TAXING TOP INCOMES

Laurence Ales
Christopher Sleet*

Tepper School of Business, Carnegie Mellon University

This Version: August 21, 2015

Abstract

We model high income earners as sellers of quality services in a competitive assignment framework. Sellers (high income earners) are differentiated by ability; buyers by their taste for the service. There is assortative matching of buyers and sellers. We show that conventional optimal tax formulas are modified both by a social concern for buyers and an altered mapping of the talent into the income distribution. We quantitatively apply the model to the taxation of CEOs. We find that firm value and CEO income data is consistent with a talent distribution that has a thin tail and bounded support and, given sufficient concern for non-CEO firm claimants very low and, perhaps, zero optimal marginal tax on top incomes. (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.
1 Introduction

Recent models of high earning CEOs depart from the standard framework in which heterogeneous workers sell labor to a representative firm. They assume one-to-one matching of differently talented sellers (CEOs) to differently sized buyers (firms) in a competitive assignment model. Top CEO incomes are generated by matches of highly talented CEOs with large firm sizes. 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 other hand, “local” competition amongst similar CEOs prevents any given CEO from extracting all of the surplus from firms. This assignment model has proven a useful framework for understanding the recent growth in CEO incomes and the interaction of firm and CEO attributes in shaping this growth.\footnote{See Gabaix and Landier (2008) and Terviö (2008) and the large literature stimulated by these papers.}

In this paper we consider the implications of assignment frameworks for optimal taxation and specifically for the taxation of CEOs. We make theoretical and quantitative contributions. On the theoretical front, we show that the assignment setting introduces a new dimension to policy making: social concern for buyers (in the CEO setting, non-CEO firm claimants).\footnote{For the large US corporations in our sample, the main non-CEO claimants are equity holders. Gallup reports that between 50 and 60 % of the US adult population holds equity directly or in retirement accounts, see \url{www.gallup.com/poll/1711/stock-market.aspx}. Guvenen (2009) reports that in 2002 half of US households held stocks.}

This concern imparts a moderating force into optimal tax formulas. On the quantitative side we apply our framework to the taxation of CEOs. The assignment model alters the mapping between the underlying distribution of talent and the elasticity of seller effort with respect to taxes on the one hand and the observed distribution of seller income and conventionally measured elasticities of seller income with respect to taxes on the other. Taking this into account, we show that if the policymaker attaches sufficient weight to non-CEO claimants, the optimal marginal tax on top CEO incomes may be very close to zero and far below the level suggested by other analyses of the tax treatment of top earners.

In an influential contribution, Saez (2001) derives a simple formula that maps the elasticity of top incomes with respect to tax rates and the tail coefficient of the income distribution to an optimal tax rate for the highest earners.\footnote{Given a real-valued random variable $w$ with distribution function $M$ and density $m$, the (local) Pareto coefficient at $w$ is defined as: $\frac{m(w)w}{1-M(w)} \in [0, \infty]$. Saez’s formula relates taxes to $A_w(w_0) = \int_{w_0}^{\infty} \frac{m(w)w}{1-M(w)} \text{d}w$ (with $w$ interpreted as income) which converges to the limiting Pareto coefficient as $w_0$ converges to $\infty$ and equals the Pareto coefficient if $M$ conforms to a Pareto distribution above $w_0$.} Empirical
evaluations of this formula suggest marginal tax rates on top incomes should be high - perhaps as high as 70\% or 80\%. In its basic form, this formula simply gives the marginal tax rate that maximizes tax revenues from high incomes. High income earners themselves receive zero weight and other actors in the economy are assumed to be affected by the top marginal tax rate only insofar as they benefit from tax revenues. In the competitive assignment model we develop, buyers of services are directly affected by taxes and to the extent that they are valued a motive for moderating optimal taxes is introduced. In particular, the Saez formula expressed in terms of conventionally measured elasticities of seller income holds only as a limiting case in which buyers are not socially valued. Otherwise the optimal marginal tax rate on the highest earners is depressed downwards.

Optimal tax formulas may alternatively be expressed in terms of structural parameters: the underlying distributions of seller talent and buyer taste and the behavioral seller effort elasticity. Cast in these terms the conventional Mirrleesian formula emerges as a limiting case when buyers are as socially valued as tax receipts. Thus, conventional tax formulas in terms of income elasticities and the income distribution hold only when buyers are not valued at all and conventional formulas in terms of the talent distribution and effort elasticities hold only when buyers are sufficiently valued. In addition, relative to standard models, the assignment framework modifies the mapping between the underlying talent distribution and effort elasticities and the induced income distribution and income elasticities.

These theoretical results have important implications for the quantitative assessment of optimal taxes. Estimates of the tail Pareto coefficient for incomes and the elasticity of taxable income with respect to marginal taxes maybe recovered directly from the data. Provided these values remain valid at the optimum, they permit direct calculation of the tax rate on top earners implied by the Saez formula. However, our theory implies that this formula is only valid if buyers are not valued at all. On the other hand, if buyers and tax receipts are equally valued the conventional formula in terms of primitives is reinstated, but the recovery of the talent distribution from the data is modified. We take these insights to the data and calculate estimates of optimal taxes for high earning CEOs. In the context of our model, sellers correspond to CEOs and buyers to (the non-CEO claimants upon)

We call $A_w$ the tail coefficient (function) for $w$.

\footnote{CEOs comprise only a fraction of top income earners. However, the firm-CEO relationship is naturally modeled as an a competitive assignment one, see Gabaix and Landier (2008) and Terviö (2008). In addition, we have good quality (non top coded) data on the incomes of CEOs. We also believe that our broad quantitative insights carry over to other “superstar” buyer-seller relationships, see Ales and Sleet (2015).}
firms. To better relate buyers in our theory to firms in the data we implement a version of the procedure outlined by Terviö (2008) that accounts for adjustable capital and permits CEO effective labor to have a long-lived effect on firm revenues. We calculate optimal income tax rates on top earning CEOs for two benchmark cases. In the first, the policymaker places zero weight on firm claimants and in the second weights them equally with tax revenues. We view these as natural benchmarks. To the extent that economic profits are dispersed more widely than to those on very high incomes, then the zero weight case seems too harsh; to the extent that taxes are directed to the most needy, the equal weight case is arguably too generous. In the first case, the Saez income-based formula for optimal marginal tax rates is applicable. We estimate the relevant Pareto coefficient by fitting a Pareto tail to CEO compensation data in Execucomp and take a conservative stance on the elasticity of CEO incomes with respect to taxes. We recover high values for optimal taxes of up to 80%. In the second case, the formula expressed in terms of the Pareto coefficient of the talent distribution and effort elasticities is valid. We use the model to disentangle the CEO talent distribution from the CEO income and firm value distributions. We find evidence that the talent distribution has a thin tail at the top and bounded support. Even with very conservative values for the effort elasticity we obtain very low and possibly zero optimal marginal taxes across a range of top incomes.

In the remainder of the paper we broaden the analysis and develop non-linear optimal tax formulas for assignment economies. Again, when no weight is placed on buyers conventional formulas in terms of the income Pareto coefficient and income elasticity emerge, whereas when buyers and taxes are equally weighted conventional formulas in terms of the talent Pareto coefficient and effort elasticity emerge. The derivations of these formulas sharpen intuition. In particular, we show that an increase in marginal taxes over a small interval of incomes depresses the effort and incomes of sellers populating that interval, but raises the incomes of more talented sellers. This second effect mitigates the first, raising tax revenues, but depressing the payoffs of buyers. It creates a motive for setting higher marginal taxes when buyers are not valued and lower when they are. We use our (nonlinear) optimal tax formulas to quantitatively characterize optimal taxes on CEOs across

5As noted in footnote 2, evidence suggests that about about 50% of US households own stock directly or indirectly through retirement accounts.

6Diamond and Saez (2011) (p. 11, footnote 9) point out that a bounded support is consistent with a thick tail (small Pareto coefficient) up to the income of the highest paid earner. Our evidence suggests this is true for CEO incomes, but not for CEO talent. In the latter case, the Pareto coefficient is large across a wide range of incomes.
a range of high incomes and firm claimant weight values. When taxes and firm claimants are equally weighted, optimal marginal tax rates decline from around 35% at an income of $10 million to 5% at an income of $50 million. At higher incomes they are even closer to zero. When firm claimants receive lower weight optimal marginal tax rates show a similar pattern of decay, but are higher and for relative welfare weights of 0.5 or 0 much higher.

The remainder of the paper proceeds as follows. Following a brief literature review, Section 2 provides our baseline model of taxation in an assignment economy. It provides initial characterization and a formulation of equilibrium suitable for our subsequent tax analysis. Section 3 derives a formula for the optimal affine tax on top incomes and compares it to the well known related formula of Saez (2001) and elaborations of that formula in Piketty et al. (2014). Section 4 uses this formula and data on CEO compensation and firm values to recover estimates of optimal income taxes on top CEOs for both the case in which firm claimants are and the case in which they are not socially valued. Section 5 theoretically characterizes the fully optimal non-linear tax, while Section 6 calculates the optimal nonlinear tax function for CEOs for a range of incomes and social weights on firm claimants. Section 7 concludes. Appendices contain proofs and additional details.

Related literature Our paper contributes to a literature in normative public finance that relaxes the traditional assumption of exogenously given wages. Although this literature goes back to Stiglitz (1982), there has been revived recent interest. Rothschild and Scheuer (2013) extend the theoretical implications of Stiglitz (1982) to a rich assignment setting. Ales et al. (2015) explore the implications for policy of technical change in such a setting. Rothschild and Scheuer (2014) further extend this framework by incorporating externalities at the level of occupations. They, hence, introduce a motive for corrective Pigouvian taxation. All of these papers feature the matching of populations of workers to occupations. Although, diminishing returns at the level of the occupation imply that the wage distribution is influenced by the collective behavior of workers (and, hence, by tax policy), an individual worker’s choice of effort cannot affect her wage. A worker is free to work as much or as little as she wants in the occupation of her choice at a given wage. Moreover, workers receive the entire surplus that they create. These

---

7See also Lockwood et al. (2014).
8In this model, 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.
features contrast with our model in which a buyer matches with a single seller, a
seller appropriates only part of the surplus she creates and a unilateral decision to
work less by the seller requires rematching with a different, less demanding buyer.

Ales et al. (2015) consider optimal taxation in a Rosen (1982) span of control set-
ting. They identify top earners with managers who match with and control teams of
workers. In contrast to our setting, firms (buyers of managerial talent) have no
exogenously given factors and all variations in firm size are attributable to vari-
tions 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 as-
suming that “firms are differentiated by important indivisible characteristics that
cannot be easily shuffled among(st them)”. Our results also extend beyond man-
gerational taxation problems to other superstar settings. However, our quantitative
section follows the approach of Ales et al. (2015) in using firm-level data to help pin
down key parameters necessary for optimal taxation of managers.

Piketty et al. (2014) consider a different optimal tax model in which top earners
are identified with managers. Their model analyses an imperfectly competitive set-
ing in which managers and workers bargain over compensation. In this 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. In an alternative direction, Stantcheva (2014) provides
a rich model that combines informational frictions between workers and firms as
well as between workers and the government. We abstract from informational fric-
tions between buyers and sellers, but enrich the model with heterogeneity on the
side of buyers and an assignment structure.

By far the most closely related paper to ours is the contemporaneous work of
Scheuer and Werning (2015). These authors develop the implications for optimal
taxation of an assignment setting similar to ours. Their model is equivalent to one
in which buyers (firms in their case) receive the same social weight as tax rev-

\textsuperscript{9}In their model all revenues generated by buyer-firms are returned to seller-workers modulo
taxes. The policy problem they consider is dual to one in which buyers and tax revenues receive
equal social weight.
Pareto optimality (with respect to different combinations of weights across sellers). These formulas are similar to and consistent with the ones we provide. In particular, in this setting with equal buyer-tax revenue weighting, they show that the test condition for Pareto optimality expressed in terms of the talent distribution is the same as that obtained in the conventional case without assignment. Moreover, they show that the formula when expressed in terms of the income distribution and seller elasticities of income is also the same provided the latter elasticities are defined appropriately (and differently from the conventional definition in many empirical public finance papers). Collectively, they cast these results as neutrality propositions. This is an interesting alternative perspective to ours. In contrast, we focus on the calculation of optimal taxes (under different weightings of buyers and tax revenues) rather than the Pareto testing of existing tax schedules. Our calculations imply that under reasonable social criteria (that place no weight on CEOs themselves) taxes may be very much lower than conventional analyses suggest.\footnote{These conventional analyses do not explicitly consider the weighting of buyers, but this point is moot in their analyses since workers (sellers) capture the entire pre-tax surplus from trade.}

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.\footnote{Terviö (2008) in particular elaborates on the nature of these assets.} 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.

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 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 sellers sell more and at a higher price to more desiring buyers. 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 buyers and sellers

We augment an assignment game of buyers and sellers with seller effort and taxes on seller incomes. While our focus in later quantitative sections will be upon the taxation of CEO’s, our use in this section of the terms ‘buyer’ and ‘seller’ emphasizes the more general nature of our results.

Buyers and Sellers  A population of buyers is described by a Lebesgue measure on the interval $I$, with $I = [0, 1]$ or $I = (0, 1]$. Let $S : I \rightarrow \mathbb{R}_+$ give the buyer’s taste for services with $S$ a smooth and decreasing function. Then $v \in I$ provides a ranking of buyers by taste with buyer $v = 1$ having the weakest taste for the service. In addition, let $G$, with $G(s) := 1 - S^{-1}(s)$, denote the distribution of buyers across the taste parameter. Each buyer matches with and purchases services from a single seller. If the $v$-th buyer purchases $z^b$ units of services from a seller for a price of $w^b$, then the buyer earns a payoff of:

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

where $V : \mathbb{R}^2_+ \rightarrow \mathbb{R}$ is assumed to be super-modular, increasing in both arguments and continuously differentiable and concave in $z^b$.

The population of sellers is also described by a Lebesgue measure on the interval $I$. Let $h : I \rightarrow \mathbb{R}_+$ give the talent of each seller with $h$ also a smooth and decreasing function. Thus, in the context of sellers, $v$ is a talent ranking. Let $F$, with $F(n) := 1 - h^{-1}(n)$, denote the distribution of sellers across talent. The amount of services

---

12When $I = [0, 1]$ buyers’ taste for service and sellers’ talent $h$ (to be defined below) are bounded above. Allowing for the possibility that $I = (0, 1]$ enables us to accommodate the case in which either buyers’ desire or seller’s talent is unbounded above, i.e. $\lim_{v \downarrow 0} S(v) = \infty$ or $\lim_{v \downarrow 0} h(v) = \infty$. 


supplied by a seller is a combination of talent and effort:

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

We assume that a seller can sell to only one buyer in a period.\(^\text{13}\) This is a natural assumption when sellers are interpreted as CEOs selling effective labor to firms. Under this assumption, the price a seller charges to its single customer and the seller’s income are identical and we use the expressions price and (seller) income interchangeably.

The utility of a seller over consumption \( c \) and effort \( e \) is given by \( U : \mathbb{R}_+ \times [0, \bar{e}] \to \mathbb{R} \), 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: \( -\frac{U_e(c, z^s/h)}{hU(c, z^s/h)} \) is decreasing in \( h \).

A seller must pay a tax \( T : \mathbb{R}_+ \to \mathbb{R} \), \( T(w) \in (-\infty, w] \), on her earnings. If a seller sells her services for a price of \( w^s \), her after-tax income and, hence, consumption is: \( c(w^s) = w^s - T[w^s] \). The \( v \)-th seller’s effective utility if she sells \( z^s \) for a price of \( w^s \) (and, hence, earns income \( w^s \)) is:

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

All sellers have an outside utility option of \( \bar{U} > U(0,0) \).

**Interpretation: Sellers as CEOs.** In this scenario, services correspond to effective labor, firms correspond to buyers and CEOs to sellers. Firms are differentiated by the size of their productive, non-transferable and indivisible assets with \( S(v) \) the assets of the \( v \)-th ranked firm. 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.\(^\text{14}\)

Adjustable capital is excluded from \( S \) (but we extend the model to include this together with long-lived effects of CEO effective labor in our later quantitative section).\(^\text{15}\) Firm surplus is a multiplicative function of assets and managerial effective labor: \( V(S, z^b) = DSz^b \).\(^\text{16}\) Thus, a larger stock of firm assets underpins a greater “taste” for CEO effective labor services. The \( v \)-th

---

\(^\text{13}\)See Ales and Sleet (2015) for the generalization of this assumption.

\(^\text{14}\)See Terviö (2008) and the references therein for additional discussion.

\(^\text{15}\)Despite the exclusion of adjustable capital, we refer to \( S(v) \) as firm \( v \)'s *assets* adding the qualifier "indivisible" when needed.

\(^\text{16}\)Firm surplus is to be understood as the resources available to firm owners and CEOs after the optimal choice of other adjustable inputs and after payments to these inputs have been made.
firm’s surplus net of the salary $w^b$ paid to the CEO it hires is:

$$DS(v)z^b - w^b.$$  

CEOs derive utility from consumption and disutility from effort and have preferences over pre-tax incomes and effective labor of the form (2).

**The Market Assignment Game** Returning to the general setting, given $T$ and $\bar{U}$, buyers and sellers play an assignment game. As a precursor to later optimal tax results, we formalize this game and characterize its equilibrium.

Let $w : I \rightarrow \mathbb{R}_+$ give the price paid by each buyer, $m : I \rightarrow I \cup \{u\}$ the match partner of each buyer (with $m(v) = u$ indicating that $v$ is unmatched) and $z : I \rightarrow \mathbb{R}_+$ the quantity of service purchased by each buyer. The functions $w$, $m$ and $z$ are assumed to be Lebesgue measurable.\(^\text{17}\) In addition, the match function $m$ is assumed to be measure-preserving, i.e. for all Lebesgue measurable sets $B \subset m^{-1}(I)$, $\mu[m(B)] = \mu(B)$, where $\mu$ denotes Lebesgue measure. This captures the one-to-one matching of buyers and sellers. Definition 1 defines an equilibrium for the assignment game. The definition requires that no buyer or seller can be made better off by unilaterally leaving the assignment market and that no buyer-seller pair can be made better off by (if necessary dissolving their current match or leaving their current unmatched state and) matching together and choosing some price-service combination.

