_w[\hat{w}]) \) and \( V_{\hat{z}}(S(v'), \hat{z})h(\mu(v)) = -\frac{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(\hat{v})))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v'))} (1 - T_w[\hat{w}] \) since if so the following contradiction emerges:

\[
V_{\hat{z}}(S(v), \hat{z})h(\mu(v')) = -\frac{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v'))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(\hat{v})))} (1 - T_w[\hat{w}]) \\
< -\frac{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(v)))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(\mu(\hat{v})))} (1 - T_w[\hat{w}]) \\
= V_{\hat{z}}(S(v'), \hat{z})h(\mu(v)) < V_{\hat{z}}(S(v), \hat{z})h(\mu(v')),
\]

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 rematching 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 \( (\mu, z, w) \) is an equilibrium. We conclude that \( \mu \) is increasing on \((0, \hat{v})\).

Finally, we show that for \( \hat{v} > 0 \), \( \mu \) is the identity map on \((0, \hat{v})\). Since \( \mu \) is measure-preserving and increasing, it is sufficient to show that there are no discontinuities in \( \mu \) and that \( \lim_{\mu \uparrow \mu} \mu(v) = 0 \). Suppose that \( \mu \) has a discontinuity at some \( \hat{v} \), but (without loss of generality) is continuous from the right. Then \( \mu(\hat{v}) > \mu(\hat{v}-) := \lim_{\mu \uparrow \mu} \mu(v) \) and CEOs between \( (\mu(\hat{v} -), \mu(\hat{v})) \) are unmatched. But for \( \tilde{m} \in (\mu(\hat{v} -), \mu(\hat{v})) \), \( U(\tilde{w}(\hat{v}) - T[\tilde{w}(\hat{v})], z(\hat{v})/h(\tilde{m})) > U(\tilde{w}(\hat{v}) - T[\tilde{w}(\hat{v})], z(\hat{v})/h(\mu(\hat{v}))) \geq \hat{U} \), contradicting the definition of equilibrium. Thus, \( \mu \) is continuous. By a very similar argument if \( \lim_{\mu \downarrow \mu} \mu(v) > 0 \), then there are unmatched CEOs in \((0, \lim_{\mu \downarrow \mu} \mu(v))\). 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 \( \mu \) at \( \hat{v} \). Suppose \( \hat{v} \in (0, 1) \) and let \( \gamma_n \uparrow \hat{v} \) and \( \gamma_n \downarrow \hat{v} \) (with
each $0 < \underline{v}_n < \tilde{\vartheta} < \overline{v}_n < 1$). We have:

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

As observed previously $w_n = w(\underline{v}_n)$, $c_n = w(\underline{v}_n) - T[w(\underline{v}_n)]$ and $z_n = z(\overline{v}_n)$ are bounded, decreasing sequences. Denote the limits of these sequences $(w_\infty, c_\infty, z_\infty)$. Since $\lim h(\underline{v}_n) - h(\overline{v}_n) \downarrow 0$ and $U$ is continuous, it follows that $U(c_n, z_n/h(\underline{v}_n)) - U(c_n, z_n/h(\overline{v}_n))$ 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 $\tilde{\vartheta}$ firm and CEO are indifferent about matching at the effective labor-income $(z_\infty, w_\infty)$.

Without loss of generality we select equilibria in which if $\tilde{\vartheta} > 0$, then $\mu(\tilde{\vartheta}) = \tilde{\vartheta}$. The next proposition simplifies the equilibrium conditions (6) to (10) 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 $(0, \tilde{\vartheta}]$ are such that no CEO $v \in I \cap [0, \tilde{\vartheta}]$ is made strictly better off exchanging places with CEO $v'(0, \tilde{\vartheta}]$ and accepting the terms $v'$ receives and similarly for firms, then no firm-CEO pair $(v, v') \in (0, \tilde{\vartheta}]^2$ 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 $\tilde{\vartheta}$-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 $(\tilde{\vartheta}, w, z)$ 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 $(\tilde{\vartheta}, z, w)$ satisfying the conditions described above.

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

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

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

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

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 $(v, v') \in (0, \tilde{\vartheta}]$ and $w' \in w((0, \tilde{\vartheta}])$ 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) \( U\left(w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(\hat{v})}\right) = \hat{U}\) and \( V(S(\hat{v}), z(\hat{v})) - w(\hat{v}) \geq 0 \) and (ii) for all \( w' \not\in w((0, \hat{v}]) \), \( T(w') = w' \), then \((w, z)\) defines an equilibrium at \((T, \hat{U})\). Conversely, if \((\hat{v}, w, z)\) is an equilibrium at \((T, \hat{U})\), then \((\hat{v}, w, z)\) satisfies (A.1) to (A.3), \( U\left(w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(\hat{v})}\right) \geq \hat{U}\) and \( V(S(\hat{v}), z(\hat{v})) - w(\hat{v}) \geq 0 \) with equality if \( \hat{v} \in (0, 1) \).

