

# The Taxation of Superstars\*

Florian Scheuer  
Stanford University

Iván Werning  
MIT

February 2016

How are optimal taxes affected by superstar phenomena? To answer this question, we extend the Mirrlees model to incorporate an assignment problem in the labor market that generates superstar effects. Perhaps surprisingly, rather than providing a rationale for higher taxes, we show that superstar effects provide a force for *lower* marginal taxes conditional on the observed distribution of earnings. Superstar effects make the earnings schedule convex, which increases the responsiveness of individual earnings to tax changes. We show that various common elasticity measures must be adjusted upwards in optimal tax formulas. Finally, we study a comparative static that does not keep the observed earnings distribution fixed: when superstar technologies are introduced, inequality increases but we obtain a neutrality result, finding optimal taxes unaltered.

## 1 Introduction

Top earners make extraordinary figures, with differences in pay that, as one moves up the scale, eclipse any apparent differences in skill. One way economists have sought to explain this phenomenon is by appealing to superstar effects—to use the term coined by [Rosen \(1981\)](#) in the seminal contribution on the topic. According to this theory, better workers end up at better jobs, with greater complementary resources at their disposal, and this magnifies workers' innate skill differences. In principle, due to superstar effects, someone barely 5% more productive, at any given task, may earn 10 times more.

Models with such superstar effects have therefore been widely used to explain the observed income inequality in a variety of labor markets. For example, [Gabaix and Landier \(2008\)](#) and [Terviö \(2008\)](#) have shown that they are extremely successful in explaining the empirical distribution of CEO compensation.<sup>1</sup> Moreover, there are reasons to believe that the impact of superstar effects on inequality has been growing recently, e.g. due to globalization

---

\*In memory of Sherwin Rosen. We thank numerous superstar economists for useful comments.

<sup>1</sup>For recent overviews, see e.g. [Edmans and Gabaix \(2015\)](#) and [Garicano and Rossi-Hansberg \(2015\)](#).

and the rising importance of information technology, which provide increasing leverage for highly skilled individuals. This raises the question: should superstars pay superstar taxes?

Our goal is to answer this question by studying the taxation of earnings in the presence of superstar effects. Following [Mirrlees \(1971\)](#), a vast theoretical and applied literature has provided insights into the forces that should shape tax schedules. Particular attention has been given to taxes at the top. Given this focus, the absence of superstar effects in the analyses is, potentially, an important omission. Our first contribution is to incorporate these effects into the standard optimal tax model.

With earnings in disproportion to inherent skills, it may appear intuitive that superstar effects tilt the calculus balancing efficiency and equality, exacerbating inequality, and leading to higher taxes. Perhaps surprisingly, our results are the opposite. Depending on the exercise, we find that superstar effects are either neutral or provide a force for lower taxes.

It is useful to break up the broader question regarding the taxation of superstars by posing two distinct and sharper questions, as follows:

**Question I.** Given a tax schedule and an observed distribution of earnings, how do the conditions for the efficiency of this tax schedule depend on superstar effects?

**Question II.** Given a distribution of skills and preferences, how are optimal tax rates affected by superstar effects that may affect the distribution of earnings?

Question I holds the observed distribution of earnings fixed. The motivation for doing so is that the earnings distribution is an empirical datum to policymakers at any moment. Indeed, the earnings distribution has been recognized as a key input into optimal tax formulas in standard analyses without superstars ([Saez, 2001](#)). If anything, what is not directly observable is the extent to which, in the background, superstar effects shape the observed distribution. Hence, it is imperative to consider hypothetical scenarios that turn superstar effects on and off while holding the earnings distribution fixed. Doing so establishes whether superstar effects matter per se, in addition to their impact on the distribution of earnings.

In contrast, Question II is based on a comparative static, one that envisions a change in technology holding other primitives fixed. As a result, the distribution of earnings is no longer held fixed: skill differences are potentiated and inequality rises when superstar effects emerge. Such a comparative static exercise becomes policy relevant if one wishes to anticipate future changes in technology driven by superstar phenomena. Indeed, many current discussions of inequality and technological change are cast in terms of a trend or adjustment process that is still ongoing, as in the literature on the growing importance of information technology and automation. For such hypothetical extrapolations, a comparative static result is precisely what is required.

We obtain three main results, of which the first two are geared towards Question I.

**Result 1.** The conditions for the efficiency of a tax schedule are *unchanged* relative to the standard Mirrlees model when expressed in terms of two sufficient statistics: the distribution of earnings and the earnings elasticities with respect to individual tax changes.

According to this result, superstar effects do not directly affect whether taxes are optimal or suboptimal, as long as we condition on the observed distribution of earnings and on the elasticities of earnings with respect to individual tax changes, at each point in this distribution. These are the same statistics that emerge in the standard Mirrlees setting, where superstar effects are absent, and the conditions for efficient taxes are identical.

We view this neutrality result as an important conceptual benchmark that stresses that, at least in theory, with just the right information nothing needs to be changed. As our second result stresses, in practice, however, things are more challenging. In particular, unless one obtains the elasticity of earnings off of superstar workers, using *individual* tax changes that do not give rise to general equilibrium effects, one must recognize that the required earnings elasticities are affected by the presence of superstar effects.

**Result 2.** Superstar effects *increase* earnings elasticities relative to measures that omit superstar effects; this provides a force for *lower* taxes for a given distribution of earnings.

According to this result, typical elasticities must be adjusted upwards and we provide formulas for undertaking such adjustments. As usual, higher elasticities provide a force for lower tax rates. In particular, the top tax rate should be lower.<sup>2</sup>

Why is it that superstar effects increase individual earnings elasticities? There are several forces at work, the most basic of which is based on the Le Chatelier principle. In assignment models, workers sort into jobs and the equilibrium earnings schedule reflects the earnings across these different jobs. The availability of different job options tends to make the earnings schedule convex. This, in turn, increases the behavioral response to any tax change. Intuitively, a worker induced to provide greater effort by way of lower taxes anticipates being matched with a better job, with better pay, and this further amplifies the incentive for effort.<sup>3</sup> Thus, earnings elasticities that do not correctly capture this reallocation, which we call “fixed-assignment” elasticities, are too low.

An additional, mechanical bias arises if one starts with earnings elasticities for non-superstar workers, for whom earnings are proportional to effort. This then identifies what we call “effort” elasticities. Suppose we try to extrapolate such effort elasticities to workers in superstar positions. Because earnings are a convex function of effort for superstar workers, any percentage change in effort induces a greater percentage change in earnings. As a result, the earnings elasticity must also be adjusted upwards relative to the effort elasticity.