**Definition 1.** A triple $(w, m, z)$ is an equilibrium of an assignment game at $(T, \bar{U})$, if:

1. **Participation:** for all $v \in m^{-1}(I)$,

   $$V(S(v), z(v)) - w(v) \geq 0 \quad \text{and} \quad U \left( w(v) - T[w(v)], \frac{z(v)}{h(m(v))} \right) \geq \bar{U}. \quad (3)$$

2. **Stability:** for all buyer-seller pairs $(v, v')$ there is no $w'$ and $z'$ such that either

   $$V(S(v), z') - w' \geq 0 \quad \text{if } m(v) = u \quad (4)$$

   or

   $$V(S(v), z') - w' \geq V(S(v), z(v)) - w(v) \quad \text{if } m(v) \in I \quad (5)$$

\(^{17}\)More precisely, $m$ is assumed $\mathcal{L}_I / \mathcal{B}_u$-measurable, where $\mathcal{L}_I$ is the Lebesgue sigma-algebra on $I$ and $\mathcal{B}_u$ is the augmentation of $\mathcal{B}_I$, the Borel sigma algebra on $I$, with sets of the form $A \cup \{u\}$, $A \in \mathcal{B}_I$.  

10
and either
\[ U \left( w' - T[w'], \frac{z'}{h(v')} \right) \geq \bar{U} \quad \text{if } v' \notin m(I) \] (6)
or
\[ U \left( w' - T[w'], \frac{z'}{h(v')} \right) \geq U \left( w(\hat{v}) - T[w(\hat{v})], \frac{z(\hat{v})}{h(v')} \right) \quad \text{if } m(\hat{v}) = v' \] (7)
with at least one of the applicable inequalities above strict.

3. **No rents to the least talented seller:** if \( 1 \in m(I) \), then:
\[ U \left( w(m^{-1}(1)) - T[w(m^{-1}(1))], \frac{z(m^{-1}(1))}{h(1)} \right) = \bar{U}. \] (8)

Note that an implication of the assumption \( \bar{U} > U(0,0) \) is that any buyer-seller match involves the trade of positive amount of service at a positive price: there are no passive matches in which nothing is done.\(^{18}\) Note also that this definition implies that any matched buyer-seller pair attains a bilateral (after-tax) Pareto efficient price-service combination. The third component of the definition asserts that there are no rents (above the outside option) to the least talented seller. This may be justified informally by assuming that the \( I \) matched sellers are the most talented members of a population of strictly greater measure and that if the least talented matched seller obtained a payoff in excess of \( \bar{U} \), then a slightly less talented unmatched seller could offer the same service at a price slightly below \( w(m^{-1}(1)) \) to buyer \( m^{-1}(1) \) and make both this buyer and herself better off. The no rents restriction removes a potential source of indeterminacy from the analysis since otherwise the allocation of the least talented seller and her partner is not determined.

We now give a proposition that characterizes equilibria. It shows that there is assortative matching of buyers and sellers with all buyers below a \( \hat{v} \) threshold matching their \( v \)-ranked seller counterparts and gives simple participation and incentive constraints that must be satisfied in equilibrium. Conversely, it shows that if a threshold \( \hat{v} \) and price and service schedules \( w \) and \( z \) are consistent with the simple participation and incentive constraints and if the tax function sufficiently penalizes choices outside of the ranges of \( w \) and \( z \), then \((\hat{v}, w, z)\) are part of an

\[ ^{18} \text{Sellers must receive a strictly positive price to match with a buyer, become a seller and give up their outside option. Since buyers must earn non-negative revenue, they must contract for a positive amount of effort from a seller.} \]
equilibrium. Thus, the stability conditions on buyers and sellers in Definition 1 are decoupled and re-expressed as separate buyer and seller incentive conditions. This is useful for our subsequent tax analysis. 

**Proposition 1.** If \((m, w, z)\) is an equilibrium, then either (a) \(m = u\), no buyer produces and all sellers take their outside option or (b) there is a \(\bar{v} \in I\) such that (i) for all \(v \in (\bar{v}, 1]\), \(m(v) = u\) and (ii) for all \(v \in I \cap [0, \bar{v})\), \(m(v) = v\). Moreover, \(w\) and \(z\) satisfy the participation conditions, for all \(v \in I \cap [0, \bar{v}]\),

\[
U \left( w(v) - T[w(v)], \frac{z(v)}{h(v)} \right) \geq \bar{U} \quad \text{and} \quad V(S(v), z(v)) - w(v) \geq 0, \quad (9)
\]

and the incentive conditions, for all \(v, v' \in I \cap [0, \bar{v}]\).

\[
U \left( w(v) - T[w(v)], \frac{z(v)}{h(v)} \right) \geq U \left( w(v') - T[w(v')], \frac{z(v')}{h(v)} \right), \quad (10)
\]

and

\[
V(S(v), z(v)) - w(v) \geq V(S(v), z(v')) - w(v'). \quad (11)
\]

Conversely, if \(\bar{v}\), \(w\) and \(z\) are such that (i) for all \(v \in I \cap [0, \bar{v}]\), \((9)\) to \((11)\) hold, (ii) \(U \left( w(\bar{v}) - T[w(\bar{v})], \frac{z(\bar{v})}{h(\bar{v})} \right) = \bar{U} \) and \(V(S(\bar{v}), z(\bar{v})) - w(\bar{v}) \geq 0\) and (iii) for all \(w' \notin w(I \cap [0, \bar{v}]), T[w'] = w'\), then \((m, w, z)\) with \(m\) such that for all \(v \in (\bar{v}, 1]\), \(m(v) = u\), and for all \(v \in I \cap [0, \bar{v}]\), \(m(v) = v\), is an equilibrium.

**Proof.** See Appendix A. \(\square\)

By standard arguments (see Lemma A.2 in Appendix A) equilibrium \(w\) and \(z\) are non-increasing on \(I \cap [0, \bar{v}]\) as are seller consumption \(c(v) = w(v) - T[w(v)]\) and seller and buyer payoffs:

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

If \(T\) is differentiable at \(w(v)\) and \(w\) and \(z\) are differentiable at \(v \in I \cap [0, \bar{v}]\) then \((10)\) implies the seller’s first order condition:

\[
U_c \left( w(v) - T[w(v)], \frac{z(v)}{h(v)} \right) w_v(v) \{1 - T[w(v)]\} + U_c \left( w(v) - T[w(v)], \frac{z(v)}{h(v)} \right) \frac{z_v(v)}{h(v)} = 0. \quad (12)
\]

Similarly, \((11)\) implies the buyer’s first order condition:

\[
V_z(S(v), z(v))z_v(v) - w_v(v) = 0. \quad (13)
\]
Combining (12) and (13), totally differentiating with respect to $v$ and denoting the compensated and uncompensated effort elasticities by $E_c$ and $E_u$ respectively, gives the following useful expression for equilibrium price growth across $v$:

$$\left( - \frac{w_v}{w} \right) = \frac{V_z}{w} \left( 1 + E^u \right) \left( - \frac{h_v}{h} \right) + E_c \frac{V_z S}{V_z} \left( - \frac{S_v}{S} \right). \tag{14}$$

Expression (14) relates price (and, hence, seller income) growth $-w_v/w$ to talent $-h_v/h$ and buyer taste $-S_v/S$ growth (across rank). These last two variables contribute to price growth directly and also indirectly through the incentives for greater effort that they create. In much of the paper we consider the special case in which seller preferences are quasilinear-constant elasticity, $U(c, z/h) = c - \Psi(z/h)$, with $\Psi(e) = \frac{e^{1+\varepsilon}}{1+\varepsilon}$, and buyer preferences are multiplicative $V(S, z) = D S z$, with $D$ a parameter. We call this the quasilinear/multiplicative case. In this case, if the tax function $T$ is locally linear (so that $T_{ww} = 0$), equation (14) and the envelope condition for buyers:

$$\pi_v(v) = D S_v(v) z(v),$$

imply a relationship between the (reciprocals of the) local Pareto coefficients of the price, talent and buyer payoff distributions. In particular, recalling the definition of the talent distribution $F$ and letting $M$ and $L$ denote the distributions of seller income and buyer payoff (with densities $m$ and $l$) respectively, we have:

$$\left(1 + \mathcal{E}\right) \frac{1 - F(h)}{hf(h)} = \frac{w}{DSz} \frac{1 - M(w)}{wm(w)} - \mathcal{E} \frac{\pi}{DSz} \frac{1 - L(\pi)}{\pi l(\pi)} \tag{15}.$$ 

Smaller values for the Pareto coefficient in the right tail of a distribution indicate a thicker or fatter tail. Thus, (15) relates the tail thickness of the (typically unobservable) seller talent distribution to those of the (typically observable) price (or seller income) and buyer payoff distributions. Notice, in particular that a fat seller income tail need not translate into a fat talent tail if sellers in the tail capture a small share of the surplus. In this case $\frac{w}{DSz}$ is small and (15) implies that the value

\[19]\text{And note that the local Pareto coefficient of the talent distribution is given by } \frac{hf(h)}{1-F(h)} \text{ and its reciprocal by } \frac{1-F(h)}{hf(h)}. \text{ Similarly, the Pareto coefficients of the price and buyer payoff distributions are, respectively, } \frac{wm(h)}{1-M(w)} \text{ and } \frac{nl(\pi)}{1-L(\pi)} \text{ and their reciprocals are } \frac{1-M(w)}{wm(w)} \text{ and } \frac{1-L(\pi)}{nl(\pi)}.

\[20]\text{In probability theory, 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.}
of $\frac{1-F(h)}{hf(h)}$ is reduced relative to that of $\frac{1-M(w)}{wm(w)}$ both because $\frac{1-M(w)}{wm(w)}$ receives a small weight and because the depressing term $\frac{1-L(\pi)}{n(\pi)}$ receives a large weight. Pareto coefficients enter into optimal tax equations. Thus, the relationship (15) shapes the determination of optimal taxes; we use it to build intuition throughout our theoretical and quantitative analysis.

3 Optimal affine taxation of sellers

Before proceeding to a derivation of optimal nonlinear taxes, we first calculate the optimal affine tax on top sellers. This calculation is simple and highlights the role of social concern for buyers in shaping and modifying optimal tax formulas.

We assume that the policymaker is restricted to tax functions that are affine above some price $w_0$:

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

The policymaker is assumed to maximize a weighted sum of tax revenues (inclusive of tax revenue $T\bar{U}$ from unmatched sellers) and buyer payoffs:

$$T\bar{U}[1 - \vartheta] + \int_0^\vartheta T[w(v)]dv + \chi \int_0^\vartheta \pi(v)dv.$$

In (17), $\chi$ is the relative welfare weight on buyers and a zero welfare weight is placed on sellers. The latter assumption parallels that made by Saez (2001) and Diamond and Saez (2011) where high income earners receive zero weight and for the purposes of comparability we maintain it here.

We focus upon the determination of the marginal tax rate $\tau$ given $w_0$ and $T[w_0]$. Let $\tau^*$ denote the optimal value of $\tau$ and let $w^*$, $z^*$ and $\pi^*$ denote the corresponding equilibrium price, service and buyer payoff functions. For concreteness, we assume that $w_0 = w^*(v_0)$ for some $v_0 \in I \cap [0, \vartheta]$. By making $w_0$ large ($v_0$ small), we can then characterize optimal taxes on the highest paid incomes. We assume throughout that the choice of $\tau$ does not affect the taxes paid by other workers (i.e. those that do not sell their services in the assignment market) either because the tax is specific to sellers of (high end) services or because other workers earn incomes below $w_0$ and, hence, justify the omission of these agents from the societal objective (17). Finally
we assume that sellers’ outside options are unaffected by the tax function.\textsuperscript{21}

Perturbations of the marginal tax rate around $\tau^*$ potentially modify the entire schedule of equilibrium prices and service amounts. Let $\bar{w}(\cdot; \delta)$ and $\bar{z}(\cdot; \delta)$ denote equilibrium price and service schedules occurring after a perturbation of the marginal tax rate on sellers from $\tau^*$ to $\tau^* + \delta$ (with $\bar{w}(v, 0) = w^*(v)$ and $\bar{z}(v, 0) = z^*(v)$). The government’s problem may be expressed as a maximization over (small) perturbations $\delta$ around $\tau^*$:

$$
\sup_{\delta \in \Delta} \int_0^{v_0} \left\{ T[w_0] + (\tau^* + \delta)(\bar{w}(v; \delta) - w_0) \right\} dv + \chi \int_0^{v_0} \left\{ V(S(v), \bar{z}(v; \delta)) - \bar{w}(v; \delta) \right\} dv,
$$

(18)

where $\Delta$ denotes an open interval around 0. In Saez (2001) the optimal linear tax at the top is chosen to maximize tax revenues only and, hence, trades off the mechanical revenue raising effect of a higher tax rate against the offsetting behavioral effect on sellers (workers). In (18) the suppressing effect of the tax rate on buyer’s payoffs creates an added cost of taxation. Let $W = \int_0^{v_0} w^*(v) dv$ denote the total income of sellers at the optimum and let $Q = \int_0^{v_0} V(S(v), z^*(v)) dv$ and $\Pi = \int_0^{v_0} \pi^*(v) dv = Q - W$, denote, respectively, the total value of services received by buyers and the total (net of price) payoffs of buyers at the optimum. Define the elasticities:

$$
E_w = \frac{1 - \tau}{W} \frac{\partial W}{\partial (1 - \tau)} \quad \text{and} \quad E_\pi = \frac{1 - \tau}{\Pi} \frac{\partial \Pi}{\partial (1 - \tau)}.
$$

(19)

Both elasticities incorporate adjustment in the equilibrium price and service schedules. To the extent that high income earners are identified with sellers, $E_w$ is the model counterpart of the elasticities of taxable income with respect to the marginal retention rate $1 - \tau$ for such earners measured in the empirical public finance literature.\textsuperscript{22} Let $\Delta W := W - w_0$ and define:

$$
A_w := \frac{W}{\Delta W}.
$$

We will refer to $A_w$ as the \textit{tail coefficient} (of seller income). If the right tail of the distribution of seller incomes above $w^*(v_0)$ is Pareto with parameter $\alpha_w$, then $A_w = \alpha_w$. More generally, making the dependence of $A_w$ on $w_0$ explicit, $\lim_{w_0 \uparrow w^*(0)} A_w(w_0) = \lim_{w_0 \uparrow w^*(0)} \frac{m[w_0]w}{1 - M[w_0]}$, i.e. the limit of $A_w(w_0)$ equals the limiting Pareto coefficient of the seller income distribution. Differentiating (18) with respect to $\tau$, using the pre-

\textsuperscript{21}These assumptions can be relaxed. We make them here to highlight our main results.

\textsuperscript{22}See Saez et al. (2012) for a survey.
ceeding definitions and rearranging gives the following expression for the optimal marginal tax:

\[ \tau^* = \frac{1}{1 + A_w \frac{\varepsilon_w + \chi \Pi \varepsilon_\pi}{1 - \chi A_w \Pi \varepsilon_\pi}}. \] (20)

In a standard labor market setting, Saez (2001) derives the related formula (re-expressed in our notation):

\[ \tau_{\text{Saez}} = \frac{1}{1 + A_w \varepsilon_w}. \] (21)

The logic behind (20) extends that behind (21) to include the impact of taxes on buyer payoffs. In particular, the standard elasticity \( \varepsilon_w \) in the classic Saez formula (21) is replaced by \( \frac{\varepsilon_w + \chi \Pi \varepsilon_\pi}{1 - \chi A_w \Pi \varepsilon_\pi} \) in (20). This last term equals \( \varepsilon_w \) if either buyers receive no social weight (\( \chi = 0 \)) or if sellers receive the entire production surplus (\( \Pi \) and, hence, \( \varepsilon_\pi = 0 \)) as in Saez (2001) and most other optimal tax analyses. In these cases the standard Saez formula is valid. Otherwise, if \( \chi > 0 \) and \( \varepsilon_\pi > 0 \) (higher marginal taxes depress buyer surplus), then marginal taxes are reduced relative to those prescribed by the standard formula.\(^{23}\)

**Quasi-linear seller and multiplicative buyer objectives** In our setting, particularly sharp results are available in the quasi-linear/multiplicative setting defined previously.\(^{24}\) In this setting, let \( \varepsilon = \frac{\psi}{\psi_{ee}} = \frac{1}{\varepsilon} \) denote the constant Frisch labor supply elasticity. Under these assumptions simple manipulations (given in Appendix B) lead to the following expressions for the elasticities:

\[ \varepsilon_\pi = \frac{\Delta \Pi}{\Pi} \varepsilon \] (22a)

and:

\[ \varepsilon_w = \left( \frac{Q}{W} - \frac{\Delta \Pi}{W} \right) \varepsilon = \frac{Q}{W} \varepsilon - \frac{\Pi}{W} \varepsilon_\pi, \] (22b)

\(^{23}\)We emphasize that the modification of the Saez formula does not stem from any externality. The assignment model is competitive and the outcome constrained efficient.