**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 \(v, v' \in (0, \hat{v}]\) and \( w' = w(\hat{v}) = \hat{w}(\hat{v}) \) 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:

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

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 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 \((0, \hat{v}]\), then each firm (resp. CEO) \( v \in (0, \hat{v}] \) is better off matched with CEO (resp. firm) \( v \), than matched with an alternative partner \( v' \in (0, \hat{v}] \) at an income \( w' \in w((0, \hat{v}]) \) 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 \( w' \in \mathbb{R}_+ \setminus w((0, \hat{v}]) \), then there is 100% taxation of any price outside of the range of \( w \) on \((0, \hat{v}]\). Clearly, no firm or CEO would wish to choose such an income and, hence, no firm or CEO in \((0, \hat{v}]\) 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 \( U\left(w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(\hat{v})}\right) = \hat{U}\) and \( V(S(\hat{v}), z(\hat{v})) - w(\hat{v}) = 0 \), then no firm or CEO \( v \in (0, \hat{v}] \) 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 \((\bar{v}, z, w)\) is an equilibrium at \((T, \bar{U})\), then it is immediate that it satisfies (A.1) to (A.3) and \(U\left( w(\bar{v}) - T[w(\bar{v})], \frac{z(\bar{v})}{h(\bar{v})} \right) \geq \bar{U}\) and \(V(S(\bar{v}), z(\bar{v})) - w(\bar{v}) \geq 0\). If \(\bar{v} \in (0,1)\), then \(U\left( w(\bar{v}) - T[w(\bar{v})], \frac{z(\bar{v})}{h(\bar{v})} \right) = \bar{U}\) and \(V(S(\bar{v}), z(\bar{v})) - w(\bar{v}) = 0\) 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.

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

**Lemma A.2.** If \((\bar{v}, z, w)\) is an equilibrium threshold and a pair of equilibrium effective labor and income functions at \((T, \bar{U})\), then \((z, w, c)\), with \(c(v) = w(v) - T[w(v)]\), \(v \in (0, \bar{v}]\) are non-increasing and there exists a function \(\omega : \mathbb{R} \rightarrow \mathbb{R}_+\) satisfying for each \(v \in (0, \bar{v}]\), \(\omega(z(v)) = w(v)\). In addition, equilibrium CEO utility \(\Phi\), \(\Phi(v) = U(c(v), z(v)/h(v))\) for \(v \in (0, \bar{v}]\), and firm profits \(\pi\), \(\pi(v) = V(S(v), z(v)) - w(v)\) for \(v \in (0, \bar{v}]\), 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 \(\omega\), 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 \(\pi\). \(\square\)
B Derivation of elasticities for Section 3

In this appendix, we derive 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 $\zeta(v, 1 - \tau)$ 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:

$$\left(1 - \tau\right) \frac{\partial \omega}{\partial z} hU_c\left(c, \frac{\zeta}{h}\right) + U_e\left(c, \frac{\zeta}{h}\right) = 0.$$  \hspace{1cm} (B.1)

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

$$\frac{\partial \omega}{\partial z} = DS.$$  \hspace{1cm} (B.2)

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

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

And so:

$$\frac{\zeta_v}{\zeta} = \left\{ 1 + E^u \right\} \frac{h_v}{h} + E^c \frac{S_v}{S}.$$  \hspace{1cm} (B.3)

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

$$\frac{\partial^2 \omega}{\partial z^2} = \frac{DS_v}{\zeta} \left\{ (1 + E^u) \frac{h_v}{h} + E^c \frac{S_v}{S} \right\}.$$
Totally differentiating the CEO’s first order condition (B.1) with respect to $1 - \tau$ gives:

$$\frac{\partial \zeta}{\partial (1 - \tau)} = -\frac{\partial \omega}{\partial z} hU_c + \left\{ w - w_0 + (1 - \tau) \frac{\partial \omega}{\partial (1 - \tau)} \right\} \left\{ (1 - \tau) \frac{\partial \omega}{\partial z} hU_c + U_{ec} \right\} + (1 - \tau) \frac{\partial^2 \omega}{\partial z^2} hU_c$$

$$\frac{(1 - \tau)^2 \partial^2 \omega}{\partial z^2} hU_c + \left\{ (1 - \tau) \frac{\partial \omega}{\partial z} \right\}^2 hU_{cc} + 2(1 - \tau) \frac{\partial \omega}{\partial z} U_{ec} + U_{ee}/h.$$

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

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

Combining the preceding conditions:

$$\frac{1 - \tau}{\zeta} \frac{\partial \zeta}{\partial (1 - \tau)} = -\left( \frac{1 - \tau}{\zeta} \right) DShU_c + \left\{ \frac{w - w_0}{w} + \frac{1 - \tau}{\partial (1 - \tau)} \right\} w \left\{ (1 - \tau) DShU_{cc} + U_{ec} \right\} \left\{ (1 - \tau) DS \right\}^2 hU_{cc} + 2(1 - \tau) DSU_{cc} + U_{ee}/h.$$

(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 $E = \frac{U_{cc}}{U_{cc} + U_{ce}}$. If income effects on effort supply are negative and $\frac{1 - \tau}{\omega} \frac{\partial \omega}{\partial (1 - \tau)} \leq 0$, then

$$\frac{1 - \tau}{\zeta} \frac{\partial \zeta}{\partial (1 - \tau)} \geq \left( \frac{U_c}{U_{ce}} \right) U_c + \frac{w}{DSh} \left\{ -\frac{U_{cc}}{U_{cc} + U_{ce}} \right\} U_{cc} - 2 \frac{U_{cc}}{U_{cc} + U_{ce}} U_{ee} > \left( \frac{U_c}{U_{ce}} \right) U_c + \left\{ -\frac{U_{cc}}{U_{cc} + U_{ce}} \right\} U_{cc} - 2 \frac{U_{cc}}{U_{cc} + U_{ce}} U_{ee},$$