---

<sup>2</sup>This does not imply that taxes should be low in absolute terms, or lower than current levels. It establishes a downward force on optimal taxes, coming from superstar effects, compared to an economy with no superstars.

<sup>3</sup>As we discuss further below, we adopt a broad notion of effort here.

Finally, we require elasticities with respect to individual tax changes. In contrast, often one starts with evidence from economy-wide tax changes, or “macro” elasticities, which induce general equilibrium effects. We show that for progressive tax schedules these general equilibrium effects create a further bias. In our benchmark cases, the macro elasticity either coincides with the effort elasticity, or is even further downward biased.

This discussion underscores that properly measuring the earnings elasticity required for Result 1 demands a number of conditions to be met. First, the tax change cannot be temporary or unexpected—so that the workers have time for any desired job reallocation. Second, we must observe a superstar worker—to make sure we estimate an earnings rather than effort elasticity. Third, tax changes must be individual in nature—to avoid general equilibrium effects in the earnings schedule. Otherwise, Result 2 provides the necessary upward adjustment factors for each of the three elasticity measures discussed above.

Our approach of expressing the optimality conditions in terms of sufficient statistics, namely the distribution and elasticities of earnings, allows us to use the parametrization of [Gabaix and Landier \(2008\)](#), who provide an application to CEOs, to inform the required adjustments. They turn out quantitatively significant. For instance, for an elasticity of .25 that ignores job reassignment, the correct earnings elasticity is 1. As discussed, even greater adjustments are required for effort and macro elasticities.

Our third result provides a sharp answer to Question II, the comparative static that introduces superstar effects without holding the distribution of earnings fixed.

**Result 3.** Superstars are *neutral* with respect to comparative statics: when holding preferences and the distribution of skills fixed, the conditions for the efficiency of a marginal tax rate schedule do not depend on whether superstar effects are present.

This exercise considers a change in technology, introducing superstar effects, which generally increases earnings inequality. The neutrality result is based on the fact that the conditions for Pareto efficiency can be expressed in terms of the primitive distribution of skills and the elasticity of effort, determined by preferences only and invariant to the presence of superstar effects, unlike the elasticity of earnings.

This suggests that when superstars are introduced optimal tax rates remain unchanged. Indeed, this is the case in natural benchmark cases. The only caveat is that when performing such a comparative static the allocation will generally change, even if taxes do not, and effort elasticities may depend on where they are evaluated. Likewise, for any given social welfare function, marginal social welfare weights may depend on where they are evaluated. Nevertheless, when elasticities are constant and marginal social welfare weights are unchanged, then marginal tax rates are unchanged. In particular, this holds at the top of the distribution when the welfare weight converges to zero, e.g. with a utilitarian social welfare function.

How does this last neutrality result relate to our two previous results? Result 1 shows that the distribution of earnings affects efficient tax schedules. Typically, the higher inequality that comes with superstar effects will justify higher tax rates. On the other hand, Result 2 shows that superstar effects increase the elasticity of earnings. According to Result 3, these two opposing effects on optimal taxes precisely cancel out.

**Related Literature.** We build on the literature on assignment models that generate superstar effects, surveyed in [Sattinger \(1993\)](#) and more recently [Edmans and Gabaix \(2015\)](#) and [Garicano and Rossi-Hansberg \(2015\)](#). This positive literature has shown how these models can explain observed patterns of occupational choice and income inequality. We add a normative perspective, characterizing optimal redistribution and taxation in a broad class of models used in the literature. Our baseline model is a general one-to-one assignment model between workers and jobs that captures, for example, [Terviö \(2008\)](#) and [Gabaix and Landier \(2008\)](#). Indeed, we inform our elasticity adjustments using the empirical results of [Gabaix and Landier \(2008\)](#), and provide a fully specified parametric example that is able to match their evidence. We also extend the analysis to other settings, such as the one-to-many span-of-control models developed by [Lucas \(1978\)](#) and [Rosen \(1982\)](#), as well as the organizational hierarchy models in [Garicano \(2000\)](#) and [Garicano and Rossi-Hansberg \(2015\)](#).

Our paper incorporates a richer labor market into the canonical Mirrlees model, which features the most basic environment where wage differences simply reflect exogenous skill differences. Recent contributions have enriched the Mirrlees model with worker sorting across jobs and their general equilibrium effects. In particular, [Rothschild and Scheuer \(2013\)](#) and [Ales et al. \(2015b\)](#) show how endogenous sectoral wages and assignments affect optimal income taxes. In these models, relative wages enter incentive constraints. The planner sets taxes with an eye towards influencing these relative wages so as to relax these constraints, generalizing the insight from the two-type model by [Stiglitz \(1982\)](#). The same effect is featured in [Scheuer \(2014\)](#) and [Ales et al. \(2015a\)](#), who focus on occupational choice between workers and managers and the role of endogenous firm size. The latter paper also works with a version of the span-of-control model of [Lucas \(1978\)](#) and [Rosen \(1982\)](#), as we do in one of our extensions, although our adaptation is different in crucial aspects.

While our paper is part of this greater agenda incorporating important features of real-world labor markets into optimal tax analysis, it is quite distinct. Crucially, we focus on superstar effects, which none of the above papers have considered. Indeed, in all these papers each individual faces a linear earnings schedule—whereas a convex schedule is the defining characteristic of superstar phenomena. To isolate the effect of superstars, we purposefully abstract from Stiglitz effects, the central focus of the above literature over the standard Mirrlees model. Moreover, although our model incorporates a richer labor market and features

an equilibrium-determined wage schedule, our results are not driven by general equilibrium effects from tax changes, as is the case with Stiglitz effects. Indeed, Results 1 and 2 call for “partial equilibrium” earnings elasticities with respect to individual tax changes.

We also abstract from externalities and other inefficiencies, such as unproductive rent-seeking or positive spillovers from innovation. For example, some incomes may result from contest-like tournaments with winner-takes-all compensation (Rothschild and Scheuer, 2016). Likewise, CEOs may capture the board, due to poor corporate governance, and set their pay at excessive levels, effectively stealing from stockholders (Piketty et al., 2014). On the other hand, some top earners, such as innovators, may not capture their full marginal product. As this literature shows, such inefficiencies call for Pigouvian adjustments to taxes, but these issues are, once again, conceptually distinct from superstar effects.

Perhaps equally importantly, we also take a fundamentally different approach in deriving our results. The literature has characterized the effects of richer labor markets on taxes by computing optimal taxes and comparing them to standard Mirrleesian ones. This involves fixing primitives, such as preference parameters and the skill distribution, and calibrating technological parameters using data, e.g. sectoral wage and firm size distributions. Instead, we take a “sufficient statistics” perspective, writing taxes directly in terms of observable distributions and elasticities. We relate the relevant elasticities in the presence of superstars to those that emerge in standard settings. We then show how these elasticities are affected by superstar effects. The effects of superstars on taxes work exclusively through these statistics.