\(^{24}\)We generalize to non-quasilinear seller and non-multiplicative buyer preferences in the later non-linear tax setting. Quasilinear-constant effort elasticity preferences purge income effects and allow us to give the simple (multiplicatively separable) form of the income elasticity in (22b).
where $\Delta \Pi = \Pi - \pi(v_0)$. Given the non-negativity and monotonicity of equilibrium buyer payoffs $\pi$ (see Lemma A.2 in the appendix), we have that $\Delta \Pi := \Delta Q - \Delta W > 0$. Then $\varepsilon_\pi > 0$ and for $\chi > 0$,

$$
\tau^* = \frac{1}{1 + A_w \hat{\varepsilon}_w^{-1}} \frac{1}{1 + A_w \hat{\varepsilon}_w^{-1}} < \frac{1}{1 + A_w \hat{\varepsilon}_w^{-1}}.
$$

An increase in the marginal tax $\tau$ deters seller effort. However, the $v_0$-th ranked buyer’s and seller’s payoffs are unaffected with a reduction in price $w(v_0)$ occurring that exactly equals the reduction in the gross payoff of the $v_0$-th ranked buyer $DS(v_0)z(v_0)$. More talented sellers see a proportional reduction in their “talent premium” $w(v) - w(v_0)$ (commensurate with the relative reduction in their effort). However, since this talent premium captures only part of the gross payoff premium $DS(v)z(v) - DS(v_0)z(v_0)$ accruing to more desiring buyers and since these latter premia are also proportionately reduced by taxation, (net of price) buyer payoffs are decreased by higher marginal taxes. Since this is socially costly when $\chi > 0$ a force for lower optimal marginal tax rates is introduced.

**A restatement of the optimal tax formula for the $\chi = 1$ case.** The formulas (22a) and (22b) relate the elasticities $\varepsilon_\pi$ and $\varepsilon_w$ to the (reciprocal of the) tail coefficient of buyer payoffs $\frac{1}{A_\pi} = \frac{\Delta \Pi}{\Pi}$. Substituting these expressions into the optimal tax formula and assuming $\chi = 1$ and, hence, that the policymaker values tax revenues and buyer payoffs equally, gives:

$$
\tau^* = \frac{1}{1 + \hat{\varepsilon}_w^{-1} \frac{1}{1 + A_w \hat{\varepsilon}_w^{-1}}}.
$$

Our earlier equation (15) related the local Pareto coefficients of the income, buyer payoff and talent distributions. If the first two correspond to $A_w$ and $A_\pi$ and the last to $A_h = \frac{H}{\Delta h} = \frac{\int h(h_{(v_0)} h f(h) dh}{\int h_{(v_0)} h f(h) dh - h(1)}$ either because the underlying distributions are tail Pareto or because limits are taken, then (15) implies that (23) can be reduced to:

$$
\tau^* = \frac{1}{1 + \hat{\varepsilon}_w^{-1} A_h}.
$$

Strikingly (24) is a conventional formula for optimal taxes in terms of the “primitive” talent distribution and the effort elasticity. Thus, when $\chi = 1$, the standard optimal
tax formula holds in terms of talents, but not incomes and when $\chi = 0$ the situation is reversed. For alternative $\chi$ values neither standard formula holds. As we have previously argued when $\chi = 1$, optimal marginal taxes are lower than those implied by the Saez formula (21), but the formula in terms of the talent distribution and the effort elasticity is valid. Equation (15) supplies the reconciliation: in the assignment model, the talent Pareto coefficient is raised relative to the income Pareto coefficient and relative to the value implied by a standard labor market model.\textsuperscript{25}

**Relation to Piketty et al. (2014).** In Piketty et al. (2014) a tax equation similar to (20) is obtained. However, both the model from which it is derived and its interpretation is very different. Piketty et al. (2014) assume that top income earners exert privately costly effort in bargaining to secure income. Income secured from bargaining over and above the high income earner’s contribution to production is netted out of the tax receipts collected from the population of other agents. The elasticity of this term with respect to $1 - \tau$ takes the place of $-\chi \Pi W E_w$ in our tax formula. Despite the similarity of Piketty et al.’s tax formula to ours, its interpretation and its relationship with the data are quite different. In Piketty et al. (2014) the analogue of $-\chi \Pi W E_w$ is a bargaining elasticity that may be positive indicating that marginal taxes should be higher than in the standard case. In this case, the additional tax can be interpreted as a Pigouvian correction for rent seeking in bargaining. This is clearly different from the interpretation of the term in our model with competitive assignment.

### 4 Quantitative analysis of optimal top tax rates

We now use the formulas from the preceding section to evaluate optimal marginal tax rates on top CEO incomes. Our motive for focussing on CEOs is two fold. First, regulatory requirements on US corporations provide us with high quality, publicly available non-top coded data. Second, as discussed earlier, the buyer/seller model of this paper has a natural interpretation within the firm/CEO relationship. Specifically, the non-CEO claimants on firm income (whom we label “firms” in the following) may be interpreted as buyers, while CEOs may be considered sellers.

We characterize optimal tax rates on top CEOs for the cases $\chi = 0$ (no concern for firm claimants) and $\chi = 1$ (equal social weighting on tax revenues and firm

\textsuperscript{25}In a standard model $DSz = w$, $\pi = 0$, and (15) implies: $(1 + E) \frac{1 - F(k)}{h_f(h)} = \frac{1 - M(w)}{w m(w)}$. If $E$ is small, then the talent and income Pareto coefficients are close.
We view these as natural benchmarks. To the extent that the economic profits of firms are dispersed more widely than to those with very high incomes (or very high wealth), the $\chi = 0$ case seems stringent. On the other hand, to the extent that tax revenues are directed to those to whom society attaches greatest weight (e.g. the poor) and to the extent that these individuals are not claimants on firm profits, the case $\chi = 1$ weights firm claimants heavily. We explore intermediate cases $\chi \in (0,1)$ in the later nonlinear tax section.

### 4.1 Optimal top tax rates with no social concern for firm claimants

If the policymaker attaches no weight to firm claimants ($\chi = 0$) or CEOs, then our preceding analysis implies that the optimal tax rate on top earning CEOs is given by the Saez formula:

$$\tau^* = \frac{1}{1+E_wA_w}, \quad (25)$$

where $E_w$ and $A_w$ are now interpreted as the elasticity with respect to the marginal retention rate of and tail coefficient for top CEO incomes (above a high threshold $w_0$).\(^{26}\) There is limited direct evidence on $E_w$ for CEO’s. Time series evidence shows a strong negative correlation between top marginal tax rates and CEO incomes. However, regressions provided by Frydman and Molloy (2011) indicate a small contemporaneous response of CEO incomes to tax reforms. Given this we use multiple values for $E_w$. In the context of top income earners (but not necessarily CEOs), Diamond and Saez (2011) select a value for $E_w$ of $1/4$. We proceed cautiously and use this as an upper bound. We also use a more conservative value of $E_w = 1/10$ (which is more in line with the evidence of Frydman and Molloy (2011)). Consistent with our choice of zero welfare weights on CEOs, our focus on low elasticities biases our results towards higher optimal marginal taxes on CEO incomes.

#### Recovering $A_w$ from CEO income data

We use data on CEO compensation from 1947 to 2011. Data from 1947 to 1991 is from Frydman and Saks (2010). Data from 1992 to 2011 is taken from ExecuComp (a CEO compensation dataset from S&P Capital IQ).\(^{27}\) The measure of compensation considered includes the amounts

\(^{26}\) Strictly, the values of $A_w$ and $E_w$ in (25) are those derived at the optimum rather than at the equilibrium implied by the data. However, if the underlying preferences are quasilinear/constant effort elasticity on the CEO’s side and multiplicative on the firm’s and if the actual tax system is linear above a threshold, then our theory implies that the empirical values of $A_w$ and $E_w$ remain valid at the optimum.

\(^{27}\) We drop data from 1936-1946. The dataset compiled by Frydman and Saks (2010) only features between 70 to 90 CEOs per year in the time period 1947 to 1991. Data from 1992 onwards feature...
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 may be calculated by either 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. However, we also report calculations using the former approach.

We proceed by fitting a Paretian tail to CEO compensation data and identifying $A_w$ with the estimated Pareto coefficient $\hat{\alpha}_w$.\(^{28}\) Specifically, we assume that the income distribution $M$ has a Paretian tail with parameter $\alpha_w$ above an income $w$:

\[
\ln(1 - M(w)) = \alpha_w \ln(w) - \alpha_w \ln(w)
\]

and use the estimated value of $\alpha_w$ to proxy $A_w$. If the threshold $w$ was known, then $\alpha_w$ could be estimated by simple OLS using compensation data. However, in practice, it is not. Consequently, we follow the approach of Clauset et al. (2009) and first estimate $\alpha_w$ by maximum likelihood at each $w$ value. We then select the $w$ value (and corresponding $\alpha_w$ estimate) that maximizes a Kolmogorov-Smirnov goodness-of-fit statistic.

In Figure 1 we display our estimates for $\alpha_w$. We note that our estimate varies between about 2.5 and 4.5. Our estimates of the parameter $\alpha_w$ are somewhat larger (implying a more compact distribution) than the numbers typically used to describe the distribution of high income earners in the taxation literature. Saez (2001) reports a value of $\hat{\alpha}_w$ equal to 2, while Diamond and Saez (2011) and Piketty et al. (2014) assume a value of 1.5.\(^{29}\) As an aside, we note two striking secular move-

\(^{28}\)As Diamond and Saez (2011) point out $A_w$ may exhibit a discontinuity at the very highest income (at which $A_w = \infty$), but be stable over the interval of incomes below the highest. In this case, our approach should be interpreted as generating an approximation to $A_w$ (and corresponding income tax rate) for top, but not the very top income earner. In addition, although not explicitly incorporated into our model, income tax rates are set before and in anticipation of the realized cross section of top incomes. Our approach gives guidance on the setting of top income tax rates before knowledge of the very highest income is available.

\(^{29}\)The lower $\alpha_w$ value we estimate from CEO data relative to that commonly cited in the literature reflects a difference in data sets (CEO vs. general population), combined with our attempts to separate earned income from other sources of income and, possibly, differences in estimation procedures. Using (partially aggregated) data from the World Top Income Data Base, after attempting to isolate earned income in the general population, we obtain estimates of $\alpha_w$ ranging from 1.6 to 4.5 and averaging a little over 2 since the early 1990s. Secular variation across time in this data
Figure 1: Estimates of $\hat{\alpha}_w$. Data from 1947-1991 is from Frydman and Saks (2010); Data from 1992 to 2011 is from ExecuComp. Compensation measures use the value of exercised options. Dashed lines represent 95 percent confidence intervals.

In the distribution of CEO compensation. First, from the mid-1960s to the late 1970s this distribution became more compact, with our parameter estimates $\hat{\alpha}_w$ increasing from 3 to just above 4. Then, from the late 1970s until 2000 the trend towards compactness reversed, the tail of the CEO compensation distribution fattened and $\hat{\alpha}_w$ fell to around 2.5. Since 2000 $\hat{\alpha}_w$ has moderately increased. Explanations for these movements lie outside the scope of the paper, but the assignment framework suggests that they might be due to variations in the underlying distribution of firm as well as CEO characteristics.

For comparison, we also compute $\hat{\alpha}_w$ using CEO compensation measures based on the value of granted options, see Figure 2(a). This gives values slightly higher (and, hence, further from the typical values used) than the previous approach. Secular patterns are similar but dampened. Finally in Figure 2(b), we display the estimated threshold value of $\widebar{w}$. Values displayed are real in 2014 dollars (CPI for all goods and all urban consumers is used). As expected the threshold is increasing until and peaking in 2005. The value ranges from 1.2 million 2014$ in the early 1950s to approximately 24 million 2014$ in 2005.

mirrors that in our CEO data. For further discussion see Appendix E.1.

30It is this evolution that has preoccupied the literature with its focus on the growth of very high CEO incomes.
Computed Optimal Tax  In Figure 3 we plot the optimal top tax rate implied by our estimates of \( \alpha_w \) and alternative values for \( \mathcal{E}_w \) using equation (25). Figure 3(a) shows that for the conservative \( \mathcal{E}_w \) value of 1/10, optimal marginal tax rates vary between 70% and 80% depending upon the estimated value of \( \hat{\alpha}_w \) in any given year. The \( \mathcal{E}_w \) value of 1/4 used by Diamond and Saez (2011) implies more moderate (but still high relative to current policy) marginal tax rates of between 48% and 62%.\(^{31}\) In Figure 3(b) we show the optimal tax rates implied by alternative compensation measures based on the value of options exercised (our benchmark) and the value of options granted respectively. Both experience similar secular movements, but the higher \( \hat{\alpha}_w \) values implied by the latter measure translate into slightly higher optimal marginal tax rates.

4.2 Optimal top tax rates with equally weighted firm claimants and tax revenues

We now assume that firm claimants and tax revenues receive equal weight in the social objective (\( \chi = 1 \)) and specialize to the quasilinear/multiplicative case previously considered. Recall that in this case the classical optimal tax equation in

\(^{31}\)Note that our higher estimates for \( A_w \) imply that this choice of \( \mathcal{E}_w \) gives lower optimal tax numbers than Diamond and Saez (2011).
Figure 3: Top tax rates.

(a) High vs. low elasticity

(b) Options granted vs. exercised. $\E_w = 1/10$

Connecting the tails of the talent, CEO income and firm value distributions

In the abstract buyer-seller assignment model, equation (15) related the Pareto coefficient of the talent distribution to those of the seller income and buyer payoff distributions. To obtain an empirical counterpart of firm (buyer) payoffs, we follow the strategy of Terviö (2008). This strategy allows for CEO effective labor to have a long lasting effect on firm surplus and for empirical measures of firm size to absorb returns to adjustable capital as well as returns to the indivisible asset $S$. We extend the strategy to incorporate CEO effort which is omitted from Terviö’s model. It leads to a modified and empirically operational version of (15) that relates the Pareto coefficient of the talent distribution to those of CEO incomes and firm market values, both of which can be recovered from the data.
Terviö (2008) assumes that the impact of CEO effective labor on a firm’s surplus is long-lived but decays at a rate $\lambda$. Thus, the amount of CEO effective labor used at date $t$ is:

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

Tervio further assumes that all firms grow at a common rate $g$ so that the per period surplus of a firm with asset $S$ is $(1+g)^tSz_t$, that CEO outside options grow at a corresponding rate and that firms discount their future surpluses at a common rate $r$. Under these assumptions a firm hires the same amount of CEO effective labor in each period in equilibrium. Thus, a firm with asset $S$ hiring constant effective labor $z$ obtains a present discounted value of firm surplus: $\frac{Sz}{1-B}$, where $B = \frac{1+g}{1+r}$ is the growth adjusted discount factor. This surplus is split between CEO’s and firm owners according to:

$$\frac{w}{1-B} + p = \frac{Sz}{1-B},$$  

where $p$ denotes the capitalized sum of economic profits. The latter need not equal observed firm market value $q$ since $q$ also includes the value of optimally chosen adjustable capital. If firm surplus is Cobb-Douglas in $Sz$ and adjustable capital with capital share $\theta$, then the present discounted value of surplus (net of payments to adjustable capital and after maximization over adjustable capital) remains equal to $\frac{Sz}{1-B}$. In addition, economic profit $p$, after deducting optimal capital from observed firm market value, is:

$$p = \xi q - (1 - \xi) \frac{w}{1-B},$$  

where $\xi = \frac{1-\theta}{1-\theta + \xi(1-B)}$. Combining (28) and (29) and totally differentiating gives “Terviö’s accounting formula”:

$$\frac{S_{v'}}{S} + \frac{z_v}{z} = \frac{\xi w_v + (1-B)\xi q_v}{\xi w + (1-B)\xi q},$$  

which relates firm asset $S$ and effective labor $z$ growth (over rank) to growth in a composite of CEO income and firm market value. Setting (30) aside, consider the firm’s problem. Given the long-lived nature of CEO effective labor, a firm with indivisible assets $S$ that perturbs its effective CEO labor from the equilibrium value.