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 $\frac{1 - \tau}{\omega} \frac{\partial \omega}{\partial (1 - \tau)}$. 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 $\frac{1 - \tau}{\zeta} \frac{\partial \zeta}{\partial (1 - \tau)}$ 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 $\Phi_0$ 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, $\pi_0$ firm $v_0$’s profit and $S_0$ its asset size (all prior to the tax change). Then:

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

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 $\Phi_0$ after the change, then $\frac{\partial \omega}{\partial (1 - \tau)}(z_0, 1 - \tau) = 0$. Furthermore,

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

Consider three cases. First $v_0 = \tilde{v} = 1$ 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 $\pi_0$. 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:

$$E_{w}(v_0) := \frac{1 - \tau}{w} \frac{d \omega}{d (1-\tau)} \bigg|_{v_0} = \frac{1 - \tau}{w_0} \frac{\partial \zeta}{\partial z} \frac{d \zeta}{d (1-\tau)}(v_0, 1-\tau) \bigg|_{v_0} = \frac{DS_0}{w_0} \xi_0,$$

where $\xi_0$ is the compensated effort elasticity of CEO $v_0$ in equilibrium. Define $v(z, 1-\tau)$, $z \in \zeta([0, 1], 1-\tau)$, to be the rank of the CEO exerting effective labor $z$ when the retention rate is $1-\tau$, i.e.

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

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

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

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 $\frac{1-\tau}{\zeta} \frac{\partial \zeta}{\partial (1-\tau)}$ (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 for $z \in \zeta([0, 1], 1-\tau)$,

$$\omega(z, 1-\tau) = \omega(z_0, 1-\tau) + \int_{z_0}^{z} DS(v(z', 1-\tau)) dz'. \quad \text{(B.9)}$$
Differentiating (B.9) with respect to $1 - \tau$ (and using $\frac{\partial \omega}{\partial (1 - \tau)}(v_0) = 0$) gives:

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

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 $\frac{1 - \tau}{z}\frac{\partial \zeta}{\partial (1 - \tau)}$ that if the uncompensated behavioral elasticity for effort is positive over an interval of effective labors and there is no bunching, then $\frac{1 - \tau}{z}\frac{\partial \zeta}{\partial (1 - \tau)}$ and $\frac{\partial \omega}{\partial (1 - \tau)}$ 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: $DS(v) = \frac{\partial \omega}{\partial \nu}(\zeta(v, 1 - \tau), 1 - \tau)$. 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.) Thus, 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:

$$\mathcal{E}_w(v) = \left(\frac{\zeta}{w} \frac{\partial \omega}{\partial \zeta} \right) \left(\frac{1 - \tau}{\zeta} \frac{\partial \zeta}{\partial (1 - \tau)} \right) + \left(\frac{1 - \tau}{w} \frac{\partial \omega}{\partial (1 - \tau)} \right), \quad (B.11)$$

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

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

The preceding formulas are simplified in the absence of income effects. Then $\frac{1 - \tau}{z\frac{dz}{d(1-\tau)}} = \mathcal{E}^c = \mathcal{E}^u = \mathcal{E} := \frac{u'}{u'_{w}}$ and, after an integration by parts,

$$-\frac{1 - \tau}{w} \frac{\partial \omega}{\partial (1 - \tau)}(v) = \frac{DSz - DS_zv_0z_0}{w} \mathcal{E} + \frac{1}{w} \int_{v_0}^{v_0} DS(v')z(v')\mathcal{E}(v')\mathcal{M}(v')dv', \quad (B.13)$$

with:

$$\mathcal{M}(v') := \left(1 + \mathcal{E}(v') + \frac{1 - \tau}{\mathcal{E}} \frac{\partial \mathcal{E}}{\partial (1 - \tau)}(v') \right) \frac{h_v}{h}(v') + \left(\mathcal{E}(v') + \frac{1 - \tau}{\mathcal{E}} \frac{\partial \mathcal{E}}{\partial (1 - \tau)}(v') \right) \frac{S_v}{S}(v')$$
In the constant elasticity case, (B.13) is further reduced to:

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

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

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

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:

$$\text{E}_w(v) = \left\{ \frac{DS_0z_0 - w_0}{w(v)} + 1 \right\} E \geq \text{E} \geq 0. \quad (B.14)$$

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

**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:

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

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

$$\text{E}_W = \text{E} \int_0^{v_0} w(v) \left\{ \frac{DS_0z_0 - w_0}{w(v)} + 1 \right\} dv = \left( \frac{R}{W} - \frac{\Delta \Pi}{W} \right) E \geq 0, \quad (B.15)$$

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

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

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:

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

52
where the inequality uses the monotonicity of $\pi$ 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 $\Pi$ with respect to $1 - \tau$ is:

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

Again this is positive if the underlying uncompensated behavioral effort elasticity.

In the quasilinear/constant elasticity case, this formula reduces to:

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

**Optimal tax formulas** The preceding “aggregate” elasticities can be substituted into the tax formula (29) 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 (29) to allow for societal concern for top CEOs. We then consider the case $\chi = 1$ in detail and derive (33) in the main text.