This approach is inspired by Saez (2001), who expressed optimal tax formulas in terms of the income distribution and elasticities, and Werning (2007), who developed tests for the Pareto efficiency of a tax schedule as a function of these statistics. This testing approach is particularly useful to derive Results 1 and 2, because it allows us to hold the earnings distribution fixed when characterizing the effects of superstars. In contrast, computing optimal taxes for given redistributive motives requires taking into account that the earnings distribution changes when varying taxes and technology, as we do when deriving Result 3.

In parallel and independent work, Ales and Sleet (2015) pursue a more structural, quantitative approach to optimal taxation in a similar assignment model. Their focus is on how the social weight placed on profits affects the optimal top marginal tax rate. In some cases, they also observe a downward force on this tax rate, even though through a different channel, adjusting the Pareto tail of the inferred skill distribution rather than elasticities. Their goal of quantifying this force is an important complement to our qualitative analytical results.

The paper is organized as follows. Section 2 introduces our baseline model and Section 3 is devoted to our Results 1, 2 and 3. Section 4 discusses extensions, Section 5 provides a quantitative illustration and Section 6 concludes. All proofs are relegated to an appendix.

## 2 Superstars in an Assignment Model

We first lay out our baseline model, based on a one-to-one assignment setting similar to [Terviö \(2008\)](#). Assignment models capture the essence of the superstar phenomena introduced by [Rosen \(1981\)](#).<sup>4</sup> In Section 4, we discuss other settings, including first- and second-generation span-of-control models. All our results go through in these more general settings.

### 2.1 Setup and Equilibrium

Our model blends the canonical Mirrlees model with an assignment model, by introducing an effort decision in the latter.

**Preferences.** There is a continuum of measure one of workers with different ability types  $\theta$ . The utility function for type  $\theta$  is  $U(c, y, \theta)$ , where  $c$  is consumption and  $y$  denotes the effective units of labor supplied to the market, or ‘effort’ for short. The latter is best interpreted as a catch-all for what a worker brings to the table, including what a worker does while on the job, e.g. the hours and intensity of their work effort, as well as what they do before they even reach the job market to make themselves more productive, e.g. their human capital investment.<sup>5</sup> We assume standard conditions on the utility function, including differentiability with  $U_c > 0$ ,  $U_y < 0$  and the single-crossing property, which requires that the marginal rate of substitution

$$MRS(c, y, \theta) \equiv -\frac{U_y(c, y, \theta)}{U_c(c, y, \theta)}$$

is decreasing in  $\theta$ , so that more able agents (higher  $\theta$ ) find it less costly to provide  $y$ . The canonical case studied by [Mirrlees \(1971\)](#) assumes  $U(c, y, \theta) = u(c, 1 - y/\theta)$  for some utility function  $u$  over consumption and leisure. Ability is distributed according to c.d.f.  $F(\theta)$  with density  $f(\theta)$  and support  $\Theta$ .

**Technology.** There is a unit measure of firms indexed by  $x$ , distributed according to c.d.f.  $G(x)$ . A single consumption good is produced through one-to-one matching between individuals and firms. If an individual that provides  $y$  units of effort is matched with firm  $x$ , the output of this match is  $A(x, y)$ , where the function  $A$  is non-negative, increasing, twice differentiable and strictly supermodular:  $A_{xy}(x, y) > 0$ .

---

<sup>4</sup>[Rosen \(1981\)](#) considered a setting where agents choose the scale of their production subject to given costs. This model is close in spirit to the span-of-control models that we consider in Section 4.

<sup>5</sup>As we show in the richer models in Section 4, this can also involve a manager surrounding himself with a team of more and better workers, presumably an important part of his effort in practice. While our notion of effort incorporates human capital, by focusing on income taxation we abstract from the optimal tax treatment of human capital (see e.g. [Bovenberg and Jacobs, 2005](#), and [Stantcheva, 2016](#), for recent work on this issue).

Following [Terviö \(2008\)](#) and [Gabaix and Landier \(2008\)](#), an important benchmark is when  $A$  is linear in  $y$ , so that  $A(x, y) = a(x)y$ . Indeed, when  $A$  can be expressed as a product  $A(x, y) = a(x)\gamma(y)$  one can always renormalize  $y$  so that  $\gamma(y) = y$ .

An allocation specifies  $c(\theta), y(\theta)$  and a measure-preserving assignment rule, mapping worker  $\theta$  to firm  $x = \sigma(\theta)$ . The resource constraint is then

$$\int c(\theta)dF(\theta) \leq \int A(\sigma(\theta), y(\theta))dF(\theta). \quad (1)$$

**Labor Markets.** Individuals are paid according to a schedule  $W(y)$  that is endogenously determined in equilibrium, but is taken as given by both individuals and firms.

Individuals choose effort  $y$  to maximize utility, taking as given the earnings schedule and a nonlinear tax schedule  $T(w)$  set by the government. They solve

$$\max_{c, y} U(c, y, \theta) \quad \text{s.t.} \quad c = W(y) - T(W(y)). \quad (2)$$

with solution  $c(\theta), y(\theta)$ , which by single-crossing are nondecreasing functions of  $\theta$ .

Firm  $x$  maximizes profits, solving

$$\max_y \{A(x, y) - W(y)\}. \quad (3)$$

Denote the solution by  $Y(x)$ , which is monotone by the supermodularity of  $A$ . The necessary first-order condition is

$$A_y(x, y) = W'(y). \quad (4)$$

Because firms are in fixed supply, they will earn profits. We assume equal ownership of firms across individuals. Alternatively, we may assume that firm profits accrue to the government, either because it owns firms directly or because it fully taxes their profits, as in [Diamond and Mirrlees \(1971a\)](#). These standard assumptions allow us to sidestep distributional effects from firm ownership, which are not at the heart of the issue of superstar workers.<sup>6</sup>

**Positive Assortative Matching.** As we have noted, on the worker side  $y$  and  $\theta$  are positively related, due to the single-crossing assumption. On the firm side  $x$  and  $y$  are positively related, due to supermodularity of  $A$ . Together, these two facts imply that there is also a positive relation between  $x$  and  $\theta$ , captured by an increasing assignment function  $\sigma(\theta)$ . We incorporate this fact in the definition of an equilibrium below.

---

<sup>6</sup>Profits are possible because we consider a fixed distribution of firms. With endogenous entry of firms, ex ante or average profits would be dissipated, negating their distributional effects. Such an extension would take us away from the standard assignment models and is beyond the scope of the present paper.