\footnote{Note that throughout firm surplus is net of payments to adjustable inputs such as non-CEO labor that generate income neither for CEOs nor for firm owners.}
to \( z' \) obtains a present value of surplus equal to:
\[
\frac{\lambda}{\lambda + 1 - B} S(z' - z) + \frac{S_z}{1 - B}.
\]

Thus, the firm will choose \( v \) to maximize:
\[
DS_z(v) - w(v) \quad \text{with} \quad D := \frac{\lambda}{\lambda + 1 - B}.
\]
The firm’s first order condition implies:
\[
\frac{w_v}{w} = \left( \frac{DS_z}{w} \right) \frac{z_v}{z} = \left( \frac{\lambda \xi}{\lambda + 1 - B} \right) \left( \frac{w + (1 - B)q}{w} \right) \frac{z_v}{z'}, \tag{32}
\]
where (28) and (29) are used to replace \( S_z \) and obtain the second equality. If the actual tax system is linear above a given \( w_0 \), then for sufficiently highly ranked CEOs, the CEO’s first order condition gives:
\[
\frac{z_v}{z} = (1 + \mathcal{E}) \frac{h_v}{h} + \mathcal{E} \frac{S_v}{S}. \tag{33}
\]

Combining the accounting formula (30) and the first order conditions (32) and (33) together with \(-h_v = \frac{1}{f(h)}\), \(-w_v = \frac{1}{w(m)}\) and \(-q_v = \frac{1}{k(q)}\), where \( k \) is the firm market value density and \( K \) its distribution, we have the following relation:
\[
\frac{1 - F(h)}{f(h)h} = N_h \left( \frac{w}{w + (1 - B)q} \right) \frac{1 - M(w)}{m(w)w} + P_h \left( \frac{(1 - B)q}{w + (1 - B)q} \right) \frac{1 - K(q)}{k(q)q}, \tag{34}
\]
with constants \( N_h = \left( \frac{\lambda + 1 - B}{\lambda \xi} \right) - \frac{\mathcal{E}}{1 + \mathcal{E}} \) and \( P_h = -\frac{\mathcal{E}}{1 + \mathcal{E}} \). We also note that the reciprocal of the Pareto coefficient of the \( S \) distribution is given by:
\[
\frac{1 - G(S)}{g(S)S} = N_s \left( \frac{w}{w + (1 - B)q} \right) \frac{1 - M(w)}{m(w)w} + P_s \left( \frac{(1 - B)q}{w + (1 - B)q} \right) \frac{1 - K(q)}{k(q)q}, \tag{35}
\]
with constants \( N_s = 1 - \left( \frac{\lambda + 1 - B}{\lambda \xi} \right) \) and \( P_s = 1 \). Equations (34) and (35) indicate that the reciprocals of the Pareto coefficients for talent and firm assets are weighted sums of those for CEO incomes and firm market values. Note, however, that while both enter positively into (35), the firm market value reciprocal Pareto coefficient enters negatively into (34): given CEO income growth (across rank), more rapid firm market value growth indicates a greater contribution of firm assets \( S \) to the scaling up of CEO incomes and a correspondingly reduced contribution of talent to income growth. In addition large empirical values for \( q \) relative to \( w \) depress the positive weight on \( \frac{1 - M(w)}{m(w)w} \) and further reduce the negative weight on \( \frac{1 - K(q)}{k(q)q} \) lowering the inferred values for \( \frac{1 - F(h)}{f(h)h} \).
Equation (34) may be reformulated as a differential equation for \( h \) (as a function of rank) and integrated to obtain CEO talent as a function of CEO incomes, firm market values, the parameters \( \lambda, \xi, B \) and \( E \) and an initial condition for \( h \). An equation for indivisible firm assets \( S \) may be similarly recovered from (35). We use these equations to recover empirical proxies for \( h(v)/h(1) \) and \( S(v)/S(1) \) and hence the tail coefficients \( A_h \) and \( A_s \). We focus on the time period 1992-2011. We obtain CEO income data as before. In addition, for each CEO in the ExecuComp dataset we use information from the Center for Research in Security Prices (CRSP) database on the number of shares outstanding and average monthly share price to compute an average market capitalization for the CEO’s firm. We equate this value with market value \( q \). We continue to rank CEOs and firms by CEO compensation. We follow Terviö (2008) and select values for \( g = 0.025, r = 0.05, \lambda = 0.5 \) and \( \theta = 0.4 \) (additional robustness tests with respect to the parameters are performed in Appendix E.1). These pin down the parameters \( B \) and \( \xi \) in (34). We assume constant elasticity preferences and set the effort elasticity \( E \) to the low value \( 1/15 \).

Empirical characterization of the tail of the talent distribution Figure 4 plots computed values for \( \frac{A_h}{A_h-1} \) and for comparison \( \frac{A_w}{A_w-1} \) for 2011. As the figure indicates \( \frac{A_h}{A_h-1} \) is very close to one over all of its domain, implying that the tail coefficient \( A_h \) is very large. This in turn means that the talent distribution is not thick tailed and is not likely to be well described by a distribution with a Pareto right tail. In contrast \( \frac{A_w}{A_w-1} \) is fairly stable between 1.5 and 2 for a range of incomes from $10 million to $90 million; only at the very top income in our sample (of $155 million) does \( \frac{A_w}{A_w-1} \) fall to zero. Thus \( A_w \) corresponds to the (hypothetical) description given in Diamond and Saez (2011) p. 11 of a distribution whose tail behavior is well approximated by a Pareto distribution except at the very top.

In the next subsections we describe two statistical characterizations of the tail of the CEO talent distribution (and the firm asset distribution). First, we fit a Pareto distribution to the talent tail following the procedure of Clauset et al. (2009). Then we fit a generalized extreme value distribution that allows for more flexible tail behavior. In Appendix E.3, we calculate Hill and Pickands estimators of the talent

---

33 This is due to the much larger sample of CEOs available from 1992 onwards.

34 Terviö (2008) uses data from CompuStat focusing on the top 1000 firms only, our sample uses all firms available in ExecuComp. In addition our measure of market capitalization is the average monthly average capitalization, while to impute market capitalization Terviö (2008) considers the reference year average monthly December price and share outstanding. Finally Terviö (2008) ranks CEOs by the market capitalization of the firm they manage as opposed to their compensation.

35 This choice is consistent with the low income elasticity \( E_w \) of \( 1/10 \) assumed in Subsection 4.1. Recall that our conservative elasticity values bias our results towards high marginal taxes.
distribution’s tail index and its reciprocal for various years. All of these approaches point to the same result: the talent distribution lacks a thick tail and $A_h$ is very large over a range of incomes. This suggests in turn that for the $\chi = 1$ case marginal taxes on high CEO incomes (and not just the very highest) should be negligible. We emphasize that if the data was treated as if it were generated by a standard model in which CEO income was identified with effective labor and CEO preferences were quasilinear with a small elasticity of labor supply, then very different results would be obtained: $A_h$ estimates would be similar to those of $A_w$ creating a rationale for high optimal marginal taxes at all incomes (except the very highest).

**Results: Fitted Pareto Distribution** We first proceed as in Subsection 4.1 and assume that the $h$ and $S$ distributions conform to a Pareto distribution above threshold talent and asset levels. We then identify $A_h$ with the corresponding estimated Pareto parameter $\hat{\alpha}_h$. We proceed similarly with $A_s$, the tail coefficient for firm assets $S$, and identify it with the estimated Pareto parameter $\hat{\alpha}_s$. The estimation results for the years 1992-2011 are shown in Figure 5. For comparison purposes, we also show the estimated income Pareto parameter $\hat{\alpha}_w$.\(^{36}\) Our estimates of $\alpha_w$ range between 2 or 3 as do those of $\alpha_s$; however our estimates of $\alpha_h$ are between 20 and 40. As anticipated, the estimates of $\alpha_h$ are clearly much larger than those of $\alpha_w$.

\(^{36}\)We re-estimate this using the slightly smaller sample of CEOs for which we have compensation and corresponding firm values.
\( \alpha_w \) and \( \alpha_s \). The average value of \( \hat{\alpha}_h \) over the sample is about 30. Combined with our

![Graph showing estimates of the Pareto coefficients \( \hat{\alpha} \) for the distributions of \( h, s \) and \( w \).](image)

Figure 5: Estimates of the Pareto coefficients \( \hat{\alpha} \) for the distributions of \( h, s \) and \( w \).

(conservative) value of the elasticity of \( \mathcal{E} \) of 1/15, this implies an optimal tax rate of 33%. In 2011, \( \hat{\alpha}_h \) takes a value of 45. Using this gives an optimal tax rate of 25%. However, the size and volatility suggest of the Pareto estimates suggest that the Pareto tail assumption that underpins this estimation procedure is misspecified. Alternative and more flexible estimation approaches give much larger values for \( A_h \) and, hence, much lower values for taxes.

**Results: Generalized Extreme Value Distribution** We now relax the parametric Pareto assumption on the distribution of \( h \) and assume instead that the tail behavior of \( h \) is described by a generalized extreme value (GEV) distribution with
distribution function:\textsuperscript{37,38}

\[ G(x|\eta) = \begin{cases} 
\exp \left\{ -\left(1 + \eta x\right)^{-1/\eta} \right\}, & \text{if } \eta \neq 0 \\
\exp \left\{ -\exp\{-x\} \right\}, & \text{if } \eta = 0.
\] (GEV)

The GEV restriction permits a wider range of tail behaviors: if the extreme value index \( \eta \) is strictly positive then the GEV is tail equivalent to a Frechét distribution; if it is 0 then \( G \) is tail equivalent to a Gumbel distribution, while if \( \eta \) is negative, then \( G \) is tail equivalence to a Weibull distribution. Lemma 1 connects the limit properties of the GEV distribution with the limiting Pareto coefficient (for talents).

**Lemma 1.** Let \( h \) be distributed according to a GEV distribution. Then
\[ \lim_{h \uparrow h(0)} \frac{f(h)}{1-F(h)} = \frac{1}{\eta} \text{ if } \eta > 0 \text{ and } \infty \text{ if } \eta \leq 0. \]

**Proof.** In Appendix E.2

From Lemma 1, the behavior of the Pareto coefficient and, hence, \( A_h \) at top incomes is determined by \( \eta \). Table 1 displays maximum likelihood estimates for the GEV distribution of \( h \). The estimate of \( \eta \) for the talent distribution is small, negative and precisely estimated. The confidence intervals indicate that the hypothesis that \( \eta \) is less than zero is highly statistically significant. In Table 1 we also include estimates obtained from fitting a GEV distribution to the distribution of firm assets \( S \). The estimate (0.682) is this time positive indicating that the distribution of \( S \) exhibits a tail behavior similar to distributions (like the Pareto) in the domain of attraction of the Frechét distribution. This result for \( S \) is consistent with our earlier Figure 5.

In Appendix E.3 we conduct additional robustness tests. We first show that varying the value of \( \lambda \), \( \theta \) and \( g \) impacts only slightly the above GEV estimates. We then use the non-parametric Hill procedure to obtain an estimate of the limiting Pareto coefficient and the Pickands procedure to obtain an estimate of its reciprocal. The former gives point estimates that are over 100 on average for the years

\textsuperscript{37}The Pareto distribution can be generated by characterizing the (normalized) distribution of the maximum of the \( N \) draws of an underlying random variable \( X \). This interpretation is appealing in this setting since the underlying talent of a worker can be thought as a multidimensional random variable. The observed income of a worker is then governed by the best talent characteristic of the worker, perhaps via a process of learning and specialization on the worker’s part. In a well known result Fisher and Tippett (1928) determine that there exists three possible type of distributions that characterize the (normalized) maximum of \( N \) draws: Frechét (to which the Pareto distribution is related), Weibull and Gumbel. Each of the three limiting distribution is associated with a maximum domain of attraction that relate the underlying distribution function for \( X \) to the respective limiting extreme value distribution.

\textsuperscript{38}The GEV distribution is a convenient single parameter representation that encompasses all of the distributions resulting from the Fisher and Tippett (1928) theorem. For additional detail refer to De Haan and Ferreira (2007).
Table 1: GEV Distribution Estimates

<table>
<thead>
<tr>
<th>Distribution</th>
<th>η</th>
<th>µ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talent</td>
<td>-0.332</td>
<td>1.106</td>
<td>0.0425</td>
</tr>
<tr>
<td></td>
<td>(-0.348, -0.316)</td>
<td>(1.104, 1.108)</td>
<td>(0.0411, 0.0439)</td>
</tr>
<tr>
<td>Firm Assets</td>
<td>0.682</td>
<td>21.60</td>
<td>19.13</td>
</tr>
<tr>
<td></td>
<td>(0.617, 0.747)</td>
<td>(20.50, 22.70)</td>
<td>(17.99, 20.34)</td>
</tr>
</tbody>
</table>

Notes: Maximum likelihood estimates of the three parameter GEV distribution with density: $f(x|\eta, \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\left(1 + \eta \frac{x-\mu}{\sigma}\right)^{-1/\eta}\right) \left(1 + \eta \frac{x-\mu}{\sigma}\right)^{-1-1/\eta}$ for $\eta \neq 0$. Here we augment the basic GEV form (GEV) to include location $\mu$ and scale $\sigma$, as well as shape $\eta$, parameters. 95 percent confidence intervals in parenthesis.

1992 to 2011 (and that peak at around 250) for the Pareto coefficient; the latter gives negative point estimates for the reciprocal of the coefficient, consistent with the above GEV estimates.

**Tax implications** Collectively the results suggest that $A_h$ diverges to $\infty$ (or converges to a very large limiting value), that the distribution of talent is thin tailed and most likely in the domain of attraction of the Weibull distribution. Thus, for the $\chi = 1$ case, we interpret the results as consistent with a negligible and, perhaps, zero optimal marginal tax rate on very high CEO incomes.

This section has focused on top tax rates for the benchmark cases $\chi = 0$ and $\chi = 1$. The next section expands the theoretical analysis and considers the fully non-linear optimal tax schedule for a range of $\chi$ values. This permits us to revisit the results of this section for intermediate values of $\chi$ as well as considering optimal marginal tax rates across the entire distribution of large CEO incomes.

## 5 Optimal non-linear taxation

We now generalize our earlier results and consider the policymaker’s optimal choice of non-linear tax function over (all) incomes. We revert again to the language of buyers and sellers to emphasize the general applicability of our results. In the general non-linear setting, the simplest and most direct way of deriving the optimal non-linear 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 buyer taste distributions. We then derive formulas in terms of the seller income and buyer payoff distributions and the seller income elasticity via direct perturbation of the tax function. Note that the latter is complicated relative to Saez (2001) by the endogeneity of the price schedule.

The policymaker’s mechanism design problem

It is convenient to reformulate a tax equilibrium in terms of a tuple \((\bar{v}, z, w, \Phi)\), where \(\Phi\) gives the seller’s utility with \(\Phi(v) = U(c(v), z(v)/h(v))\). Let \(C(\phi, z/h)\) be the consumption of a seller when her utility is \(\phi\) and her effort \(z/h\). We focus on smooth allocations and relax the global seller and buyer incentive constraints (10) and (11), replacing them with, respectively, the seller’s envelope condition and the buyer’s first order condition.\(^{39}\)

To further simplify matters and to align our work more closely with Saez (2001) and Diamond and Saez (2011), we (continue to) focus on the case in which sellers receive zero welfare weight.\(^{40}\) Finally, we assume that it is optimal to induce all buyers to match (so that \(\bar{v} = 1\)). The government’s problem may then be expressed as: \(^{41}\)

\[
\sup_{\Phi, z} \int_0^1 \{\chi V(S(v), z(v)) + (1 - \chi)w(v) - C[\Phi(v), z(v)/h(v)]\} \, dv \tag{36}
\]

subject to for \(\Phi(1) \geq \bar{U}, V(S(1), z(1)) \geq w(1)\), and for \(v \in I\),

\[
\Phi_v(v) = -U_e \left[ C \left[ \Phi(v), \frac{z(v)}{h(v)} \right] \right] \frac{z(v) \, h_v(v)}{h(v)} \tag{37}
\]

\[
w_v(v) = V_z(S(v), z(v))z_v(v). \tag{38}
\]

The optimal tax formula in terms of primitive distributions

After formulating (36) as an optimal control problem and after conventional manipulations (see Appendix C for details), we obtain the following expression for optimal marginal taxes in terms of local Pareto coefficients of the seller talent distribution \(f(h)h\) and the buyer taste distribution \(g(S)S\), the compensated and uncompensated elasticities of

\(^{39}\)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.

\(^{40}\)We assume that buyers receive a constant, but possibly positive weight.

\(^{41}\)The case \(\chi = 1\) is somewhat easier to solve since \(w\) no longer appears in (36). It can be formulated as an optimal control problem with two rather than three state variables. In the context of an economy in which sellers are managers who also collectively own the firms to which they sell their labor (so that all pre-tax output is distributed to manager-owners), problem (36) also has an interpretation as the dual problem of a policymaker who is Rawlsian with respect to sellers.
seller effort $E^c$ and $E^u$, the welfare weight $\chi$ and the co-state $p^\Phi$, \(^{42}\)

$$T^*_{w[w]} = \frac{1 + (1 - \chi) \frac{1-F(h)}{p^\Phi} \frac{E^c}{1+\frac{E^u}{1-F(h)}}}{1 + \frac{1-F(h)}{p^\Phi} \frac{E^c}{1+\frac{E^u}{1-F(h)}}},$$

where $p^\Phi$ satisfies:

$$p^\Phi(v) = \int_v^1 \left[ (1 - \gamma(v')U_c(v')) \frac{1}{U_c(v')} \exp \left\{ - \int_v^{v'} \frac{U_c(v''')}{U_c(v''')} z(v''') \frac{h_v(v''')}{h(v''')} dv''' \right\} \right] dv'. \quad (40)$$

As in the affine case it follows from (39) that higher values of $\chi$ are a force for lower marginal tax rates. In addition, if $\chi = 1$, then (again as in the affine case) (39) reduces to the standard formula found in conventional Mirrleesian analysis:

$$T^*_{w[w]} = \frac{1}{1 + \frac{1-F(h)}{p^\Phi} \frac{E^c}{1+\frac{E^u}{1-F(h)}}}, \quad (41)$$

Inspection of the derivation of equations (39) and (41) enhances intuition. When $\chi = 1$ the policymaker attaches equal unit weight to buyer surplus and tax revenues. Seller income detracts from buyer surplus, but (holding seller consumption fixed) augments tax revenues. Thus, it nets out and receives zero direct weight in the policymaker’s objective (36). As in the conventional Mirrlees model, the social marginal value of a (compensated) increase in service provided by seller $v$ must be equated to the social marginal cost of supplying more utility (informational rents) to higher ranked sellers. In the assignment setting the social marginal value of an increase in service is modified from the usual $1 + \frac{U_c(v)}{h(v)U_c(v)}$ to $V_z(S(v), z(v)) + \frac{U_c(v)}{h(v)U_c(v)}$. However, the compensated (pre-tax) marginal value to a seller from supplying such a unit is similarly modified from $1 + \frac{U_c(v)}{h(v)U_c(v)}$ to $V_z(S(v), z(v)) + \frac{U_c(v)}{h(v)U_c(v)}$ in equilibrium.\(^{43}\) Consequently, the conventional optimal tax formula holds.

If $\chi < 1$, then the policymaker places less weight on buyer payoffs and, as previously noted, implements a corresponding higher income tax rate. Inspection of (36) indicates that compared to the $\chi = 1$ case seller incomes now receive separate positive weight in the social objective. Thus, the policymaker has an incentive to further increase seller incomes and, so, tax revenues at the expense of buyer surplus. The derivation of (39) indicates that this is achieved by raising marginal

\(^{42}\)All functions below are understood to be evaluated at (the allocation associated with) the seller’s rank $v$. To economize on notation the $v$ argument is dropped.

\(^{43}\)The latter is a consequence of the buyer’s incentive condition which equates marginal changes in seller income per unit of service to $V_z$. 