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

$$-\int_0^{v_0} \{\omega(z^*(v); 1 - \tau^*) - w_0\} dv$$

$$+ \int_0^{v_0} \psi(v) U_c(v) \left\{\omega(z^*(v); 1 - \tau^*) - w_0 + (1 - \tau^*) \frac{\partial \omega}{\partial (1 - \tau)} (z^*(v); 1 - \tau^*)\right\} dv$$

$$+ \tau^* \int_0^{v_0} \left\{\frac{\partial \omega}{\partial z} (z^*(v); 1 - \tau^*) \frac{\partial \zeta}{\partial (1 - \tau)} (v, 1 - \tau^*) + \frac{\partial \omega}{\partial (1 - \tau)} (z^*(v); 1 - \tau^*)\right\} dv$$

$$- \chi \int_0^{v_0} \frac{\partial \omega}{\partial (1 - \tau)} (z^*(v); 1 - \tau^*) dv = 0,$$

where $z^*(v) := \zeta(v, 1 - \tau^*)$. 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:

$$\tau^* = \frac{1}{1 + A^*_W \frac{\Pi^*_W \epsilon^*_W}{1 - \phi - (\chi - \phi) A^*_W \frac{\Pi^*_W \epsilon^*_W}{1 - \phi}}},$$

(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:

$$\tau^* = \frac{1}{1 + A^*_W \frac{\Pi^*_W \epsilon^*_W}{1 - A^*_W \frac{\Pi^*_W \epsilon^*_W}{1}}},$$

(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 (33) in the main text. After simple calculations:

\[
\frac{\Pi^*}{W^*}e_{\Pi}^* = \frac{R^*}{W^*}e_{R}^* - e_{W}^*,
\]

(B.20)

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

\[
E_{\tau}^* := 1 - \tau \frac{dR}{R} \bigg|_{1-\tau^*} = \int_{v_0}^{v_0} \frac{r^*(v)}{R^*} \zeta(v, 1-\tau^*) \frac{d\zeta(v, 1-\tau^*)}{d(1-\tau)} dv,
\]

(B.21)

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

\[
\tau^* = \frac{1}{1 + \frac{\mathcal{E}_{R}^*}{\mathcal{E}_{\Pi}^*}}.
\]

(B.22)

From the definition of \( A_{W}^* \),

\[
\frac{W^*}{A_{W}^*} = \Delta W^* = \int_{v_0}^{v_0} \{w^*(v) - w_0\} dv.
\]

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 \( v = 1 - M(w^*(v)) \),

\[
\int_{0}^{v_0} \frac{w^*(v)}{\alpha_{w}^*(v)} dv = \int_{0}^{v_0} \frac{1 - M(w^*(v))}{m(w^*(v))} dv = \int_{0}^{v_0} w^*_0(v) \nu dv = \int_{0}^{v_0} \{w^*(v) - w_0\} dv,
\]

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

\[
\frac{W^*}{A_{W}^*} = \int_{0}^{v_0} \frac{w^*(v)}{\alpha_{w}^*(v)} dv.
\]

(B.23)

Denoting the local Pareto coefficients for CEO talent and firm profit (at the optimum) by \( \alpha_h \) and \( \alpha_{\pi}^* \) respectively and using (20) in the main body of the paper gives:

\[
\frac{w}{\alpha_{w}^*} = DSz(1 + \mathcal{E}_u \nu) \frac{1}{\alpha_h} + \pi \mathcal{E}_c \frac{1}{\alpha_{\pi}^*}.
\]

(B.24)

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

\[
\frac{\Delta W^*}{R^*} = \frac{1}{R^*} \frac{W^*}{A_{W}^*} = \int_{0}^{v_0} \frac{r^*(v)}{R^*} (1 + \mathcal{E}_u^*) \frac{1}{\alpha_h} dv + \int_{0}^{v_0} \frac{\tau^* \mathcal{E}_c^*}{R^*} \frac{1}{\alpha_{\pi}^*} dv.
\]
Consequently,

\[
\frac{1}{R^*} \frac{W^*}{A^*_W} - \frac{\Pi^*}{R^*} \mathcal{E}^*_\Pi = \int_0^{v_0} \frac{r^* (1 + \mathcal{E}''^*)}{\alpha_h} \frac{1}{R^*} dv + \frac{1}{R^*} \int_0^{v_0} \pi^* \mathcal{E}^c^* \frac{1}{\alpha^*_\pi} dv
\]

\[
+ \frac{1 - \tau^*}{R^*} \int_0^{v_0} \int_{z_0}^{z^*(v)} \left\{ DS_v(v^*,1 - \tau) \frac{\partial v}{\partial (1 - \tau)} (z^*,1 - \tau^*) \right\} dz'dv.
\]

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

\[
\tau^* = \frac{1}{1 + \frac{1}{\xi} \frac{\frac{d\xi}{R^* (1 + \mathcal{E}''^*)}}{\alpha_h} dv + \frac{1}{R^*} \int_0^{v_0} \mathcal{E}^c \frac{1}{\alpha^*_\pi} dv + \frac{1}{R^*} \int_0^{v_0} \int_{z_0}^{z^*(v)} \left\{ DS_v(v^*,1 - \tau) \frac{\partial v}{\partial (1 - \tau)} (z^*,1 - \tau^*) \right\} dz'dv}
\]