**Equilibrium Definition.** An equilibrium consists of a tax schedule  $T(w)$ , an earnings schedule  $W(y)$ , firms' demands  $Y(x)$ , workers' consumption and effort  $c(\theta), y(\theta)$ , and an increasing assignment function  $\sigma(\theta)$  mapping workers to firms satisfying: (i)  $Y(x)$  solves (3) given  $W(y)$ ; (ii)  $c(\theta), y(\theta)$  solve (2) given  $W(y)$  and  $T(w)$ ; (iii) goods market clearing: the resource constraint (1) holds with equality; and (iv) labor market clearing:

$$F(\theta) = G(\sigma(\theta)), \quad y(\theta) = Y(\sigma(\theta)).$$

Condition (iv) ensures that the match between worker  $\theta$  and firm  $x = \sigma(\theta)$  is consistent with their supply and demand choices,  $y(\theta) = Y(\sigma(\theta))$ , and that  $G(x)$ , the measure of firms demanding  $Y(x)$  or less, coincides with  $F(\theta)$ , the measure of workers supplying  $y(\theta)$  or less.

## 2.2 Superstar Effects

Denoting by  $\Gamma(y)$  the inverse of  $y(\theta)$ , (4) becomes

$$W'(y) = A_y(\sigma(\Gamma(y)), y). \quad (5)$$

For a given effort schedule  $y(\theta)$ , this condition pins down the earnings schedule  $W(y)$  up to a constant.<sup>7</sup> This well-known marginal condition, which obtains in a wide set of assignment models, is fundamental to superstar effects. To see this, differentiate to obtain

$$W''(y) = A_{yx}(\sigma(\Gamma(y)), y) \cdot \sigma'(\Gamma(y)) \cdot \Gamma'(y) + A_{yy}(\sigma(\Gamma(y)), y). \quad (6)$$

By supermodularity, the first term is positive, providing a force for the earnings schedule to be convex in effort. When  $A$  is linear in  $y$ , then  $A(\theta, y) = a(x)y$  for some increasing function  $a$  and so  $W'(y) = a(\sigma(\Gamma(y)))$  is increasing in  $y$ . In this case, output is linear in  $y$  at the firm level, but earnings are convex,  $W''(y) > 0$ , as individuals with higher  $y$  get matched to higher- $x$  firms, which is reflected in the earnings schedule. This is illustrated in Figure 1.

## 2.3 Efficient Taxes and Planning Problem

We now turn to the study of Pareto efficient taxes. We first establish a connection between incentive compatible allocations and allocations that are part of an equilibrium with some taxes  $T$ . This allows us to formulate a planning problem in terms of allocations directly. Our results in later sections will exploit the optimality conditions of this planning problem.

---

<sup>7</sup>In our model, with an equal number of firms and workers, this constant may not be uniquely determined. With more firms than workers, only the top firms are active and earn positive profits. The lowest active firm makes zero profits, pinning down the constant. If  $A(0, 0) = 0$ , this implies  $W(0) = 0$ .

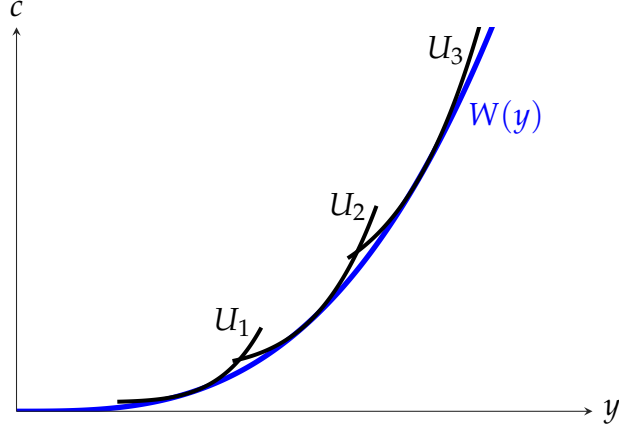


Figure 1: Convex equilibrium earnings schedule.

**Incentive Compatibility and Tax Implementation.** For any allocation  $(c(\theta), y(\theta))$ , define the utility assignment

$$V(\theta) = U(c(\theta), y(\theta), \theta). \quad (7)$$

An allocation is *incentive compatible* when

$$V(\theta) = \max_{\theta'} U(c(\theta'), y(\theta'), \theta) \quad \forall \theta. \quad (8)$$

The following lemma relates equilibria with taxes and incentive compatible allocations.

**Lemma 1.** *An allocation  $(c(\theta), y(\theta), \sigma(\theta))$  is part of an equilibrium for some tax schedule  $T(w)$  if and only if it is resource feasible (1), incentive compatible (7)–(8) and  $\sigma(\theta) = G^{-1}(F(\theta))$ .*

The proof is in Appendix A.1. The key to the ‘only if’ part is that, although the earnings schedule is determined by the labor market and the government cannot tax effort  $y$  directly, it can, for any monotone earnings schedule  $W(y)$ , create incentives using the tax schedule to tailor consumption  $W(y) - T(W(y))$  at will, subject only to incentive compatibility.

By single-crossing, the global incentive constraints (7) and (8) are equivalent to the local constraints

$$V'(\theta) = U_{\theta}(c(\theta), y(\theta), \theta) \quad \forall \theta \quad (9)$$

and the requirement  $y(\theta)$  is non-decreasing. (10)

**Planning Problem.** An allocation is (constrained) *Pareto efficient* if it solves

$$\max_{c, y, V} \int V(\theta) d\Lambda(\theta) \quad (11)$$

subject to (7), (9), (10) and

$$\int (B(\theta, y(\theta)) - c(\theta)) dF(\theta) \geq 0 \quad (12)$$

for some c.d.f. of Pareto weights  $\Lambda(\theta)$ , where we have introduced the shorthand notation  $B(\theta, y) \equiv A(\sigma(\theta), y)$  for the output of type  $\theta$  under the equilibrium assignment rule  $\sigma(\theta) = G^{-1}(F(\theta))$ . The function  $B$  inherits the properties of  $A$ , i.e. monotonicity and supermodularity. We call an earnings tax  $T$  Pareto efficient if it induces a Pareto efficient allocation. We consider cases where constraint (10) is not binding, abstracting from bunching.