32
taxes on lower seller incomes, suppressing service by less talented sellers and, thus, raising the income increment that more desiring buyers must pay for higher levels of service. To understand this better, consider a locally “convexifying” shift to the service schedule $z$ that first reduces (at lower $S(v)$ values) and then raises (at higher $S(v)$ values) $z_v$ over a small interval leaving the value of $z$ unchanged at either end of the interval. The buyer’s incentive condition $w_v = V_z(S(v), z(v))z_v$ and the super modularity of $V$ imply that the association of higher values of $z_v$ with higher $S$ values (and lower values of $z_v$ with lower $S$ values) raises the overall increase in seller’s incomes over the interval. So raising the incomes of higher ranked sellers and tax revenues collected from these sellers (given the tax function over these incomes). A small increase in the marginal tax over a small range of incomes has exactly this “convexifying” effect on the service schedule $z$: seller service and $z_v$ is first depressed over the income range to which the marginal tax increase is applied and then increased at slightly higher incomes. Notice that such local convexifying effects fail to raise higher ranked seller incomes if, as in standard models, $S$ is constant and buyers are homogeneous. Note also that if $\chi > 1$, then these same forces imply that the policymaker has an incentive to reduce taxes on seller incomes below those in the $\chi = 1$ case and, perhaps, even to subsidize seller incomes.

The optimal tax formula in terms of the income distribution. In the affine case we presented an optimal tax formula in terms of the elasticities and tail coefficients of seller incomes and buyer payoffs. We obtained this formula 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 the “primitive” formulas given above to nonlinear tax formulas in terms of induced seller income and buyer payoff variables. We focus on the quasi-linear/multiplicative preferences case.

In Appendix D we construct the implications for tax revenues and seller incomes of an explicit perturbation of a tax function in the assignment setting. Following Saez (2001), the perturbation decomposes the overall effect of a variation in the marginal income tax $T_w$ at an income level $w_0$ on tax revenues into “mechanical” and “behavioral” parts.\footnote{We perturb the marginal retention rate $1 - T_w$ over a small interval of incomes. We successively take derivatives with respect to the perturbation parameter and with the length of the interval (i.e. we normalize the impact on revenues by the size of the parameter and take the limit as the parameter goes to zero). Perturbation of a marginal tax “at” an income can also be approached via the theory of distributions and generalized functions. This approach, of course, gives identical results.} The first of these gives the impact on tax revenues of
the perturbation holding the price schedule $w$ fixed. As is standard, it is given by $1 - M(w_0)$ and is interpreted as the additional tax revenues raised from sellers with incomes above $w_0$ by an increase in marginal taxes at $w_0$. 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 entire schedule of incomes. Let $v_0$ denote the buyer-seller rank corresponding to the equilibrium wage $w_0$ prior to the tax perturbation and let $z_0$ denote the corresponding seller service amount (i.e. $w_0 = w(v_0)$ and $z_0 = z(v_0)$, where $w$ and $z$ are the pre-perturbation equilibrium schedules). Then the behavioral effect is given by:

$$\begin{align*}
- \frac{T_w[w_0]}{1 - T_w[w_0]} m(w_0) w_0 & \frac{\mathcal{E} S(v_0) z_0}{w_0} \frac{w_0}{1 + \mathcal{E} T(v_0) S(v_0) z_0} \\
+ \frac{T_w[w_0]}{1 - T_w[w_0]} m(w_0) w_0 & \frac{\mathcal{E} S(v_0) z_0}{w_0} \frac{w_0}{1 + \mathcal{E} T(v_0) S(v_0) z_0} \frac{1 - G(S(v_0))}{g(S(v_0)) S(v_0)} \\
\times \frac{1}{(1 - v_0) T_w[w_0]} \int_0^{v_0} \frac{T_w[w(v)]}{1 + \mathcal{E} T(v) S(v) z(v)} \exp \left\{ - \int_v^{v_0} \frac{S(v')}{S(v')} \frac{S(v') z(v')}{w(v')} dv' \right\} dv,
\end{align*}$$

where $\mathcal{T}(v) = \frac{T_{w_0}[w(v)]}{1 - T_{w_0}[w(v)]}$ denotes the elasticity of the marginal retention function. The first term in (42) (on the first line) gives the marginal reduction in tax revenues stemming from the decrease in seller service and price at $w_0$. Note that here as in Saez (2001) the perturbation has a direct effect on the marginal tax rate paid and an indirect effect that stems from the reduction in seller income and the non-linearity of the tax function. This indirect effect is captured by the denominator $1 + \mathcal{E} T(v_0) S(v_0) z_0 / w_0$ in the first term in (42) (and is absent if the tax rate $T$ is locally linear at $w$). The component $\frac{\mathcal{E} S(v_0) z_0 / w_0}{(1 + \mathcal{E} T(v_0) S(v_0) z_0 / w_0)}$ of the first term can be interpreted as the “local” elasticity of seller income at $w_0$ in response to a change in the marginal retention rate $1 - T_w$. The second and more complicated term in (42) is absent in standard models. In the assignment setting, a higher marginal tax rate at $w_0$ induces a shift in the equilibrium income schedule that raises the incomes of those above $w_0$. The second term in (42) captures the corresponding increase in tax revenues. It stems from the local convexifying effect of higher marginal taxes at an income (or over a small interval of incomes) described previously.

Consider for the moment a simple case in which the underlying tax function $T$ around which the perturbation is occurring is linear above $w_0$. At $w_0$, seller effort and service is suppressed by the higher marginal tax rate, while above $w_0$ sellers face the same marginal tax rate as before and, given the quasilinearity of seller
preferences, do not change their service supply. Thus the marginal tax increase locally convexifies the service schedule - first depressing and then steepening it. Higher incomes to those above \( w_0 \) ensue. In particular in this linear case \( T = 0 \) and the complicated second term in (42) reduces to:

\[
\frac{T_w[w_0]m(w_0)w_0}{1 - T_w[w_0]} \mathcal{E} \frac{S(v_0)z_0}{w_0} \left( -\frac{S_v(v_0)}{S(v_0)} \right) (1 - M(w_0)),
\]

where we replace \( 1/g(S(v_0)) \) with \(-S_v(v_0)/S(v) \) and \( 1 - G(S(v_0)) \) with \( 1 - M(w_0) \). Then

\[
\frac{m(w_0)w_0}{1 - T_w[w_0]} \mathcal{E} \frac{S(v_0)z_0}{w_0} \left( -\frac{S_v(v_0)}{S(v_0)} \right)
\]

gives the size of the income increase, \( 1 - M(w_0) \) the fraction of agents receiving this increase and \( T_w[w_0] \) the common marginal tax rate paid by these individuals. In combination the term gives the extra tax revenues collected from those above \( w_0 \) as a result of the equilibrium shift in the income schedule. If the underlying tax schedule around which the perturbation occurs is non-linear then this term takes the more complicated form in (42) that incorporates the fact that sellers above income \( w_0 \) pay different marginal tax rates \( T_w[w(v)] \) and that increases in income affect the marginal tax rates they pay inducing a further adjustment in their service supply. In particular, consolidating terms, the overall behavioral effect in (42) may be written as:

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

where \( \mathcal{E}_w(w_0) \) is a weighted elasticity of seller incomes at and above \( w_0 \) with respect to \( 1 - T_w[w_0] \):

\[
\mathcal{E}_w(w_0) := \mathcal{E} \frac{S(v_0)z_0}{w_0} \left( 1 - \frac{1 - G(S(v_0))}{g(S(v_0))} \right) \frac{1}{S_v(v_0)} \int_{v_0}^{T_w[w(v)]} \left\{ \frac{S_v(v')}{S(v')} \exp \left\{ - \frac{1}{1 + \mathcal{E} T(v')} \int_{v_0}^{v'} \frac{S(v'')z''}{w''(v'')} \frac{dv''}{1 + \mathcal{E} T(v'')} \right\} \right\} dv'.
\]

(43)

**Optimal tax implications** When \( \chi = 0 \), the policymaker is concerned only with maximizing tax revenues. Consequently, in this case the sum of the mechanical and behavioral impacts derived above must equal zero at the optimum and so combining terms:

\[
T_w[w_0] = \frac{1}{1 + \frac{m(w_0)w_0}{1 - M(w_0)} \mathcal{E}_w(w_0)}
\]
Moreover, we show in Appendix D that at this optimum $\tilde{E}_w(w_0)$ reduces to:

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

For $\chi$ values greater than 0, the policymaker is also concerned with the impact of marginal tax rates on buyer payoffs. Consequently, the behavioral term is augmented with an extra component that captures the depressing effect of higher marginal taxes at $w_0$ on the payoffs of buyers above $v_0$:

$$-\chi \frac{m(w_0)w_0}{1 - T_w[w_0]} + \xi T(v_0)\frac{S(v_0)z_0}{w_0} g(S(v_0))/S(v_0) \left\{ 1 - \frac{1 - G(S(v_0))}{g(S(v_0))/S(v_0)} \right\}$$

The expression in (44) stems from the local convexifying effect of taxes on the service schedule described previously. This effect raises the incomes and prices of sellers ranked above $v_0$ and, thereby, depresses the payoffs of buyers. In addition, in the presence of nonlinear taxes, these changes to the incomes of sellers affect the marginal taxes they pay leading to further adjustments of $z$. The term (44) is the analogue of $\chi \Pi_1 \xi \Pi_2$ in the derivation of (20). Adding it to the other behavioral and mechanical terms leads to an optimal tax equation analogous to (20) for the nonlinear setting.

### 6 Computing non-linear taxes

Equation (39) may be used to compute optimal marginal taxes over the full range of CEO incomes. As in Section 4.2 we specialize the analysis to quasi-linear/constant elasticity CEO preferences and multiplicative firm production functions. In this case we can rewrite (39) as:

$$T^*_w[w] = \frac{1 + (1 - \chi) \frac{\xi}{1 + \xi} \frac{1 - G(S)}{g(S)S} f(h)h}{1 + \frac{\xi}{1 + \xi} \frac{f(h)h}{1 - F(h)}}.$$ 

We assume the distributions for $h$ and $S$ are of the form (GEV) and use our estimates of $\eta$ for each of these distributions to recover values for the local Pareto coefficients $\frac{f(h)h}{1 - F(h)}$ and $\frac{g(S)S}{1 - G(S)}$ over their entire range. We maintain the conservative benchmark
of the previous quantitative section and set $\epsilon = 1/15$. The results are displayed in Figure 6. Here we plot optimal marginal tax rates against CEO compensation ($w$). The units for $w$ are in millions of 2014 dollars. The figure shows that when $\chi = 1$

\[ \begin{align*}
0 & \leq \chi = 0.5 \\
0 & \leq \chi = 0
\end{align*} \]

optimal tax rates decay from about 35% at incomes of $10$ million to under 5% by incomes of $50$ million. Thereafter they asymptote to zero (as anticipated by the results of the previous section). The range of incomes for which optimal taxes are close to zero is very large. Moreover, over the entire income range they are much lower than the values obtained when $\chi = 0$ or $\chi = 0.5$. With no concern for firm claimants, marginal tax rates decline from 84% to 75%, while in the intermediate $\chi = 0.5$ case, they fall form about 60% to a little below 40%.

\section{Conclusion}

In this paper we have extended the analysis of optimal taxation to assignment settings. In these settings, buyers extract surplus from trade and social concern for buyers emerges as an additional policy relevant parameter. When there is no such concern conventional optimal tax formulas expressed in terms of standard elasticities of income and the tail coefficient of income remain valid, but formulas expressed in terms of the tail coefficient for talents are not. The reverse is true when social concern for buyers equals the social value attached to tax revenues. More generally, (given an arbitrary concern for buyers) optimal nonlinear tax equations in the assignment setting involve Pareto coefficients for both seller talent and buyer
tastes when expressed in terms of primitives.

We apply the theory to the setting of high earning CEO’s. In this case buyers correspond to firm claimants (primarily firm equity holders) and sellers to CEOs. When the policymaker has concern neither for firm claimants nor very highly paid CEOs (and when effort or income elasticities are set to very low values) optimal tax rates on the CEOs have high values of between 70% and 80%. On the other hand, when the policymaker attaches equal weight to tax revenues and firm claimants (but still none to CEO’s and maintaining a low effort elasticity) the optimal tax rate on CEOs is pushed close to zero on a range of high incomes. Underpinning the latter result are our estimates of the tail coefficient of the talent distribution which suggest that this distribution, unlike the income distribution, is not thick tailed.\textsuperscript{45,46} We do not take a stand on the appropriate value for the firm claimant weight, though, as we have noted, evidence suggests that direct and indirect firm equity holdings are fairly dispersed in the US. More generally, the assignment framework directs us towards consideration of the impact of taxes on buyers when sellers do not capture all of the surplus. Our paper has pursued this idea in the context of one-to-one buyer and seller matching, a natural framework with which to model the firm-CEO relationship. In Ales and Sleet (2015) we apply it to other classes of superstar earners such as entertainers, athletes and entrepreneurs, who can affect both the segment of the market they trade in through the quality of their output and the size of this segment through replication (for example, through recording and promotion of a performance).

References


\textsuperscript{45}And as we have emphasized the talent and CEO income distributions are divorced in the assignment framework.\textsuperscript{46}In deriving our quantitative optimal tax results we have been conservative with respect to effort and income elasticities relative to the literature; we have also placed a zero welfare weight on top earning CEOs. On the other hand we have abstracted from the bargaining frictions in Piketty et al. (2014) or externalities in Rothschild and Scheuer (2014) that operate in the other direction and push for higher marginal taxes. A full integration of these forces with those we have emphasized remains to be done.


Online Appendix

A Assignment economy proofs

We now give a 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 \((w, m, z)\) is an equilibrium at \((T, \bar{U})\), then either (a) \(m = u\), no buyer produces and all candidate sellers take their outside option or (b) there is a \(\tilde{v} \in I\) such that (i) for all \(v \in (\tilde{v}, 1]\), \(m(v) = u\) and (ii) for all \(v \in I \cap [0, \tilde{v})\), \(m(v) = v\).

Proof of Lemma A.1. If \(v' > v\), \(m(v) = u\) and \(m(v') \in I\), then \(V(S(v), z(v')) - \omega(v') > V(S(v'), z(v')) - \omega(v') \geq 0\) and buyer \(v\) is made strictly better off matching with seller \(m(v')\) at the price and service supply terms she receives from \(v'\). Since buyer \(m(v')\) is obviously no worse off, this cannot be an equilibrium outcome and, hence, if \(m(v') \in I\), then \(m(v) \in I\). It follows that either (i) \(m = u\) or (ii) there is some \(\tilde{v} \in I\) such that for \(v \in I \cap [0, \tilde{v})\), \(m(v) \in I\) and for \(v \in (\tilde{v}, 1]\), \(m(v) = u\).

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

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

and:

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

If the first of these conditions did not hold, then seller \(m(v)\) could make herself and buyer \(v'\) slightly better off by offering to supply buyer \(v'\) for a price very slightly below \(w(v')\). Similarly, if the second did not hold, then seller \(m(v')\) could make herself and buyer \(v\) slightly better off by offering to supply buyer \(v\) at a price 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 buyer would pay a higher price to obtain a lower after-tax income and consumption for its seller, \(w(v') \geq w(v)\). Thus, \(m(v')\) supplies more service to \(v'\) than \(m(v)\) supplies to \(v\). If \(m(v')\) instead works for \(v\) and supplies the same service amount \(z(v')\) and accepts the same income as before, then the net of price payoff of buyer \(v\) 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 buyer \(v\) is made better off by the partner swap, while \(m(v')\) is no worse off and so \((w, m, z)\) cannot be an equilibrium. If this is change is non-positive and \(z(v') > z(v)\), then:

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

41
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, buyer $v'$ is made strictly better off by swapping partners with buyer $v$, which again contradicts the requirement that $(w, m, z)$ 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 buyers $v$ and $v'$ swap partners and continue to pay the same prices and require the same service amounts from their sellers, then no buyer or seller is made worse off. Denote the common service amount amount by $\hat{z}$ and the common income by $\hat{w}$ and, to simplify the exposition suppose that the tax function $T$ is differentiable at $\hat{w}$ with derivative $T_w[\hat{w}]$. It cannot be that: $V_z(S(v), \hat{z})h(m(v')) = -\frac{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(m(v')))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(m(v')))}(1 - T_w[\hat{w}])$ and $V_z(S(v'), \hat{z})h(m(v)) = \frac{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(m(v)))}{U_c(\hat{w} - T[\hat{w}], \hat{z}/h(m(v)))}(1 - T_w[\hat{w}])$, since if so the following contradiction emerges:

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

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

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

Finally, we characterize $m$ at $\hat{v}$. Suppose $\hat{v} \in (0, 1)$ and let $\underline{v}_n \uparrow \hat{v}$ and $\overline{v}_n \downarrow \hat{v}$ (with each $0 < \underline{v}_n < \hat{v} < \overline{v}_n < 1$). We have:

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

As observed previously $w_n = w(\underline{v}_n)$, $c_n = w(\underline{v}_n) - T[w(\underline{v}_n)]$ and $z_n = z(\underline{v}_n)$ are bounded, decreasing sequences. Denote the limits of these sequences $(w_\infty, c_\infty, z_\infty)$. 

42
Since \( \lim h(\bar{v}_n) - h(\bar{v}_n) \downarrow 0 \) and \( U \) is continuous, it follows that \( U(c_n, z_n/h(\bar{v}_n)) - U(c_n, z_n/h(\bar{v}_n)) \) converges to 0. Hence, \( W_n \downarrow \bar{U} \). By a similar argument, \( V(S(\bar{v}_n), z(\bar{v}_n)) - w(\bar{v}_n) \downarrow 0 \). It follows that if \( T \) is continuous, then the \( \bar{\sigma} \) buyer and seller are indifferent about matching at the price-service amount \((w_\infty, z_\infty)\).