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

\[
\frac{1}{R^*} \frac{W^*}{A^*_W} - \frac{\Pi^*}{R^*} \mathcal{E}^*_\Pi = (1 + \mathcal{E}) \int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv + \frac{\mathcal{E}}{R^*} \int_0^{v_0} \pi^* \frac{1}{\alpha^*_\pi} dv
\]

\[
- \frac{\mathcal{E}}{R^*} \int_0^{v_0} \left\{ DS(v) z^*(v) - DS_0 z_0 + \int_{v_0}^{v_0} w^*(v') dv' \right\} dv
\]

\[
= (1 + \mathcal{E}) \int_0^{v_0} \frac{r^*}{R^*} \frac{1}{\alpha_h} dv + \frac{\mathcal{E}}{R^*} \int_0^{v_0} \pi^* \frac{1}{\alpha^*_\pi} dv
\]

\[
- \frac{\mathcal{E}}{R^*} \int_0^{v_0} \left\{ DS(v) z^*(v) - w^*(v) - \{ DS_0 z_0 - w_0 \} \right\} dv.
\]

Then using \( \pi^*(v) = DS(v) z^*(v) - w^*(v), \pi_0 = DS_0 z_0 - w_0 \) and \( \int_0^{v_0} \pi^* dv = \int_0^{v_0} \{ \pi^*(v) - \pi_0 \} dv \), 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:

\[
\tau^* = \frac{1}{1 + \frac{\mathcal{E}}{1 + \mathcal{E}} A_h}.
\]

where \( A_h \) is defined as in the main text.
Independence of $E_W A_W$ from the marginal tax rate  We conclude this appendix by showing that the product $E_W A_W$ 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 $E_W A_W$ 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:

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

and

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

with $H(v)$ independent of marginal taxes. Hence,

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

Similarly,

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

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

C Derivation of optimal nonlinear tax formulas

When firm claimants have welfare weight $\chi$, CEOs have welfare weights $\psi$, the policymaker’s (relaxed) mechanism design problem reduces to:

$$\sup_{\Phi, z, w, \vartheta} \int_0^{\vartheta} \{\chi V(S(v), z(v)) + (1 - \chi) w(v) - C[\Phi(v), z(v)/h(v)] + \psi(v) \Phi(v)\} dv$$

$$+ \int_{\vartheta}^1 \{\psi(v) + T^0\} dv,$$  \hspace{1cm} (C.1)

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

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

$$w_v(v) = V_Z(S, z) h_v(v).$$

In the relaxed problem (C.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, $\pi$) for the relaxed problem are checked ex post in all of our numerical calculations. Problem (C.1) may be formulated as an optimal control problem in which $\Phi$, $z$ and $w$ are the state variables, $z_v$ is the control variable and $\bar{\varrho}$ is a choice variable. The Hamiltonian for this optimal control problem is:

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

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

$$p^z + p^w V_z(S, z) = 0. \tag{C.2}$$

Differentiating (C.2) with respect to $v$ gives:

$$p^z_v + p^w_v V_z(S, z) + p^w [V_{zS}(S, z)S_v + V_{zz}(S, z)z_v] = 0. \tag{C.3}$$

The optimal co-state equations are:

$$p^\Phi_v = \frac{1}{U_c} + p^\Phi U_{ec} \frac{z h_v}{U_c h} - \psi \tag{C.4}$$

$$p^w_v = - (1 - \chi) \tag{C.5}$$

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

The first order condition for $\bar{\varrho}$ is:

$$\mathcal{H}(\bar{\varrho}) - \psi(\bar{\varrho})\bar{U} - T^0 + q^V V_S(S(\bar{\varrho}), z(\bar{\varrho}))S_v(\bar{\varrho}) \geq 0, \tag{C.7}$$

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

$$p^w(\bar{\varrho}) = -q^V \quad p^z(\bar{\varrho}) = q^V V_z(S(\bar{\varrho}), z(\bar{\varrho})) \quad p^\Phi(\bar{\varrho}) = q^U. \tag{C.8}$$

There are also transversalities at $v = 0$. This is because (unlike typical optimal control problems), there are no initial conditions for $w$ and $\Phi$. 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 (C.4) gives:

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

57
while integrating (C.5) gives
\[ p^{w}(v) = -(1 - \chi)v. \] (C.10)

Combining conditions (C.3) to (C.6) and (C.10) implies:
\[
V_{z} + \frac{U_{e}}{U_{c}} h = -\frac{p^{\Phi}}{h} \left\{ \frac{U_{ee}}{U_{c}} \frac{z}{h} + U_{e} \right\} - \frac{h_{v}}{h}
+ (1 - \chi)v V_{zS}(-S_{v}).
\] (C.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 (15) imply:
\[
(1 - T_{w}[w]) V_{z} h U_{c} = -U_{e}.
\] (C.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 (C.11). If \( \chi = 1 \), then the right hand side of (C.11) equals the usual marginal informational rents term from the Mirrlees model and combining (C.11) and (C.12) and the definition of \( \alpha_{h} \) the standard formula for optimal marginal tax rates obtains:
\[
T_{w}[w] = \frac{1}{1 + \frac{\varepsilon^{e}_{c}}{\varepsilon^{e}_{u} \alpha_{h}}},
\] (C.13)

where \( \hat{p}^{\Phi} = \frac{U_{c} p^{\Phi}}{1 - F(h)} \) is the normalized co-state. More generally, using (C.11) and (C.12) and the definitions of \( \alpha_{h} \) and \( \alpha_{S} \), the optimal marginal tax rate is:
\[
T_{w}[w] = \frac{1 + (1 - \chi) \frac{\varepsilon^{e}_{c}}{\varepsilon^{e}_{u} \alpha_{h}} V_{S} \alpha_{h}}{1 + \frac{\varepsilon^{e}_{c}}{\varepsilon^{e}_{u} \alpha_{h}}}.
\] (C.14)