Observe that, perhaps surprisingly, the Pareto problem for our superstar economy reduces to a standard Mirrleesian problem, except for the output function  $B$  in the resource constraint, which summarizes production by agent  $\theta$  *in equilibrium*. The Pareto problem simplifies because we are taking into account that the assignment, at the end of the day, must be positively assorted between  $y$  and  $x$ . This reflects the fact that in this baseline model the equilibrium assignment is unaffected by the rest of the allocation  $c(\theta), y(\theta)$ .<sup>8</sup>

However, this does not imply that assignment plays no role. Indeed, *off the equilibrium* a single agent  $\theta$  may deviate and increase  $y$  discretely (surpassing its immediate peers and neighbors in an interval  $[\theta, \theta + \varepsilon)$ ), in which case this agent will match with a higher- $x$  firm, i.e. the one corresponding to the rank in  $y$ , not  $\theta$ . Hence,  $B$  summarizes output only in equilibrium, not off equilibrium. In other words, in equilibrium everyone falls in line, but workers are not predestined and can reach for the stars if they wish. Since agents have the potential to match with different firms, this shapes competition, the earnings schedule, incentives and the response of individual earnings to taxes.

## 3 Results

### 3.1 Result 1: Neutrality

We now investigate the implications of superstar effects for taxes. We use the first-order conditions of the Pareto problem (11) to derive a general efficiency test for earnings taxes  $T(w)$  in our assignment model, extending [Werning \(2007\)](#). Since first-order conditions are necessary, any Pareto efficient allocation needs to pass this test.<sup>9</sup>

---

<sup>8</sup>This is not crucial for our analysis. As we show in Section 4, our results extend to settings where the equilibrium assignment is endogenous to taxes.

<sup>9</sup>Under additional assumptions, e.g. with separable utility  $U(c, y, \theta) = u(c) - \phi(y, \theta)$ , the first-order conditions can be shown to be sufficient and any allocation that passes the test is Pareto efficient.

### 3.1.1 Elasticity Concepts

We will derive conditions for Pareto efficiency and express them in terms of elasticities. A crucial question will be *which elasticities* will be relevant for the different questions we ask.

The first elasticity concept that will be useful are total *earnings* elasticities. In particular, for a given earnings schedule  $W$  and tax  $T$ , define the earnings function by

$$w(1 - \tau, I) \in \arg \max_w U \left( (1 - \tau)w - T(w) + I, W^{-1}(w), \theta \right),$$

where  $W^{-1}$  is the inverse of  $W$ . Here, we increase the marginal tax rate by a small amount  $\tau$  and the intercept by a small amount  $I$  to measure the earnings response. The *uncompensated earnings elasticity* is

$$\varepsilon^u(w) = \left. \frac{\partial w}{\partial(1 - \tau)} \right|_{\tau=I=0} \frac{1 - T'(w)}{w}. \quad (13)$$

*Income effects on earnings* are captured by the parameter

$$\eta(w) = - \left. \frac{\partial w}{\partial I} \right|_{\tau=I=0} (1 - T'(w)), \quad (14)$$

and the *compensated earnings elasticity* is, by the Slutsky equation,

$$\varepsilon^c(w) = \varepsilon^u(w) + \eta(w). \quad (15)$$

Three observations are in order. First, these are elasticities for *earnings*, so they take into account the shape of the equilibrium earnings schedule  $W(y)$ . Their empirical counterparts, in principle, are therefore *elasticities of taxable income* (see [Keane \(2011\)](#), [Saez et al. \(2012\)](#) and [Chetty \(2012\)](#) for recent surveys of the empirical ETI literature).<sup>10</sup> Second, these elasticities are *micro*, partial equilibrium elasticities that hold the earnings schedule  $W(y)$  fixed. At the *macro* level, of course, general equilibrium effects imply that the earnings schedule is endogenous to the tax schedule. Third, these elasticities measure earnings responses under a given (nonlinear) tax schedule  $T$  in place. We will return to a detailed discussion of these elasticities in Section 3.2 and contrast them with other elasticity concepts.

### 3.1.2 An Efficiency Test

We denote the equilibrium *earnings distribution* under a given tax schedule by  $H(w)$  with corresponding density  $h(w)$ . This allows us to state our first result.

---

<sup>10</sup>This is in contrast to pure *effort* elasticities (for instance, labor supply elasticities as measured by hours).

**Proposition 1.** Any Pareto efficient earnings tax  $T(w)$  satisfies

$$\frac{T'(w)}{1 - T'(w)} \varepsilon^c(w) \left( \rho(w) - \frac{d \log \left( \frac{T'(w)}{1 - T'(w)} \varepsilon^c(w) \right)}{d \log w} - \frac{\eta(w)}{\varepsilon^c(w)} \right) \leq 1 \quad (16)$$

at all equilibrium earnings levels  $w$ , where

$$\rho(w) \equiv - \left( 1 + \frac{d \log h(w)}{d \log w} \right)$$

is the local Pareto parameter of the earnings distribution.

Proposition 1, proved in Appendix A.2, reveals three insights. First, for a given tax schedule  $T$ , the distribution of earnings  $h$  as well as the earnings elasticities  $\varepsilon^c$  and  $\varepsilon^u$  (which imply  $\eta = \varepsilon^c - \varepsilon^u$ ) are *sufficient* to evaluate the efficiency test. Conditional on these sufficient statistics, no further knowledge about the structural parameters of the economy and the superstar technology are required. Among these sufficient statistics, the distribution of earnings is directly observable; in principle, the elasticity of earnings may be estimated.

Second, condition (16) is the *same* as in a standard Mirrlees model with linear production (see [Werning, 2007](#)). Of course, as is common with sufficient statistics, the earnings distribution and elasticities are endogenous in general, and we will later show how they are affected by superstar effects. Nonetheless, once we have measures for them from the equilibrium under a given tax schedule, the formula is independent of whether there are superstars or not. In this sense, superstars are *neutral* conditional on the sufficient statistics.

Third, the relevant elasticities turn out to be precisely the *earnings elasticities* defined in the previous subsection, taking the equilibrium earnings schedule  $W(y)$  and the tax  $T$  as given. This is the key channel through which superstar effects enter the test formula.

**Intuition.** Condition (16) provides an *upper bound* for Pareto efficient marginal tax rate at  $w$ . The intuition can be grasped by imagining a reduction in taxes at  $w$ , and checking whether the induced behavioral responses lead to an increase in tax revenue; if they do, there is a local Laffer effect, and the original tax schedule cannot have been Pareto efficient. As in [Werning \(2007\)](#), this comes down to comparing the (positive) revenue effect from those who initially earn less than  $w$  and increase their earnings to benefit from the tax reduction at  $w$ , and the (negative) effect of those initially above  $w$  who reduce their earnings.