Without loss of generality we select equilibria in which if \( \bar{\sigma} > 0 \), then \( m(\bar{\sigma}) = \bar{\sigma} \).

The next proposition simplifies the equilibrium conditions (4) to (8) in Definition 1 in a way that is convenient for tax analysis. It shows that given \((T, \bar{U})\), if a pair of price and service functions \((w, z)\) on a domain \( I \cap [0, \bar{\sigma}] \) are such that no seller \( v \in I \cap [0, \bar{\sigma}] \) is made strictly better off exchanging places with seller \( v' \in I \cap [0, \bar{\sigma}] \) and accepting the terms \( v' \) receives and similarly for buyers, then no buyer-seller pair \((v, v') \in \{I \cap [0, \bar{\sigma}]\}^2 \) can benefit (both weakly and at least one side strictly) from re-matching and selecting an arbitrary service amount and a price in the codomain of \( w \). Furthermore, if the \( \bar{\sigma} \) seller and buyer 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 \((\bar{\sigma}, w, z)\) is an equilibrium at \((T, \bar{U})\). Thus, the stability conditions on buyers and sellers in are decoupled and re-expressed as separate buyer and seller incentive conditions. In addition, Proposition A.1 supplies a converse result: associated with any (non-trivial) equilibrium is a \((\bar{\sigma}, w, z)\) satisfying the conditions described above.

**Proposition A.1.** Let \( T : \mathbb{R}_+ \rightarrow \mathbb{R} \) be a tax function, \( \bar{\sigma} \) be a number in \( l \) and \( w : I \cap [0, \bar{\sigma}] \rightarrow \mathbb{R}_+ \) and \( z : I \cap [0, \bar{\sigma}] \rightarrow \mathbb{R}_+ \) be a pair of price and service functions satisfying the participation conditions, for all \( v \in I \cap [0, \bar{\sigma}] \).

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

and the incentive conditions, for all \( v, v' \in I \cap [0, \bar{\sigma}] \),

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

and

\[
V(S(v), z(v)) - w(v) \geq V(S(v), z(v')) - w(v'). \tag{A.3}
\]

Then there is no tuple \((v, v', w', z')\) with \((v, v') \in I \cap [0, \bar{\sigma}] \) and \( w' \in w(I \cap [0, \bar{\sigma}]) \) 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

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

with at least one of these inequalities strict. In addition, if \( \text{(i)} \) \( U \left( w(\bar{\sigma}) - T[w(\bar{\sigma})], \frac{z(\bar{\sigma})}{h(\bar{\sigma})} \right) = \bar{U} \) and \( V(S(\bar{\sigma}), z(\bar{\sigma})) - w(\bar{\sigma}) \geq 0 \) and \( \text{(ii)} \) for all \( w' \notin w(I \cap [0, \bar{\sigma}]) \), \( T(w') = w' \), then \((\bar{\sigma}, w, z)\) defines an equilibrium at \((T, \bar{U})\). Conversely, if \((\bar{\sigma}, w, z)\) is an equilibrium at \((T, \bar{U})\), then \((\bar{\sigma}, w, z)\) satisfies (A.1) to (A.3), \( U \left( w(\bar{\sigma}) - T[w(\bar{\sigma})], \frac{z(\bar{\sigma})}{h(\bar{\sigma})} \right) \geq \bar{U} \) and \( V(S(\bar{\sigma}), z(\bar{\sigma})) - w(\bar{\sigma}) \geq 0 \) with equality if \( \bar{\sigma} \in (0, 1) \).
Proof of Proposition A.1. Suppose the first claim in the proposition is false and that there is a tuple \((v,v',w',z')\) with \(v,v' \in I \cap [0,\bar{v}]\) and \(w' = w(\bar{v}) \in w(I \cap [0,\bar{v}])\) such that:

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

and

\[ V(S(v),z') - w' \geq V(S(v),z(\bar{v})) - w(v), \]

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

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

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

\[ V(S(v),z(\bar{v})) - w(v) \geq V(S(v),z(\bar{v})) - w(\bar{v}) > V(S(v),z') - w(\bar{v}), \]

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

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

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

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

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

The next lemma provides monotonicity results for equilibrium price, service and seller consumption functions. It also proves the existence of a function \( \omega \) relating equilibrium service to price.

**Lemma A.2.** If \((\hat{v},w,z)\) is an equilibrium threshold and a pair of equilibrium price and service functions at \((T,\hat{U})\), then \((w,z,c)\), with \( c(v) = w(v) - T[w(v)] \), \( v \in I \cap [0,\hat{v}] \) are non-increasing and there exists a function \( \omega : z(I \cap [0,\hat{v}]) \to \mathbb{R}_+ \) satisfying for each \( v \in I \cap [0,\hat{v}] \), \( \omega(z(v)) = w(v) \). In addition, the equilibrium seller payoff function \( \Phi, \Phi(v) = U(c(v),z(v)/h(v)) \) for \( v \in I \cap [0,\hat{v}] \), and seller payoff function \( \pi, \pi(v) = V(S(v),z(v)) - w(v) \) for \( v \in I \cap [0,\hat{v}] \), 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 buyer would choose to buy from a seller at price \( w(v) \) (they could strictly reduce the price they pay and weakly raise their seller’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 (2013), 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 \( \Psi \).

\[ \square \]

**B Derivation of elasticities for Section 3**

In this appendix, we derive elasticities used in Section 3 under the assumption of quasi-linear/constant effort elasticity seller preferences and a multiplicative buyer objective. Substituting (13) into (12) and using the quasi-linear assumption and the fact that the \( v \)-th seller matches with the \( v \)-th buyer gives:

\[
(1 - \tau)S(v)h(v) - \Psi_c(e(v)) = 0.
\]

(B1)
It follows immediately that:

$$\frac{1 - \tau}{e(v)} \frac{\partial e(v)}{\partial (1 - \tau)} = \mathcal{E}. $$

In other words, since each seller remains attached to their original buyer and since each seller (buyer) has a constant marginal valuation of consumption (service), a standard elasticity of effort with respect to the marginal income retention rate $1 - \tau$ obtains. The elasticity of the $v$-th buyer’s payoff from service consumption with respect to $1 - \tau$ is then simply:

$$\mathcal{E}_q(v) = \frac{1 - \tau}{S(v)} \frac{\partial S(v)z(v)}{\partial (1 - \tau)} = \mathcal{E}. $$

The elasticity of the seller’s income with respect to $1 - \tau$, $\frac{1 - \tau}{w(v)} \frac{\partial w(v)}{\partial (1 - \tau)}$, is more complicated. We focus on the case $v_0 = 1$. However, the arguments are easily modified for $v_0 \in (0, 1)$ (noting that a tax perturbation above $w(v_0)$ does not affect buyer-seller behavior at ranks $v \in (v_0, 1]$ so that the following calculation goes through with $\Phi(v_0) = U(w(v_0) - T[w(v_0)], z(v_0)/h(v_0))$ replacing $\tilde{U}$).

Since competition amongst sellers drives the lowest talent seller in the market down to his or her outside option, we have after a tax perturbation around the rate $\tau$ that is applied to the incomes of all sellers:

$$(1 - \tau + \delta)\hat{w}(1; \delta) - \{T[w(1)] + (\tau - \delta)w(1)\} - \Psi \left( \frac{z(1, \delta)}{h(1)} \right) = \tilde{U}. \tag{B2}$$

Evaluating (B1) at the perturbed tax function and allocation at $v = 1$ and combining with (B2) gives:

$$\hat{w}(1; \delta) = \frac{1}{1 - \tau + \delta} \left[ T[w(1)] + (\tau - \delta)w(1) + \tilde{U} + B \left( (1 - \tau + \delta)S(1)h(1)) \right) \right],$$

where $B := \Psi \circ \Psi^{-1} \circ e (so that $B_e = \Psi_e / \Psi_{ee})$. Differentiating the previous expression with respect to $\delta$ and evaluating at $\delta = 0$, we obtain:

$$\frac{(1 - \tau)}{w} \frac{\partial w}{\partial (1 - \tau)}(1) = \frac{S(1)z(1)}{w(1)} \mathcal{E}. \tag{B3}$$

Equation (B3) reflects the fact that in equilibrium a marginal change in seller income induced by higher effort equals the marginal change in the buyer’s payoff from consuming or using the service generated by this effort.

Totally differentiating the seller’s first order condition (B1) with respect to $v$ (appealing to the implicit function theorem) gives:

$$(1 - \tau)S(v)h(v) \left[ \frac{S_v(v)}{S(v)} + \frac{h_v(v)}{h(v)} \right] + \Psi_{ee} e(v) = 0. \tag{B4}$$
Combining (B1) and (B4) and rearranging gives:
\[
\frac{e_v(v)}{e(v)} = \mathcal{E} \left\{ \frac{S_v(v)}{S(v)} + \frac{h_v(v)}{h(v)} \right\}.
\]

It then follows from the definition of \( z \) that:
\[
\frac{z_v(v)}{z(v)} = \frac{h_v(v)}{h(v)} + \frac{e_v(v)}{e(v)} = (1 + \mathcal{E}) \frac{h_v(v)}{h(v)} + \mathcal{E} \frac{S_v(v)}{S(v)}.
\]

(B5)

Combining (B5) with the buyer’s first order condition \( w_v(v) = S(v)z_v(v) \) implies that:
\[
w_v(v) = \mathcal{M}(v)e(v),
\]
where:
\[
\mathcal{M}(v) = S(v)h_v(v) \left[ 1 + \mathcal{E} \left( 1 + \frac{S_v(v)}{S(v)} \right) \right].
\]

Then using \( w(v) = w(1) + \int_0^1 w_v(v')dv' \), we have:
\[
\mathcal{E}_w(v) = \frac{1-\tau}{w(v)} \frac{\partial w(v)}{\partial (1-\tau)} = \frac{1-\tau}{w(v)} \frac{\partial w(1)}{\partial (1-\tau)} - \frac{1-\tau}{w(v)} \int_0^1 \frac{\partial w_v(v')}{\partial (1-\tau)} dv'
\]
\[
= \frac{1-\tau}{w(v)} \frac{\partial w(1)}{\partial (1-\tau)} - \frac{1-\tau}{w(v)} \int_0^1 \mathcal{M}(v') \frac{\partial e(v')}{\partial (1-\tau)} dv'
\]
\[
= \frac{1-\tau}{w(v)} \frac{\partial w(1)}{\partial (1-\tau)} - \mathcal{E} \int_0^1 \mathcal{M}(v)e(v')dv'
\]
\[
= \left( \frac{w(v) + S(1)z(1) - w(1)}{w(v)} \right) \mathcal{E}
\]
\[
= \left\{ \frac{S(v)z(v)}{w(v)} - \left( \frac{\pi(v) - \pi(1)}{w(v)} \right) \right\} \mathcal{E},
\]
where \( \pi(v) = S(v)z(v) - w(v) \). And the elasticity of the buyer’s net of price payoff is:
\[
\mathcal{E}_\pi(v) := \frac{q(v)}{\pi(v)} \mathcal{E}_q(v) - \frac{w(v)}{\pi(v)} \mathcal{E}_w(v) = \left( \frac{\{S(v)z(v) - w(v)\} - \{S(1)z(1) - w(1)\}}{\pi(v)} \right) \mathcal{E}
\]
\[
= \left( \frac{\pi(v) - \pi(1)}{\pi(v)} \right) \mathcal{E}.
\]

The formulas for \( \mathcal{E}_w \) and \( \mathcal{E}_\pi \) in the main text follow directly from the above expressions according to:
\[
\mathcal{E}_w = \frac{1}{W} \int_0^\infty w(v) \mathcal{E}_w(v) dv \quad \text{and} \quad \mathcal{E}_\pi = \frac{1}{\Pi} \int_0^\infty \pi(v) \mathcal{E}_\pi(v) dv.
\]
C Derivation of optimal nonlinear tax formulas

When buyers have positive welfare weight $\chi$ the government’s problem is:

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

subject to for $v \in I \cap [0, \bar{v}]$, $\Phi(1) \geq \bar{U}$, $V(S(1), z(1)) \geq w(1)$, and

$$\Phi_v(v) = -U_c(C[\Phi(v), z(v)/h(v)]) \frac{z(v) h_v(v)}{h'(v) h(v)},$$

$$w_v(v) = V_z(S, z)v(v)$$

Now, $z_v$ is a control and $\Phi, z$ and $w$ are additional states. The Hamiltonian for (C.1) is:

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

$$+ p^z(v)z_v(v) + p^w(v)V_z(S(v), z(v))v(v)$$

$$+ \chi V(S(v), z(v)) + (1 - \chi)w(v) - C \left[ \Phi(v), \frac{z(v)}{h(v)} \right],$$

with co-states $p\Phi$, $p^z$ and $p^w$. The first order condition for $z_v$ implies that:

$$p^z + p^w V_z(S, z) = 0.$$

From which:

$$p^z + p^w V_z(S, z) + p^w[V_zS(S, z)S_v + V_{zz}(S, z)z_v] = 0.$$

The optimal co-state equations are:

$$p^\Phi_v = \frac{1}{U_c} + p\Phi \frac{U_{ec} z}{U_c} \frac{h_v}{h}$$

$$p^w_v = -(1 - \chi)$$

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

The transversalities at $v = 1$ are:

$$p^w(1) = -q^V$$

$$p^z(1) = q^V V_z(S(1), z(1))$$

$$p^\Phi(v) = q^U.$$
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:

$$p^w(0) = p^z(0) = p^\Phi(0) = 0.$$ 

Hence,

$$p^\Phi(v) = \int_0^v \left[ \frac{1}{U_c(u)} \exp \left\{ - \int_v^u \frac{U_{ec}(u')}{U_c(u')} \left( \frac{e(u')}{h(u')} \right) h_v(u') du' \right\} \right] du. \quad (C.2)$$

and

$$p^w(v) = -(1 - \chi)(1 - M(w(v))).$$

Combining conditions:

$$-p^\Phi \left\{ \left[ U_{ec} \left( -\frac{U_e}{U_c} \right) + U_{ec} \right] \frac{z}{h} + U_e \right\} \frac{h_v}{h} + [V_z - (1 - \chi)(1 - M(w))V_{zs}(-S_v)] h + \frac{U_e}{U_c} = 0. \quad (C.3)$$

Recall the seller’s first order condition from the assignment economy:

$$(1 - T_w[\omega(z)]) \omega_z h U_e = -U_e,$$

and the buyer’s first order condition:

$$V_z(S, z) = \omega_z.$$

Combining them gives:

$$(1 - T_w[\omega(z)]) V_z(S, z) h U_e = -U_e \quad (C.4)$$

Combining (C.3) and (C.4) gives:

$$\frac{T_w}{1 - T_w} [\omega] = p^\Phi \left\{ \left[ U_{ec} \left( -\frac{U_e}{U_c} \right) + U_{ec} \right] e + U_e \right\} \frac{h_v}{h} \left( -\frac{U_e}{U_c} \right)$$

$$+ (1 - \chi)(1 - M(w)) V_{zs} \left( -\frac{S_v}{S} \right) S h \left( -\frac{U_e}{U_c} \right).$$

Thus, after rearrangement (and use of the definitions of the distributions $G$ and $F$) we have:

$$T_w[\omega] = \frac{1 + (1 - \chi) \frac{1 - F(h)}{p^\phi} \frac{\xi^c}{1 + \xi^g} \frac{1 - G(S)}{g(S)S} V_{zs} \frac{f(h)h}{1 - F(h)}}{1 + \frac{1 - F(h)}{p^\phi} \frac{\xi^c}{1 + \xi^g} \frac{f(h)h}{1 - F(h)}}. \quad (C.5)$$

\[47\] Strictly, when the domain is $(0, 1]$, then these conditions hold as limits as $v \downarrow 0$.  