**Determining \( \hat{\varphi} \)** Substituting the definition of the Hamiltonian and (C.8) into (C.7) gives:
\[
-p^{\Phi}(\hat{\varphi}) U_{c}(\hat{\varphi}) \frac{z(\hat{\varphi})}{h(\hat{\varphi})} h_{v}(\hat{\varphi}) + \chi V(S(\hat{\varphi}), z(\hat{\varphi})) - T^{0}
+ (1 - \chi) w(\hat{\varphi}) - C \left[ \hat{U}_{r} \frac{z(\hat{\varphi})}{h(\hat{\varphi})} \right] + q V_{S}(S(\hat{\varphi}), z(\hat{\varphi})) S_{v}(\hat{\varphi}) \geq 0.
\] (C.15)

If \( \chi = 1 \), then the preceding expression reduces to:
\[
V(S(\hat{\varphi}), z(\hat{\varphi})) - C \left[ \hat{U}_{r} \frac{z(\hat{\varphi})}{h(\hat{\varphi})} \right] - T^{0} - p^{\Phi}(\hat{\varphi}) U_{c}(\hat{\varphi}) \frac{z(\hat{\varphi})}{h(\hat{\varphi})} h_{v}(\hat{\varphi}) + q V_{S}(S(\hat{\varphi}), z(\hat{\varphi})) S_{v}(\hat{\varphi}) \geq 0.
\] (C.16)
The term $V(S(\breve{\varsigma}), z(\breve{\varsigma})) - C [\breve{U}, z(\breve{\varsigma})/h(\breve{\varsigma})] - T^0$ captures the direct loss of social surplus from shutting down the $\breve{\varsigma}$ firm. The remaining terms capture the benefits to closing the firm down in terms of reduced rents to firms and CEOs ranked above $\breve{\varsigma}$. Note that (C.16) does not pin down $w(\breve{\varsigma})$, which may take any value in $[0, V(S(\breve{\varsigma}), z(\breve{\varsigma})))].$ When $\chi = 1$, the policymaker does not care whether the social surplus at $\breve{\varsigma}$ is realized as firm profit $V(S(\breve{\varsigma}), z(\breve{\varsigma})) - w(\breve{\varsigma})$ (perhaps to be captured by taxation on claimants) or realized as CEO income tax revenues $T[w(\breve{\varsigma})] = w(\breve{\varsigma}) - C [\breve{U}, z(\breve{\varsigma})/h(\breve{\varsigma})]$. From (C.8) and (C.10), $q^V = -p^w(\breve{\varsigma}) = (1 - \chi)\breve{\varsigma} \geq 0$. Thus, if $\chi < 1$, then $q^V > 0, V(S(\breve{\varsigma}), z(\breve{\varsigma})) - w(\breve{\varsigma}) = 0$ and (C.15) reduces to:

$$V(S(\breve{\varsigma}), z(\breve{\varsigma})) - C [\breve{U}, z(\breve{\varsigma})/h(\breve{\varsigma})] - T^0 + (1 - \chi)\breve{\varsigma}V_S(S(\breve{\varsigma}), z(\breve{\varsigma}))S_{\breve{\varsigma}}(\breve{\varsigma}) - p^\Phi(\breve{\varsigma})U_c(\breve{\varsigma}) \frac{z(\breve{\varsigma}) h_{\breve{\varsigma}}(\breve{\varsigma})}{h(\breve{\varsigma})} \geq 0.$$  

(C.17)

Inequality (C.17) is interpreted similarly to (C.16). Now, however, firm $\breve{\varsigma}$ generates no profit; everything above the consumption amount necessary to give CEO $\breve{\varsigma}$ utility $\breve{U}$ is captured by the policymaker in the form of taxation on CEO $\breve{\varsigma}$.

**D  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 (check consistency of language with our) of the distribution of talent using evidence on distribution of firm size and evidence on 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 “Robert’s Law ” (the empirical regularity implying that CEO compensation goes up by roughly 1/3 percent for every percent increase in firm size). These two facts imply a tail index of CEO talent equal to -2/3. We see our approach as complementary to the one in Gabaix and Landier (2008). First of all, Gabaix and Landier (2008) focus on the tail properties of CEO talent, we recover the entire functions mapping quantile to local Pareto coefficient for CEO talent and firm asset size. Knowledge of both these distributions is needed for the computation of the non-linear tax schedule. In addition, with respect to the model of Gabaix and Landier (2008) we allow for elastic labor and (as in Tervio (2008)) for long lasting CEO effects and adjustable capital.

**E  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:

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

Let \( \tilde{w} \) and \( \tilde{z} \) 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:

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

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:

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

where the notation makes the dependence of the functions \( \tilde{w} \) and \( \tilde{z} \) on \( \delta \) 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:

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

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:

\[ R(\delta) := \int_0^{v_2(\delta)} \{ T[\hat{\omega}(v; \delta)] + \delta \} dv + \int_{v_2(\delta)}^{v_0} \{ T[\hat{\omega}(v; \delta)] + \sqrt{\delta[\hat{\omega}(v; \delta) - w_0]} \} dv. \]  \hspace{1cm} (E.1)

**Mechanical Effect**  The mechanical effect is obtained from the terms:

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

Totally differentiating this with respect to \( \delta \) and setting \( \delta \) to zero gives:

\[ 1 - M(w_0). \]