If the earnings density falls quickly at  $w$  (the local Pareto parameter is positive and large) the first effect is likely to outweigh the second, and we are more likely to reject Pareto efficiency based on (16). Similar to a high earnings elasticity  $\varepsilon^c$  (in front of the brackets), this reduces the upper bound on the marginal tax at  $w$ . On the other hand, if the tax schedule is

very progressive ( $T'$  increases quickly) or the earnings elasticity increases quickly at  $w$ , then we lose more revenue from the second group than we gain from the first. Hence, a local Laffer effect becomes less likely, allowing for higher taxes as captured by the second term in brackets. The same is true if income effects are large (so  $\eta$  is positive and large).

**The role of earnings elasticities and the earnings distribution.** The usefulness of Proposition 1 is best seen in the simpler case when there are no income effects, so  $\varepsilon^c$  and  $\varepsilon^u$  coincide, the elasticity is locally constant and the tax schedule is locally linear at  $w$ , with marginal tax rate  $\tau$  (such as in the top bracket of the tax schedule).<sup>11</sup> Then (16) reduces to

$$\frac{\tau}{1-\tau}\varepsilon\rho \leq 1 \quad \Leftrightarrow \quad \tau \leq \frac{1}{1+\varepsilon\rho}. \quad (17)$$

The upper bound on  $\tau$  now only depends on the product of the local Pareto parameter  $\rho$  of the earnings distribution and the earnings elasticity  $\varepsilon$ .<sup>12</sup> For a given tax schedule  $T$  in place and the corresponding equilibrium earnings distribution, both  $\tau$  and  $\rho$  are an empirical datum. Hence, whether superstar effects make it more or less likely that the efficiency test (17) is passed, i.e., whether they push for lower or higher taxes, crucially depends on how they affect the earnings elasticity  $\varepsilon$ . We explore this next.

## 3.2 Result 2: Superstars and Earnings Elasticities

We have seen that the test for the efficiency of a tax schedule only depends on the (observable) earnings distribution and the earnings elasticities of superstars under this tax. We now show how these elasticities differ from standard non-superstar elasticities.

### 3.2.1 Estimated versus Required Elasticities

To apply the efficiency test, we require earnings elasticities for superstars, as defined in Section 3.1.1. Suppose, however, that the empirical estimates of earnings elasticities we have access to do not properly capture superstar effects. This could be for the following reasons:

1. The estimates measure *fixed-assignment* elasticities, where individuals are unable to switch matches in response to tax changes, but only vary effort at their current match. This is likely if the estimates are based on temporary and/or unexpected tax changes.

<sup>11</sup>We provide examples with such properties in Section 5.

<sup>12</sup>The upper bound also describes the optimal top tax rate when no social welfare weight is placed on top earners (Diamond and Saez, 2011). This is satisfied, for example, by the commonly used utilitarian welfare criterion with marginal utility of consumption declining to zero. With positive social weight  $\lambda > 0$  attached to top earners, the optimal top rate is  $\tau = (1 - \lambda)/(1 - \lambda + \varepsilon\rho)$ , i.e. strictly less than the upper bound.

2. The elasticities are estimated based on samples of *non-superstar* individuals. For example, the estimates could come from individuals lower in the earnings distribution, where superstar effects may be less prevalent.
3. The estimates are *macro elasticities*, which measure earnings responses to tax changes including the general equilibrium effects on the earnings schedule  $W(y)$ . For instance, this occurs when identification comes from tax reforms with aggregate effects. In contrast, the required elasticities measure earnings responses holding  $W(y)$  fixed.

For each of these scenarios, we next show how the estimated elasticities need to be adjusted.

### 3.2.2 Fixed-Assignment Elasticities

Consider the equilibrium under a given earnings tax  $T$  in place, with equilibrium earnings schedule  $W(y)$  and effort allocation  $y(\theta)$ . Pick an individual in this equilibrium with observed earnings  $w_0$ . We can back out this individual's effort  $y_0 = W^{-1}(w_0)$  and type  $\theta_0 = \Gamma(y_0)$ . We can always write earnings as  $w_0 = B(\theta_0, y_0) - \pi(\theta_0)$ , i.e., as the difference between output and profits at the firm type  $\theta_0$  is matched with.

**Fixed-Assignment Earnings Schedules.** To formalize the idea that the estimated elasticity of earnings at  $w_0$  captures behavioral responses under fixed assignment only, suppose that it is based on the following “non-superstar” earnings schedule

$$\hat{W}_0(y) = B(\theta_0, y) - \pi(\theta_0). \quad (18)$$

$\hat{W}_0$  is the schedule that results from letting the individual with earnings  $w_0$  stay at the firm she is assigned to in the superstar equilibrium. In fact, by (5),  $\hat{W}'_0(y_0) = W'(y_0)$ , so  $\hat{W}_0(y)$  both coincides with and is tangent to the original superstar earnings schedule  $W(y)$  at  $w_0$ , but otherwise follows the shape of the output function  $B(\theta_0, y)$  of the firm that type  $\theta_0$  is matched with in equilibrium. Hence, this is precisely the relationship between effort and earnings that this individual would perceive when confined to staying with the same firm.

In the case where  $A$  is linear in  $y$  (and hence  $B(\theta, y) = b(\theta)y$  for some increasing function  $b$ ), this is particularly simple:  $\hat{W}_0(y)$  is linear, and given by the line tangent to  $W(y)$  at  $w_0$ :

$$\hat{W}_0(y) = w_0 + b(\theta_0)(y - y_0) = w_0 + W'(y_0)(y - y_0).$$

In this sense,  $\hat{W}_0$  precisely captures the absence of superstar effects, as it ignores the convexity of the true earnings schedule  $W$  and instead describes a standard, linear relationship between earnings and effort, as illustrated in Figure 2.

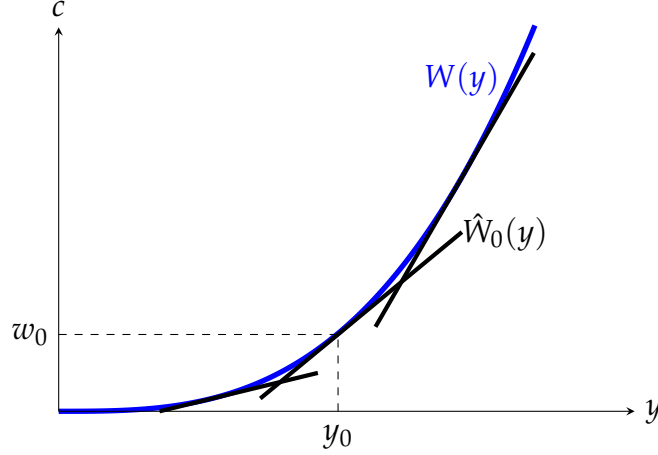


Figure 2: Equilibrium schedule  $W(y)$  versus fixed-assignment schedule  $\hat{W}_0(y)$ .