49
Or, equivalently,

\[ T_w[w] = \frac{1}{1 + \frac{1-F(h)}{p}\chi \left( 1 + \frac{V_{z}S}{\frac{g(S)}{f(h)}} \right)} \quad \text{(C.6)} \]

Completely differentiating the seller’s first order condition with respect to \( v \) gives:

\[ w_v = V_z\left( 1 + \chi \frac{h_c}{h} + \chi \frac{V_{z}S}{\frac{g(S)}{S}} \right) \quad \frac{1}{1 + \chi \left( \frac{T_{w,\chi}V_z}{w} + \frac{V_{z}S}{V_z} \right)} (1 - M(w)) \]

And so:

\[ \frac{1 - M(w)}{wm(w)} = \left( -\frac{V_z}{w} \left( 1 + \chi \frac{h_c}{h} + \chi \frac{V_{z}S}{\frac{g(S)}{S}} \right) \frac{T_{w,\chi}V_z}{w} + \frac{V_{z}S}{V_z} \right) (1 - M(w)) \]

Combining terms:

\[ T_w[w] = \frac{1}{1 + \frac{1-F(h)}{p}\chi \left( \frac{V_{z}S}{\frac{g(S)}{f(h)}} \right) \left( \frac{1-\chi}{\chi} \frac{V_{z}S}{\frac{g(S)}{S}} \right) \frac{1}{f(h)h}} \quad \text{(C.7)} \]

When the policymaker attaches unit weight to the buyer (\( \chi = 1 \)) this formula reduces to:

\[ T_w[w] = \frac{1}{1 + \frac{1-F(h)}{p}\chi \left( \frac{V_{z}S}{\frac{g(S)}{f(h)}} \right) \left( \frac{1-\chi}{\chi} \frac{V_{z}S}{\frac{g(S)}{S}} \right) \frac{1}{f(h)h}} \quad \text{(C.8)} \]

In the quasilinear/multiplicative preferences case (C.7) further simplifies to:

\[ T_w[w] = \frac{1}{1 + \frac{\frac{S_z}{w}}{\chi \left( \frac{\frac{V_{z}S}{\frac{g(S)}{f(h)}}}{\frac{1}{\chi} \frac{V_{z}S}{\frac{g(S)}{S}} \frac{1}{f(h)h}} \right)}} \quad \text{(C.9)} \]

If the policymaker attaches zero weight to buyers (C.7) reduces to:

\[ T_w[w] = \frac{1}{1 + \frac{1-F(h)}{p}\chi \left( \frac{V_{z}S}{\frac{g(S)}{f(h)}} \right) \left( \frac{1-\chi}{\chi} \frac{V_{z}S}{\frac{g(S)}{S}} \right) \frac{1}{f(h)h}} \quad \text{(C.10)} \]
In the case of quasilinear/multiplicative case (C.10) simplifies to:

\[
T_w[w] = \frac{1}{1 + \frac{S_z}{w} \left( \frac{\epsilon}{1 + \epsilon \frac{\partial w}{\partial w}} \right) \frac{m(w)w}{1 - M(w)} \left( 1 - \frac{(1-G(S))}{g(S)} \right)}.
\]

(C.11)

D  Computing the mechanical and behavioral impacts of local marginal tax rate changes in the assignment setting

Here we compute the mechanical and behavioral impacts of marginal tax rate changes in the assignment setting. We focus on the quasilinear/constant elasticity seller- multiplicative buyer preference setting. Let \( T \) denote a tax function and consider the following perturbed function:

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

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

Sellers (and buyers) can be partitioned into four groups which we label groups 0, 1, 2 and 3 respectively. Group 0 consists of the least productive sellers with types in \([v_0, 1] \). Their behavior is unaffected by the tax perturbation. Group 1 sellers with types in \([v_1(\delta), v_0) \) bunch at the kink point in the tax schedule. They earn \( w_0 \) and supply service \( z_0 \). The threshold \( v_1(\delta) \) satisfies:

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

Group 2 sellers have types \((v_2(\delta, \epsilon), v_1(\delta)) \) earn incomes between \( w_0 \) and \( w_0 + \epsilon \) and pay the higher marginal tax \( T_w[w] + \delta \). Their first order condition for service supply is given by:

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

where we use the buyer’s first order condition to replace the derivative of seller income with respect to service with \( S(v) \) and make the dependence of the functions \( \tilde{w} \) and \( \tilde{z} \) on \( \delta \) explicit. (Note that at a given \( v \) in this interval \( \epsilon \) does not affect the incomes earned and effective labor supplied by sellers and is not included as a parameter of \( \tilde{w} \)). The threshold \( v_2(\delta, \epsilon) \) is given by:

\[v_2(\delta, \epsilon) = \sup \{ v : \tilde{w}(v; \delta) \geq w_0 + \epsilon \}.
\]
Sellers in group 3 have ranks \( I \cap [0, v_2(\delta, \epsilon)] \). Their marginal taxes are determined by the original tax schedule and their first order conditions are given by:

\[
(1 - T_w[\tilde{w}(v; \delta, \epsilon)])S(v) = \Psi_2 \left( \frac{\tilde{z}(v; \delta, \epsilon)}{h(v)} \right).
\]

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

\[
S(v_2)\tilde{z}(v_2(\delta, \epsilon); \delta, \epsilon) - w(v_2(\delta, \epsilon); \delta, \epsilon) = S(v_2(\delta, \epsilon); \delta)\tilde{z}_+(v_2(\delta, \epsilon); \delta) - w_+(v_2(\delta, \epsilon); \delta),
\]

where \( \tilde{z}_+(v_2(\delta, \epsilon)) \) and \( \tilde{w}_+(v_2(\delta, \epsilon)) \) are, respectively, the right limits of \( \tilde{z} \) and \( \tilde{w} \) at \( v_2(\delta, \epsilon) \) (i.e. the limits of incomes earned and service supplied by group 2 sellers). Note that for \( v \) below \( v_2(\delta, \epsilon) \), \( \tilde{w} \) and \( \tilde{z} \) do depend on \( \epsilon \) and we make this dependence explicit in the notation.

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

\[
R(\epsilon, \delta) := \int_0^{v_2(\delta, \epsilon)} \{ T[\tilde{w}(v; \delta, \epsilon)] + \delta \epsilon \} dv + \int_{v_2(\delta, \epsilon)}^{v_0} \{ T[\tilde{w}(v; \delta)] + \delta[\tilde{w}(v; \delta) - w_0] \} dv. \tag{D.1}
\]

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

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

Differentiating this with respect to \( \delta \) and setting \( \delta = 0 \) gives:

\[
\epsilon[1 - M(v_2(0, \epsilon))] + \int_{v_2(0, \epsilon)}^{v_0} [\tilde{w}(v; 0) - w_0] dv.
\]

Differentiating this with respect to \( \epsilon \) and setting \( \epsilon \) to zero gives:

\[
1 - M(w_0).
\]

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

\[
\int_0^{v_2(\delta, \epsilon)} T[\tilde{w}(v; \delta, \epsilon)] dv + \int_{v_2(\delta, \epsilon)}^{v_0} T[\tilde{w}(v; \delta)] dv. \tag{D.2}
\]

The second term captures the behavioral effect stemming from group 1 and group 2 sellers (although the first of these turns out to be negligible). We first keep \( \epsilon \) fixed (and suppress it in the notation). Differentiating the second term with respect to \( \delta \) and setting \( \delta \) to 0 gives:

\[
-T[\tilde{w}(v_2(0))] \frac{\partial v_2}{\partial \delta} + \int_{v_2(0)}^{v_0} T_w[\tilde{w}(v)] \frac{\partial \tilde{w}}{\partial \delta}(v; 0) dv
\]
The first of these terms cancels against the corresponding term from the group 3 agents (see below). We focus on the second. We must compute 
\frac{\partial w}{\partial \delta}(v;0). Using the buyer’s first order condition \( w_v = S(v)z_v \) and integrating by parts:

\[
\begin{align*}
    w(v;\delta) &= w_0 - \int_v^{v_1(\delta)} S(v') \hat{z}_v(v';\delta) dv' \\
    &= w_0 - \left[ S(v_1(\delta)z_0 - S(v)\hat{z}(v;\delta) - \int_v^{v_1(\delta)} S_v(v') \hat{z}(v';\delta) dv' \right] \\
    &= w_0 - S(v_1(\delta))z_0 + S(v)\hat{z}(v;\delta) + \int_v^{v_1(\delta)} S_v(v') \hat{z}(v';\delta) dv'.
\end{align*}
\]

Thus, canceling terms and using \( v_1(0) = v_0 \),

\[
\frac{\partial w(v;0)}{\partial \delta} = S(v) \frac{\partial \hat{z}(v;0)}{\partial \delta} + \int_v^{v_0} S_v(v') \frac{\partial \hat{z}}{\partial \delta}(v';0) dv'.
\]

Now, we have from the group 2 seller’s first order condition:

\[
(1 - T_w[\hat{w}(v;\delta)] - \delta)S(v)h(v) - \Psi_e \left( \frac{\hat{z}(v;\delta)}{h(v)} \right) = 0.
\]

Totally differentiating this and evaluating at \( \delta = 0 \) gives:

\[
-T_{ww}[\hat{w}(v)]S(v)h(v) \frac{\partial \hat{w}}{\partial \delta}(v;0) - \Psi_{ee} \left( \frac{\hat{z}(v)}{h(v)} \right) \frac{1}{h(v)} \frac{\partial \hat{z}}{\partial \delta}(v) - S(v)h(v) = 0.
\]

Substituting and rearranging:

\[
-\frac{\partial \hat{z}(v)}{\partial (1 - T_w[\hat{w}(v)])} = \frac{\partial \hat{z}(v;0)}{\partial \delta} = -\mathcal{E} \mathcal{T}(v) \frac{\hat{z}(v)}{\hat{w}(v)} \frac{\partial \hat{w}}{\partial \delta}(v) - \frac{\mathcal{E}}{1 - T_w[\hat{w}(v)]} \hat{z}(v),
\]

where \( \mathcal{T}(v) = \frac{T_{ww}[\hat{w}(v)]}{1-T_w[\hat{w}(v)]} \). There are two special cases. If \( T_{ww} = 0 \), as we considered previously, then:

\[
1 - T_w[\hat{w}(v)] \frac{\partial \hat{z}(v)}{\partial (1 - T_w[\hat{w}(v)])} = \mathcal{E}.
\]

In the other special case \( \frac{\partial \hat{w}}{\partial \delta} = \frac{\partial \hat{z}}{\partial \delta} \) and \( \hat{z} = \hat{w} \). This is considered in Saez. Then:

\[
1 - T_w[\hat{w}(v)] \frac{\partial \hat{z}(v)}{\partial (1 - T_w[\hat{w}(v)])} = \frac{\mathcal{E}}{1 + \mathcal{E} \mathcal{T}}.
\]

However, in our case:

\[
\frac{\partial \hat{w}}{\partial \delta}(v;0) = -\mathcal{E} \mathcal{T}(v) \frac{S(v)\hat{z}(v)}{\hat{w}(v)} \frac{\partial \hat{w}}{\partial \delta}(v;0) - \frac{\mathcal{E}}{1 - T_w[\hat{w}(v)]} S(v)\hat{z}(v)
\]

\[
- \int_v^{v_0} S_v(v') \left[ \mathcal{E} \mathcal{T}(v') \frac{S(v')\hat{z}(v')}{\hat{w}(v')} \frac{\partial \hat{w}}{\partial \delta}(v';0) + \frac{\mathcal{E} S(v')\hat{z}(v')}{1 - T_w[\hat{w}(v')]} \right] dv'.
\]

53
At \( v = v_0 \) this formula simplifies to:

\[
\frac{\partial \tilde{w}}{\partial \delta}(v_0;0) = -\mathcal{E} T(v_0) \frac{S(v_0) z_0}{w_0} \frac{\partial \hat{w}}{\partial \delta}(v_0;0) - \frac{\mathcal{E}}{1 - T_w[w_0]} S(v_0) z_0 \\
= - \frac{\mathcal{E} S(v_0) z_0}{1 - T_w[w_0]} \frac{1}{1 + \mathcal{E} T(v_0) \frac{S(v_0) z_0}{w_0}}.
\]

Now, returning to:

\[
\int_{v_0} v^2(T[v]) \frac{\partial \tilde{w}}{\partial \delta}(v;0) dv,
\]

suppressing the dependence on \( \delta \) in the notation, differentiating with respect to \( \varepsilon \) and setting \( \varepsilon \) to zero gives:

\[
\frac{\partial}{\partial \varepsilon} \int_{v_0} v^2(T[v]) \frac{\partial \hat{w}}{\partial \delta}(v) dv = -T_w[w_0] \frac{\partial v^2}{\partial \varepsilon}(0) \frac{\partial \hat{w}}{\partial \delta}(v_0) = m(w_0).
\]

Recall that \( \hat{w}(v_2(\varepsilon)) = w_0 + \varepsilon \), we have:

\[
\frac{\partial v^2}{\partial \varepsilon}(0) = \frac{1}{\hat{w}_v(v_0)} = -m(w_0).
\]

Putting things together:

\[
\frac{\partial}{\partial \varepsilon} \int_{v_0} v^2(T[v]) \frac{\partial \hat{w}}{\partial \delta}(v) dv = -T_w[w_0] \frac{\mathcal{E} S(v_0) z_0}{1 - T_w[w_0]} \frac{1}{1 + \mathcal{E} T(v_0) \frac{S(v_0) z_0}{w_0}} m(w_0).
\]

We now turn to the first term in (D.2). This concerns the behavioral impact on tax revenues collected from group 3 agents. Differentiating this term with respect to \( \delta \) and setting \( \delta \) to 0 gives:

\[
T[\hat{w}(v_2(0))] \frac{\partial v^2}{\partial \delta} + \int_{v_0} v^2(0) T_w[v]\frac{\partial \hat{w}}{\partial \delta}(v;0) dv.
\]

The first term cancels against the corresponding term from the group 2 agents and we are left with:

\[
\int_{v_0} v^2(0) T_w[v]\frac{\partial \hat{w}}{\partial \delta}(v;0) dv.
\]

Similar to before:

\[
\hat{w}(v;\delta) = \hat{w}(v_2(\delta);\delta) - S(v_2(\delta)) \tilde{z}(v_2(\delta);\delta) + S(v) \tilde{z}(v;\delta) + \int_{v} v^2(\delta) S_v(v') \tilde{z}(v';\delta) dv'.
\]
Differentiating with respect to $\delta$, evaluating at $\delta = 0$ and suppressing dependence on $\delta = 0$ in the notation gives:

\[
\frac{\partial \hat{w}}{\partial \delta}(v) = \frac{\partial \hat{w}}{\partial \delta}(v_2) + \left[ \hat{w}_v(v_2) - S_v(v_2) \hat{z}(v_2) - S(v_2)z^*_v(v_2) \right] \frac{\partial v_2}{\partial \delta} \\
- S(v_2) \frac{\partial \hat{z}}{\partial \delta}(v_2) + S(v) \frac{\partial \hat{z}}{\partial \delta}(v) \\
+ S_v(v_2) \hat{z}(v_2) \frac{\partial v_2}{\partial \delta} + \int_{v_2}^{v} S_v(v') \frac{\partial \hat{z}}{\partial \delta}(v') dv'.
\]

Using $S(v_2) \hat{z}_v(v_2) = \hat{w}_v(v_2)$, the previous equation simplifies to:

\[
\frac{\partial \hat{w}}{\partial \delta}(v) = \frac{\partial \hat{w}}{\partial \delta}(v_2) - S(v_2) \frac{\partial \hat{z}}{\partial \delta}(v_2) + S(v) \frac{\partial \hat{z}}{\partial \delta}(v) + \int_{v_2}^{v} S_v(v') \frac{\partial \hat{z}}{\partial \delta}(v') dv'.
\]

Now, for group 3 agents:

\[
(1 - T_v[\hat{w}(v; \delta)]) S(v) h(v) - \Psi_e \left( \frac{\hat{z}(v; \delta)}{h(v)} \right) = 0.
\]

Totally differentiating this and evaluating at $\delta = 0$ gives:

\[
-T_v[\hat{w}(v)] S(v) h(v) \frac{\partial \hat{w}}{\partial \delta}(v; 0) - \Psi_{ee} \left( \frac{\hat{z}(v)}{h(v)} \right) \frac{1}{h(v)} \frac{\partial \hat{z}}{\partial \delta}(v; 0) = 0.
\]

Substituting and rearranging:

\[
\frac{-\frac{\partial \hat{z}(v)}{\partial (1 - T_v[\hat{w}(v)])}}{-\frac{\partial \hat{z}(v; 0)}{\partial (1 - T_v[\hat{w}(v)])}} = \frac{\partial \hat{z}}{\partial \delta}(v; 0) = -\mathcal{E} \frac{T_v[\hat{w}(v)] \hat{w}(v) \hat{z}(v)}{1 - T_v[\hat{w}(v)]} \frac{\partial \hat{w}}{\partial \delta}(v; 0).
\]

Hence,

\[
\left[ 1 + \mathcal{E} T(v) \frac{S(v) \hat{z}(v)}{\hat{w}(v)} \right] \frac{\partial \hat{w}}{\partial \delta}(v; 0) = \left[ 1 + \mathcal{E} T(v) \frac{S(v_2) \hat{z}(v_2)}{\hat{w}(v_2)} \right] \frac{\partial \hat{w}}{\partial \delta}(v_2; 0)
\]

\[
- \int_{v_2}^{v_2(0)} \frac{S_v(v') \mathcal{E} T(v') \frac{S'(v') \hat{z}(v')}{\hat{w}(v')}}{S(v')} \frac{\partial \hat{w}}{\partial \delta}(v'; 0) dv'.
\]

Let:

\[
N(v) := \frac{S_v(v') \mathcal{E} T(v) \frac{S(v) \hat{z}(v)}{\hat{w}(v)}}{1 + \mathcal{E} T(v) \frac{S(v) \hat{z}(v)}{\hat{w}(v)}}
\]

\[
y(v) := - \int_{v}^{v_2(0)} \left\{ \frac{S_v(v') \mathcal{E} T(v') \frac{S'(v') \hat{z}(v')}{\hat{w}(v')}}{S(v')} \right\} \frac{\partial \hat{w}}{\partial \delta}(v'; 0) dv'
\]

\[
A := \left[ 1 + \mathcal{E} T(v_2) \frac{S(v_2) \hat{z}(v_2)}{\hat{w}(v_2)} \right] \frac{\partial \hat{w}}{\partial \delta}(v_2; 0).
\]
Thus, from the definition of $y$,

$$\frac{dy}{dv} = \left\{ \frac{S_v(v)}{S(v)} \mathcal{E}(v) \frac{S(v)\tilde{z}(v)}{\tilde{w}(v)} \right\} \frac{\partial \tilde{w}(v)}{\partial \delta} (v; 0).$$

And from (D.3):

$$\frac{1}{N(v)} \frac{dy}{dv} = A + y(v).$$

Integrating this last equation and using $y(v_2(0)) = 0$ gives,

$$y(v) = A \exp \left( - \int_v^{v_2(0)} N(v')dv' \right) - A.$$

And so:

$$\frac{dy}{dv} = AN(v) \exp \left( - \int_v^{v_2(0)} N(v')dv' \right).$$

Thus, using the definitions of $y$, $A$ and $N$:

$$\frac{\partial \tilde{w}}{\partial \delta}(v; 0) = \frac{1 + \mathcal{E}(v_2)\frac{S(v_2)\tilde{z}(v_2)}{\tilde{w}(v_2)}}{1 + \mathcal{E}(v)\frac{S(v)\tilde{z}(v)}{\tilde{w}(v)}} \exp \left\{ - \int_v^{v_2(0)} \frac{S(v')\mathcal{E}(v')\frac{S(v')\tilde{z}(v')}{\tilde{w}(v')}}{1 + \mathcal{E}(v')\frac{S(v')\tilde{z}(v')}{\tilde{w}(v')}} dv' \right\} \frac{\partial \tilde{w}}{\partial \delta} (v_2; 0).$$