**Behavioral effect**  Next we turn to the behavioral effect.\(^{44}\) It is obtained from the remaining terms in (E.1):

\[ \int_0^{v_2(\delta)} T[\hat{\omega}(v; \delta)] dv + \int_{v_2(\delta)}^{v_0} T[\hat{\omega}(v; \delta)] dv. \]  \hspace{1cm} (E.2)

Totally differentiating this with respect to \( \delta \) and evaluating the limit as \( \delta \) converges to 0 yields:

\[ - \frac{T_w[w_0]}{1 - T_w[w_0]} \frac{\mathcal{E} S_{w_0}^{w_0}}{w_0} m(w_0) w_0 \]

\[ + \frac{T_w[w_0]}{1 - T_w[w_0]} \frac{\mathcal{E} S_{w_0}^{w_0}}{w_0} m(w_0) w_0 \]

\[ \times \frac{1}{g(S_0) S_T w[w_0]} \int_0^{v_0} \frac{T_w[\hat{\omega}(v)]}{1 + \mathcal{E} T[\hat{\omega}(v)] S_{w}[\hat{\omega}(v)]} \exp \left\{ - \int_{v_0}^{v} \frac{S_{w}(v')}{S(v')} \mathcal{E} T[\hat{\omega}(v')] \frac{S_{w}(v')}{[\hat{\omega}(v')]} dv' \right\} dv. \]  \hspace{1cm} (E.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 (E.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 (E.3).

Adding the mechanical term as well, setting the sum to zero and rearranging

\(^{44}\)In 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.
gives the optimal marginal tax rate for $\chi = 0$,

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

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

$$\tilde{E}_w(w_0) := \frac{\varepsilon S_0z_0}{1+E[T[w_0]S_0]} \left\{ 1 - \frac{1}{g(S_0)T[w_0]} \int_0^{v_0} \frac{T[w(v)]}{1+E[T[w(v)]S_0]} \exp \left\{ -\int_0^{v_0} \frac{S(v')E[T[w(v')]S(v')}{1+E[T[w(v')]S_0]} dv' \right\} dv \right\}. \tag{E.4}$$

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

$$\tilde{E}_w(w_0) := \frac{\varepsilon S_0z_0}{1+E[T(v_0)S_0]} \left\{ 1 - \frac{1-G(S_0)T[w_0]}{g(S_0)S_0} \right\}.$$

When $\chi > 0$, the impact of the marginal tax rate change on firm profits is also relevant. The impact of such a change is obtained from:

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

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

$$m(w_0) \frac{1}{g(S_0)S_0} \frac{E S_0z_0}{1-T[w_0]} \frac{1}{1+E[T[w_0]S_0]} \left\{ \int_0^{v_0} \exp \left\{ -\int_0^{v_0} \frac{S(v')E[T[w(v')]S(v')}{1+E[T[w(v')]S_0]} dv' \right\} dv \right\}. \tag{E.6}$$

The term in (E.6) is the analogue of $\chi \frac{\Pi}{1-\chi} \tilde{E}_\Pi$ in the derivation of (29). Adding it to the other behavioral and mechanical terms leads to an optimal tax equation for the non-linear setting analogous to (29).

### F 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 weight to firm claimants ($\chi = 0$) or CEOs, then in the affine setting the optimal tax rate on top earning CEOs is given by the Diamond-Saez formula:

$$\tau^* = \frac{1}{1 + \tilde{E}_W^* A_W^*}. \tag{F.1}$$

The values of the elasticity $\tilde{E}_W^*$ 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 $E_W A_W$ 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 $E_W A_W$. We then recover the tax rate implied by formula (F.1).

**Selecting a value for $E_W$** There is limited direct evidence on $E_W$ 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% of US income earners using tax return data. In addition, they find that executives, managers, supervisors and financial professionals account for 60% of the top 0.1% 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 $E_W$ 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 $E_W$ of 0.25. Given this range of values we use multiple values for $E_W$. We proceed cautiously and use the Diamond-Saez value of 0.25 as an upper bound. We also use a more conservative value of $E_W = 0.1$.

**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 ExecuComp for the year 2011.\textsuperscript{45} 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.\textsuperscript{46}

\textsuperscript{45}We have computed estimates of $A_W$ for all years from 1947 to 2011 using data from ? 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 percent confidence intervals for the 2011 value. Details are available on request.

\textsuperscript{46}We 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.
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 $\alpha_w$ 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 $\tilde{?}$. This entails first estimating $\alpha_w$ by maximum likelihood at each fixed $\bar{w}$ and then selecting the $\bar{w}$ value (and corresponding $\alpha_w$ estimate) that maximizes a Kolmogorov-Smirnov goodness-of-fit statistic. An estimate for $\alpha_w$ of 2.1 with a 95 percent 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 $\alpha_w$ 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 $\alpha_w$ equal to 2, while Diamond and Saez (2011) and Piketty et al. (2014) assume a slightly lower value of 1.5. This suggests that the right tail of the US CEO income distribution is not very different from the tail of the general population’s income distribution.47

**Computed Optimal Tax** Together a value of $A_W$ equal to 2.1 and an elasticity of taxable income $\tilde{\epsilon}_W$ 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%. The more conservative taxable income elasticity $\tilde{\epsilon}_W = 0.1$ together with $A_W = 2.1$ implies a

---

47In Appendix F.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.
marginal income tax on incomes above the threshold of close to 76%. 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 $E_W = 0.1$ and 0.25.