In terms of empirical counterparts, one way to interpret the fixed-assignment elasticity is as a short-run elasticity of taxable income, which does not observe individuals' behavioral responses to a tax change for long enough to detect the earnings effect of moving to a better job.<sup>13</sup> Instead, it only measures the earnings effect of a change in effort holding job quality fixed.<sup>14</sup> Of course, just like the correct earnings elasticity, it does not include changes due to misreporting for tax avoidance/evasion (Slemrod, 1996) or due to retiming or relabelling of incomes (Goolsbee, 2000) (see Saez et al. (2012) for an overview of these issues).

**Adjusting Fixed-Assignment Elasticities.** Denote the resulting earnings elasticities and income effect by  $\hat{\varepsilon}^u(w_0)$ ,  $\hat{\eta}(w_0)$  and  $\hat{\varepsilon}^c(w_0)$ , defined as in (13)–(15) when replacing  $W$  by  $\hat{W}_0$ . The next result shows how they differ from the elasticities required for the test (16):

**Proposition 2.** *For any given earnings level  $w_0$  in an equilibrium with earnings tax  $T$ , let  $\hat{\varepsilon}^c(w_0)$  and  $\hat{\eta}(w_0)$  be the compensated earnings elasticity and income effect, respectively, based on the earnings schedule  $\hat{W}_0$  defined in (18). Let  $y_0$  be such that  $w_0 = W(y_0) = \hat{W}_0(y_0)$  and  $\theta_0 = \Gamma(y_0)$ . Then*

$$\varepsilon^c(w_0) = \frac{\hat{\varepsilon}^c(w_0)}{\hat{\Phi}(w_0)} > \hat{\varepsilon}^c(w_0) \quad \text{and} \quad \frac{\eta(w_0)}{\varepsilon^c(w_0)} = \frac{\hat{\eta}(w_0)}{\hat{\varepsilon}^c(w_0)},$$

where

$$\hat{\Phi}(w_0) \equiv 1 - \hat{\varepsilon}^c(w_0) \frac{B_{y\theta}(\theta_0, y_0) \Gamma'(y_0) w_0}{W'(y_0)^2} \in [0, 1).$$

<sup>13</sup>In the richer matching models discussed in Section 4, it can also be interpreted more broadly as holding fixed the scale of operations or the team size and team quality of a manager.

<sup>14</sup>In practice, with many jobs of similar quality, this could be the result of switching to a job that is associated with more effort, albeit not necessarily located higher up in the chain of job qualities (see e.g. Chetty et al., 2011, for estimates in a setting where hours are rigid in a given job, but jobs differ in terms of required hours).



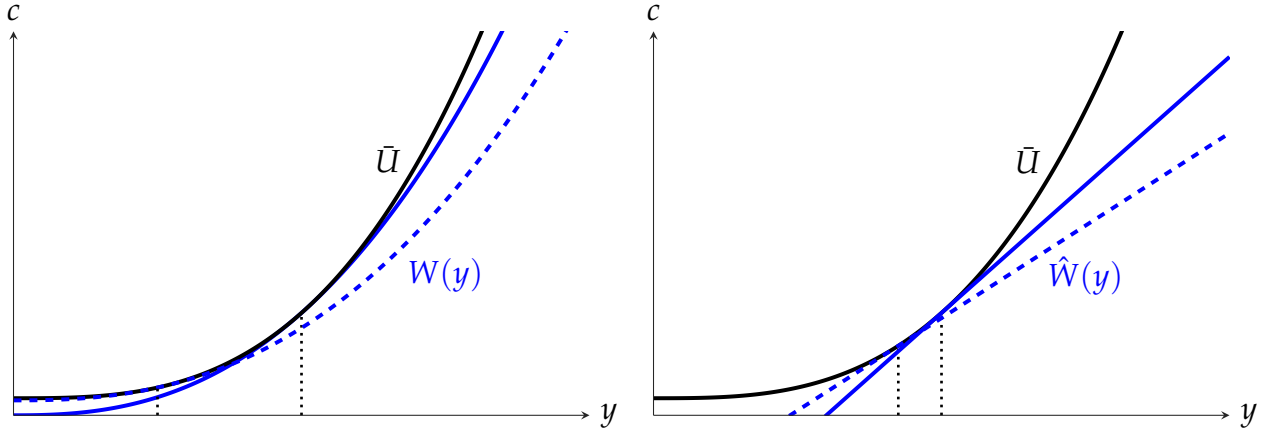


Figure 3: Compensated earnings elasticity with superstar effects (left) and without (right).

Proposition 2, proved in Appendix A.3, shows that we must adjust the compensated fixed-assignment earnings elasticity *upwards* to use it for the efficiency test; no adjustment is required for the income effect ratio  $\eta/\varepsilon^c$ . Observe that Proposition 2 ensures that  $\hat{\Phi} \geq 0$  at an equilibrium, implying an upper bound on the possible range for  $\hat{\varepsilon}^c$ . This bound is implied by the second-order condition for worker optimality.

The intuition for why superstar effects *increase* the earnings elasticity is based on a direct application of the Le Chatelier principle (Samuelson, 1947). The equilibrium earnings schedule  $W(y)$  is the upper envelope of the non-superstar, fixed-assignment earnings schedules  $\hat{W}_0(y)$  across all earnings levels  $w_0$ , as depicted in Figure 2. As a result, it is more convex than each fixed-assignment schedule. This is particularly transparent when  $B$  is linear in  $y$ . Then by (6) the adjustment becomes

$$\hat{\Phi}(w_0) = 1 - \hat{\varepsilon}^c(w_0) \frac{\frac{W''(y_0)y_0}{W'(y_0)}}{\frac{W'(y_0)y_0}{W(y_0)}} \in [0, 1]. \quad (19)$$

The fixed-assignment elasticity ignores the convexity of the earnings schedule. As Figure 3 shows, a convex earnings schedule leads to higher compensated response in effort; implying a higher response in earnings. More generally, the adjustment factor in Proposition 2 captures the additional convexity of the earnings schedule that is due to the supermodularity of the output function  $B_{y\theta} > 0$  (which is precisely due to superstar effects).

In Section 5, we provide a parametric illustration for these adjustments based on the CEO application in Gabaix and Landier (2008), which suggests that these effects can be quantitatively important. Moreover, if a large part of a manager's effort is to affect his scale of operations or his team size and quality, such as in the richer superstar models in Section 4, the difference between the fixed-assignment and the full earnings elasticity is likely even

more important than in the baseline model here.