Substituting we have:

$$\int_0^{v_2(0)} T_w[\tilde{w}(v)] \frac{\partial \tilde{w}}{\partial \delta}(v; 0) dv = \left\{ \frac{1 + \mathcal{E}(v_2)\frac{S(v_2)\tilde{z}(v_2)}{\tilde{w}(v_2)}}{1 + \mathcal{E}(v)\frac{S(v)\tilde{z}(v)}{\tilde{w}(v)}} \right\} \frac{\partial \tilde{w}}{\partial \delta} (v_2)$$

$$\times \int_0^{v_2(0)} \frac{T_w[\tilde{w}(v)]}{1 + \mathcal{E}(v)\frac{S(v)\tilde{z}(v)}{\tilde{w}(v)}} \exp \left\{ - \int_v^{v_2(0)} \frac{S(v')\mathcal{E}(v')\frac{S(v')\tilde{z}(v')}{\tilde{w}(v')}}{1 + \mathcal{E}(v')\frac{S(v')\tilde{z}(v')}{\tilde{w}(v')}} dv' \right\} dv.$$

We now turn to the determination of $\frac{\partial \tilde{w}(v_2)}{\partial \delta}$. First note that the buyers’ incentive conditions at and close to $v_2$ imply:

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

And so:

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

Totally differentiating with respect to $\delta$ (and suppressing $\delta$ in the notation) gives:

$$\frac{\partial \tilde{w}}{\partial \delta} (v_2) = S(v_2) \left[ \frac{\partial \tilde{z}}{\partial \delta} (v_2) - \frac{\partial \tilde{z}_+}{\partial \delta} (v_2) \right] + \frac{\partial \tilde{w}_+}{\partial \delta} (v_2) + \frac{\partial v_2}{\partial \delta} K.$$
It is easy to check that evaluated at \( \delta = 0, \ K = 0 \). Then using our earlier result that at \( \delta = 0 \):

\[
\frac{\partial z}{\partial \delta}(v_2) = -\mathcal{E} T(v_2) \frac{\tilde{z}(v_2)}{\tilde{w}(v_2)} \frac{\partial \tilde{w}}{\partial \delta}(v_2)
\]

and:

\[
\frac{\partial z^+}{\partial \delta}(v_2) = -\mathcal{E} T(v_2) \frac{\tilde{z}(v_2)}{\tilde{w}(v_2)} \frac{\partial \tilde{w}^+}{\partial \delta}(v_2) - \frac{\mathcal{E} \tilde{z}(v_2)}{1 - T_w[\tilde{w}(v_2)]}
\]

we have:

\[
\left\{ 1 + \mathcal{E} T(v_2) \frac{S(v_2) \tilde{z}(v_2)}{\tilde{w}(v_2)} \right\} \frac{\partial \tilde{w}}{\partial \delta}(v_2) = \left\{ 1 + \mathcal{E} T(v_2) \frac{S(v_2) \tilde{z}(v_2)}{\tilde{w}(v_2)} \right\} \frac{\partial \tilde{w}^+}{\partial \delta}(v_2) + \frac{\mathcal{E} S(v_2) \tilde{z}(v_2)}{1 - T_w[\tilde{w}(v_2)]}
\]

Returning to our definition of \( \frac{\partial \tilde{w}}{\partial \delta} \) for group 2 agents and substituting for \( \frac{\partial \tilde{w}^+}{\partial \delta} \) gives

\[
\left\{ 1 + \mathcal{E} T(v_2) \frac{S(v_2) \tilde{z}(v_2)}{\tilde{w}(v_2)} \right\} \frac{\partial \tilde{w}}{\partial \delta}(v_2) = - \int_{v_2}^{v_0} \frac{S(v')}{S(v)} \left[ \mathcal{E} T(v') \frac{S(v') \tilde{z}(v')}{\tilde{w}(v')} \frac{\partial \tilde{w}}{\partial \delta}(v') + \frac{\mathcal{E} S(v') \tilde{z}(v')}{1 - T_w[\tilde{w}(v')]} \right] dv'
\]

Denote the right hand side by \( D(v_2) \). Note that \( D(v_0) = 0 \) and using our previous formula for \( \frac{\partial \tilde{w}}{\partial \delta}(v_0) \) gives:

\[
\frac{\partial D}{\partial v_2}(v_0) = \frac{S(v_0)}{S(v_0)} \left[ \mathcal{E} T(v_0) \frac{S(v_0) \tilde{z}(v_0)}{\tilde{w}(v_0)} \frac{\partial \tilde{w}}{\partial \delta}(v_0) + \frac{\mathcal{E} S(v_0) \tilde{z}(v_0)}{1 - T_w[\tilde{w}(v_0)]} \right] = \frac{S(v_0)}{S(v_0)} \frac{\mathcal{E} S(v_0) \tilde{z}(v_0)}{1 - T_w[\tilde{w}(v_0)]} \frac{1}{1 + \mathcal{E} T(v_0) \frac{S(v_0) \tilde{z}(v_0)}{\tilde{w}(v_0)}}.
\]

We can write (making the dependence of \( v_2 \) on \( \varepsilon \) explicit):

\[
\int_0^{v_2(\varepsilon)} T_w[\tilde{w}(v)] \frac{\partial \tilde{w}}{\partial \delta}(v; \varepsilon) dv
\]

\[
= D(v_2(\varepsilon)) \int_0^{v_2(\varepsilon)} \frac{T_w[\tilde{w}(v)]}{1 + \mathcal{E} T(v) \frac{S(v) \tilde{z}(v)}{\tilde{w}(v)}} \exp \left\{ - \int_{v_2(\varepsilon)}^{v_0} \frac{S(v')}{S(v)} \mathcal{E} T(v') \frac{S(v') \tilde{z}(v')}{\tilde{w}(v')} dv' \right\} dv
\]

Differentiating this expression with respect to \( \varepsilon \) and evaluating at \( \varepsilon = 0 \) (and using \( v_2(0) = v_0 \) and \( D(v_2(0)) = D(v_0) = 0 \)), gives:

\[
\frac{\partial v_2}{\partial \varepsilon} \frac{\partial D}{\partial v_2} \int_0^{v_0} \frac{T_w[\tilde{w}(v)]}{1 + \mathcal{E} T(v) \frac{S(v) \tilde{z}(v)}{\tilde{w}(v)}} \exp \left\{ - \int_{v_0}^{v_2(\varepsilon)} \frac{S(v')}{S(v)} \mathcal{E} T(v') \frac{S(v') \tilde{z}(v')}{\tilde{w}(v')} dv' \right\} dv
\]
Further using $\hat{\omega}(v_2(\epsilon)) = w_0 + \epsilon$ so that $\hat{\omega} \frac{\partial \hat{\omega}}{\partial \epsilon} = 1$ (and recalling that $w_0^* = -\frac{1}{m(\hat{\omega})}$ and $S_v = -\frac{1}{g(S)}$),

$$
\frac{\partial}{\partial \epsilon} \int_0^{v_0} T_w[\hat{\omega}(v)] \frac{\partial \hat{\omega}}{\partial \delta}(v;0) dv
\begin{array}{c}
= \frac{\partial v_2}{\partial \epsilon} \frac{\partial D}{\partial v_2}(v_0) \int_0^{v_0} \frac{T_w[\hat{\omega}(v)]}{1 + \mathcal{E}T(v)} \frac{S(v_0)\overline{z}(v)}{\hat{\omega}(v)} \exp \left \{ - \int_v^{v_0} \frac{S(v')}{S(v')} \mathcal{E}T(\epsilon) \frac{S(v')\overline{z}(v')}{\hat{\omega}(v')} dv' \right \} dv
\end{array}
\begin{array}{c}
= m(w_0) \frac{1}{g(S(v_0))} \frac{\mathcal{E}S(v_0)z_0}{1 - T_w[w_0]} \frac{1}{1 + \mathcal{E}T(v_0)} \frac{S(v_0)z_0}{\hat{\omega}(v_0)}
\end{array}
\begin{array}{c}
\times \int_0^{v_0} \frac{T_w[\hat{\omega}(v)]}{1 + \mathcal{E}T(v)} \frac{S(v_0)\overline{z}(v)}{\hat{\omega}(v)} \exp \left \{ - \int_v^{v_0} \frac{S(v')}{S(v')} \mathcal{E}T(\epsilon) \frac{S(v')\overline{z}(v')}{\hat{\omega}(v')} dv' \right \} dv
\end{array}
$$

Adding the two behavioral terms gives (42) in the main text. Adding the mechanical term as well and rearranging gives

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

where $\hat{\mathcal{E}}w(w_0)$ is defined as in (43) in the main text. Comparing this equation to (C.11) in Appendix C immediately gives that:

$$
\hat{\mathcal{E}}w(w_0) := \frac{\mathcal{E}S(v_0)z_0}{1 + \mathcal{E}T(v_0)} \frac{1}{\frac{S(v_0)z_0}{\hat{\omega}(v_0)} - 1} \left \{ 1 - \frac{1 - g(S(v_0))}{g(S(v_0))S(v_0)} \right \}.
$$

Using the expressions for $\frac{\partial z}{\partial \delta}$ and $\frac{\partial w}{\partial v}$ previously derived and letting $\hat{\Pi}(v;\delta) := S(v)\overline{z}(v;\delta) - \hat{\omega}(v;\delta)$, we obtain:

$$
\frac{\partial \hat{\Pi}}{\partial \delta}(v_0) = 0,
$$

and for $v \in I \cap [0, \hat{v}_0)$,

$$
\frac{\partial \hat{\Pi}}{\partial \delta}(v) = - \left [ 1 + \mathcal{E} \frac{T_{ww}}{1 - \mathcal{E} \frac{Sz}{w}} \right ] \frac{\partial \hat{\omega}}{\partial \delta}.
$$

Thus,

$$
\int_0^{v_2(\epsilon)} \frac{\partial \hat{\Pi}}{\partial \delta}(v;\epsilon) dv = -D(v_2(\epsilon)) \int_0^{v_2(\epsilon)} \mathcal{E}T(v) \frac{S(v)^*z^*(v)}{w^*(v)} dv' dv.
$$
For $\chi \neq 0$, the behavioral term is augmented with the expression:

$$\chi \frac{\partial v_2}{\partial D} \int_0^{v_0} \exp \left\{ -\int_0^{v_0} \frac{S(v') \mathcal{E} \mathcal{T}(v') S(v') z^*(v')}{1 + \mathcal{E} \mathcal{T}(v')} dv' \right\} dv$$

$$= m(w_0) \frac{1}{g(S(v_0)) S(v_0)} \frac{1}{1 - T_w[w_0]} \frac{1}{1 + \mathcal{E} \mathcal{T}(v_0) \frac{S(v_0) z_0}{w_0}}$$

$$\times \int_0^{v_0} \exp \left\{ -\int_0^{v_0} \frac{S(v') \mathcal{E} \mathcal{T}(v') S(v') z^*(v')}{1 + \mathcal{E} \mathcal{T}(v')} dv' \right\} dv.$$

### E Appendix for Section 4

#### E.1 On the Pareto tail parameter

In Figures 1 and 2(a) we display the maximum likelihood estimates of the Pareto tail coefficient for CEO compensation. In this section we compare these estimates to ones reported previously in the literature and to our own based on analysis of the entire population.

Saez (2001) plots $\int_0^{w_0} w(m) dw / w_0$ (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 et al. (2013) using the World Top Incomes Database (WTID) report a value of 1.6 for the US in 2013. Subsequently, Diamond and Saez (2011) and Piketty et al. (2014) both use a value of the Pareto tail index equal to 1.5. Overall estimates using CEO data are higher than these values reported in the literature, indicating a more compact distribution of CEO compensation relative to the distribution of income of the entire population.

However it should be noted that the above values for the tail parameters 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: 90th, 95th, 99th, 99.5th, 99.9th and

---

48 topincomes.parisschoolofeconomics.eu
99.99th) both the fraction of total income due to earned income and the conditional average total income for that income group. We assume that for all individuals in a given income group $i$, earned income is related to total income according to: $w = \rho_i \cdot y$. Given information on the average $y$ within an income group $i$ we can then recover the conditional average for earned compensation $\bar{w}_i$ within the group. If the income distribution has a Pareto tail, it follows that the threshold earned income value for group $i$, $w_i$, is related to $\bar{w}_i$ according to:

$$\bar{w}_i = \frac{\alpha_w}{\alpha_w - 1} w_i.$$ 

The Pareto assumption further implies:

$$w_i = \frac{\bar{w}}{(1 - P_i)^{\frac{1}{\alpha_w}}}$$

where $P_i$ is the fraction of agents below $w_i$. Then considering two percentile categories and simplifying we obtain an estimate for $\alpha_w$ equal to:

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

In Figure 7 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.

E.2 Proof of Lemma 1

Proof. We first consider the case in which $F(h)$ is a GEV with parameter $\eta = 0$. In
this case the density $f(h)$ is given by $f(h) = F(h) \exp(-h)$. Hence both the numer-
ator and the denominator of the Pareto coefficient go to zero as $x$ becomes large.
Applying l'Hôpital’s rule we obtain:

$$\lim_{h \to \infty} \frac{F(h)e^{-h}h}{1 - F(h)} = \lim_{h \to \infty} \frac{f(h)e^{-h}h + F(h)e^{-h}[1 - h]}{-f(h)} = \lim_{h \to \infty} \frac{h}{\exp\{h\}} + h - 1 = \infty.$$ 

Consider now the case $\eta \neq 0$. Similar to before we have that:

$$\lim_{h \to \infty} \frac{f(h)h}{1 - G(h)} = \frac{\frac{d}{dh}\left\{F(h)\frac{h}{(1+\eta h)^{1+1/\eta}}\right\}}{-f(h)}.$$ 

Computing the derivative gives:

$$\lim_{h \to \infty} \frac{f(h)h}{1 - F(h)} = \lim_{h \to \infty} \frac{h - 1}{1 + \eta h} - \frac{h}{(1 + \eta h)^{1+1/\eta}}.$$ 

If $\eta > 0$ the second term goes to zero and the first term goes to $1/\eta$. If $\eta < 0$ the
second term goes to infinity. 

E.3 Robustness of tail index estimates

In this appendix we conduct additional robustness tests on the estimates of the tail
index parameters for the talent and firm asset distribution.

Sensitivity with respect to parameter values We first look at the effect of $\lambda,
\theta$ and $g$ on our estimates. For the entire set of parameter values we consider we
confirm the key finding in the body of the paper that the talent distribution belongs
to the domain of attraction of the Weibull distribution and the asset distribution
belongs to the domain of attraction of the Frechét distribution. We begin with $\lambda$.
This parameter controls the longevity of the impact of CEO effective labor on firm’s
surplus. The benchmark calibration uses a value of $\lambda = 0.5$ this value implies a
“half life” of CEO impact of $\ln(2)/\ln(1 + \lambda) = 1.71$ years. We extend the range for $\lambda$ from $0.05$ to $2$ implying a half life ranging from slightly above 14 years to 6 months.
As the figs. 8(a) and 9(a) show, the estimates for the $\eta$ for the talent and firm asset
distribution move only slightly as we change $\lambda$. Related results are obtained for
the remaining parameters. We first change the range for our parameter $\theta$ from 0.2 to 0.8 (Terviö (2008) considers 0.8 an extreme upper bound for $\theta$). The results are shown in Figures 8(b) and 9(b). Finally robustness with respect to the parameter $g$ is shown in Figures 8(c) and 9(c) here we maintain $g < r$.

![Graphs showing results for different parameter values]

Figure 8: GEV distribution estimates for the talent distribution. 95 percent confidence intervals represented by dashed lines.

![Graphs showing results for different parameter values]

Figure 9: GEV distribution estimates for the firm asset distribution. 95 percent confidence intervals represented by dashed lines.

Sensitivity with respect to estimation strategies  We now go beyond our parametric GEV estimates and explore alternative extreme value estimators of the limiting Pareto coefficient. They corroborate the parametric maximum likelihood estimators reported in the main text.

In the extreme value literature, the Hill estimator is often used to obtain empirical values for the limiting Pareto coefficient. Given data $X = \{X_i\}_{i=1}^n$ and for a given value of $k \in \mathbb{N}$, the Hill estimator is:

$$
\hat{a}_x = \frac{1}{\hat{\eta}} = \left(\frac{1}{k} \sum_{j=1}^{k} \log X_{j,n} - \log X_{k,n}\right)^{-1},
$$

(E.2)
where $X_{j,n}$ denotes the $j$-th highest value of $X$. However, application of the Hill estimator requires that $\eta > 0$. We wish to allow for the possibility that $\eta \leq 0$ (and that the talent distribution is within the maximum domain of attraction of the Weibull or Gumbel distributions). Consequently, we also consider the Pickands estimator that is applicable in this case (see Dekkers and De Haan (1989)). The Pickands estimator provides an estimate for the reciprocal of the limiting Pareto coefficient (sometimes called the extreme value index). For given data $X = \{X_i\}_{i=1}^n$ and for a given value of $k \in \mathbb{N}$, the Pickands estimator is:

$$\hat{\eta} = \frac{1}{\log 2} \log \left( \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \right).$$

(E.3)

Figure 10 displays our values for both estimators applied to our recovered talent data over time. The Hill approach gives extremely high point estimates for the limiting Pareto coefficient of over 100 on average and 250 by the end of the sample. The Pickands approach gives point estimates of $\eta$ that are small and less than zero suggesting that the distribution of $h$ lies in the maximum domain of attraction of the Weibull distribution.

---

49 The value of $k$ can be chosen arbitrary within the range $\{1, \ldots, \tilde{k}\}$, with $\tilde{k}$ the largest integer strictly less than $N$ for the Hill estimator and $N/4$ for the Pickands estimator. We provide values of estimators averaged over $k \in [1,30]$. 

63