Table 3: Tax Calculations

<table>
<thead>
<tr>
<th>Tail Properties</th>
<th>Elasticity</th>
<th>$w(v_0)$ (in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_W$</td>
<td>30</td>
</tr>
<tr>
<td>(I) Based on Pareto Distribution</td>
<td>$\frac{1}{3}$</td>
<td>56.4 [49.6 65.3]</td>
</tr>
<tr>
<td></td>
<td>$\frac{1}{10}$</td>
<td>76.4 [71.1 82.5]</td>
</tr>
<tr>
<td>(II) Based on $A_W$</td>
<td>$\frac{1}{3}$</td>
<td>62.6</td>
</tr>
<tr>
<td></td>
<td>$\frac{1}{10}$</td>
<td>80.7</td>
</tr>
</tbody>
</table>

Notes: Optimal tax rates as a function of the threshold value $w_0$ expressed in millions of 2014 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.

F.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 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 \( \alpha_w \). 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: 90\(^{th}\), 95\(^{th}\), 99\(^{th}\), 99.5\(^{th}\), 99.9\(^{th}\) and 99.99\(^{th}\)) 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 \( \bar{w}_i \) within the group. If the income distribution has a Pareto tail, it follows that the threshold earned income value for group \( i \), \( \bar{w}_i \), is related to \( \bar{w}_i \) according to:

\[
\bar{w}_i = \frac{\alpha_w}{\alpha_w - 1} \bar{w}_i.
\]

The Pareto assumption further implies:

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

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

\[
\alpha_w = \frac{\log \left( \frac{1-P_j}{1-P_i} \right)}{\log \left( \frac{\bar{w}_i}{\bar{w}_j} \right)}.
\]  

(F.2)

In Figure 5 we plot our estimates of \( \alpha_w \) 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 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 around the 70s up to 2000.

**G 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. Appendix G.1 describes the solution procedure. Results are given in Appendix G.2.
Figure 5: Estimates of \( \alpha_w \). 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 (F.2).

G.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

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

so that

\[
C[\Phi(v), z(v)/h(v)] = \Phi(v) + \frac{1}{1 + \frac{1}{\xi}} \left( \frac{z(v)}{h(v)} \right)^{1 + \frac{1}{\xi}}.
\] (G.1)

The optimal control problem is then:\(^{48}\)

\[
\sup_{\xi} \int_{0}^{1} \left\{ \chi DS(v)z(v) + (1 - \chi)w(v) - \Phi(v) - \frac{1}{1 + \frac{1}{\xi}} \left( \frac{z(v)}{h(v)} \right)^{1 + \frac{1}{\xi}} \right\} dv
\] (G.2)

\(^{48}\)Under the parameter values we consider \( \bar{\xi} = 1 \).
subject to:

\[
\Phi_v(v) = \left( \frac{z(v)}{h(v)} \right)^{1+\frac{1}{h(v)}} h(v) \tag{G.3}
\]

\[
w_v(v) = DS(v)z_v \tag{G.4}
\]

\[
\Phi(1) \geq \bar{U} \quad \text{and} \quad DS(1)z(1) \geq w(1). \tag{G.5}
\]

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

**Solving The Optimal Control Problem** We solve a scaled version of the optimal control problem in \((G.2)\) using the numerical solver DIDO version 7.3.7.\(^{49}\) 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 active affect the level of CEO consumption, but not the optimal effort allocation or optimal marginal taxes.

**G.2 The impact of CEO effort**

**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 point the profile becomes concave. Overall the top CEO works about 40% more than the bottom. While there is much anecdotal evidence concerning the very large number of hours worked and effort exerted by top CEOs, there has been little systematic empirical analysis of CEO effort. An exception is Bandiera et al. (2013). This paper studies a sample of CEOs from the largest Italian corporations (working under the Italian tax system). It reports that when CEOs are ranked by hours worked, then those at the 90th percentile work on average 20 hours (or roughly 50 percent more) than the CEOs at the 10th percentile. These authors also document a positive relationship between recorded CEO effort and firm labor productivity. The effort variation we find at the optimum is of a similar order of magnitude to that reported in Bandiera et al. (2013).

**Impact of CEO effort and talent on firm output** 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 output. 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. 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)

\(^{49}\)For details on the solution algorithm, refer to Ross and Fahroo (2003).
causes a reduction in output of approximately 27 percent. For the top firm, having a CEO with the same talent as the median (but exerting the higher effort as a highly talented CEO at the optimum), leads to a reduction in output of 17 per cent. Overall holding effort fixed at the median level has slightly larger effect. Reducing the effort elasticity $E$ further below $1/15$ brings the affect 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:

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

This 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. Note that $\Delta V(v)$ understates the additional cost.
to firm $v$ of inducing the median CEO to supply this effective labor since firm $v$ would have to pay the additional income tax incurred as a result of the needed increase in the CEO’s gross pay. 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.

Figure 8: Compensating Differentials.

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

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

We consider a counterfactual in which $T[w]$ is set to zero at all $w'$. 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 CEO

---

50 The 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.

51 The lump sum tax $T[w_0]$ imposed on the lowest ranked CEO is used by the policymaker to extract all profit from the smallest firm, since this firm must pay this tax to ensure that the CEO has a reservation utility above $\bar{U}$ and, hence, remain operational. 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.
at the bottom since competition maintains their 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 (G.3)). These gains amount to a 3% 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.

![Graph](image1)

(a) CEOs

![Graph](image2)

(b) Firms

Figure 9: Tax Burden Calculations.

**APPENDIX REFERENCES**