Taken together with Proposition 1, the upwards adjustment of the earnings elasticity in Proposition 2 provides a force for lower taxes given the earnings distribution. This is particularly clear when both marginal tax rates and elasticities are locally constant, such as in the top bracket of earnings, and in the absence of income effects, so (17) applies. Then using the non-superstar elasticity  $\hat{\varepsilon}$  instead of the superstar elasticity  $\varepsilon > \hat{\varepsilon}$  delivers a *less stringent* upper bound to the top marginal tax rate for a given value of  $\rho$ . Conversely, realizing that the earnings distribution is in fact the result of superstar effects leads to a *more stringent, lower* upper bound to the top marginal tax rate.

### 3.2.3 Effort Elasticities

We now turn to the second scenario, with *effort elasticities* estimated in a setting without superstar effects where earnings and effort coincide. The question is whether we can extrapolate these standard elasticities to superstars.

Suppose individuals work with linear technology and are on a linear part of the tax schedule, as considered by Hausman (1985) and Saez (2001). Then individuals choose effort according to

$$y(1 - \tau, I) \in \arg \max_y U((1 - \tau)y + I, y, \theta)$$

where  $\tau$  is the marginal tax rate and  $I$  is virtual income. The resulting uncompensated effort elasticity, income effect on effort, and compensated effort elasticity are

$$\tilde{\varepsilon}^u(y) = \frac{\partial y}{\partial(1 - \tau)} \frac{1 - \tau}{y}, \quad (20)$$

$$\tilde{\eta}(y) = -(1 - \tau) \frac{\partial y}{\partial I} \quad (21)$$

and

$$\tilde{\varepsilon}^c(y) = \varepsilon^u(y) + \eta(y), \quad (22)$$

respectively. These effort elasticities are only a function of the utility function  $U$ —and in general of the allocation  $c, y$  at which they are evaluated—but do not incorporate the equilibrium earnings schedule  $W$ , in contrast to the earnings elasticities from Section 3.1.1. For instance, under the quasilinear and iso-elastic preference specification  $U(c, y, \theta) = c - (y/\theta)^\gamma$ , the uncompensated and compensated effort elasticities coincide and are given by the constant structural parameter  $\tilde{\varepsilon} = 1/(\gamma - 1)$ . Empirically this corresponds to taxable income elasticities for non-superstar individuals where effort equals earnings (see e.g. Saez, 2010, for estimates based on bunching at kink points in the Earned Income Tax Credit schedule) or

hours elasticities (see e.g. [MaCurdy, 1981](#); [Eissa and Hoynes, 1998](#); [Blundell et al., 1998](#); and [Ziliak and Kniesner, 1999](#)), although of course our notion of effective effort is broader.

The following lemma shows, however, that the effort elasticities are still constrained by the earnings schedule in equilibrium, by the workers' optimality conditions. In particular, as can be seen from [Figure 1](#), in equilibrium the workers' indifference curves must have greater curvature than the earnings schedule, i.e. the necessary second-order conditions must be satisfied. The curvature of the indifference curve equals the reciprocal of the compensated effort elasticity, leading to the following bound.

**Lemma 2.** *In any equilibrium with earnings schedule  $W$ , we must have*

$$\tilde{\varepsilon}^c(y) \leq \frac{1}{\frac{W''(y)y}{W'(y)}}.$$

The proof is in [Appendix A.4](#). This property will be useful in the following, when characterizing the relationship between effort and earnings elasticities. In particular, whenever  $W(y) \rightarrow \infty$  as  $y \rightarrow b$  for some bound  $b < \infty$  (as in the example based on [Gabaix and Landier \(2008\)](#) that we will discuss in [Section 5](#)), then the right-hand side of the above inequality goes to 0, which requires that  $\tilde{\varepsilon}^c$  vanishes at the top. Intuitively, for an unbounded earnings distribution to be consistent with a bounded effort distribution, it must become increasingly costly for agents to increase effort as they approach the top, otherwise all agents would trade off a finite effort cost for an infinite earnings benefit. This unbounded cost from effort must be reflected in the elasticity and is incompatible with elasticities bounded away from zero.

**Adjusting Effort Elasticities.** How do effort elasticities relate to the earnings elasticities of a superstar with given earnings  $w_0$  and effort  $y_0$ ? For the sake of the argument, it will be useful to make the case most favorable for an extrapolation of the effort elasticities. Hence, we assume that (i)  $A$  is linear in  $y$ :  $A(x, y) = a(x)y$  (so the non-superstar counterfactual indeed involves linear production) and (ii)  $T$  is locally linear at  $w_0$  (so individuals are indeed in an income bracket with a constant marginal tax rate, as assumed in the definition of the effort elasticities). This leads to the following result, proved in [Appendix A.5](#):

**Proposition 3.** *Consider a superstar equilibrium and an individual with earnings  $w_0$  and effort  $y_0 = W^{-1}(w_0)$ . Let the compensated effort elasticity  $\tilde{\varepsilon}^c(y_0)$  and income effect on effort  $\tilde{\eta}(y_0)$  be defined as in [\(22\)](#) and [\(21\)](#), respectively. Under conditions (i) and (ii), we have*

$$\varepsilon^c(w_0) = \frac{\tilde{\varepsilon}^c(y_0)}{\tilde{\Phi}(y_0)} > \tilde{\varepsilon}^c(y_0) \tag{23}$$

$$\frac{\eta(w_0)}{\varepsilon^c(w_0)} = \frac{\tilde{\eta}(y_0)}{\tilde{\varepsilon}^c(y_0)} \frac{1}{\frac{W'(y_0)y_0}{W(y_0)}} < \frac{\tilde{\eta}(y_0)}{\tilde{\varepsilon}^c(y_0)}, \quad (24)$$

where

$$\tilde{\Phi}(y_0) = \frac{1}{\frac{W'(y_0)y_0}{W(y_0)}} - \tilde{\varepsilon}^c(y_0) \frac{\frac{W''(y_0)y_0}{W'(y_0)}}{\frac{W'(y_0)y_0}{W(y_0)}} \in [0, 1). \quad (25)$$

Because the superstar earnings schedule  $W(y)$  is convex in effort rather than linear, the Le Chatelier principle again requires an upward adjustment of the compensated elasticity, as in Proposition 2. However, there is now an additional correction, which results from the fact that we must translate the elasticity of  $y$  into an elasticity of  $W(y)$ . This explains the mechanical term  $W'(y)y/w$  (i.e. the elasticity of the earnings schedule) in (24) and (25). Since  $W$  is a convex function that goes through the origin, its elasticity is bigger than one, which implies a *further* upward adjustment of the compensated elasticity, and a *reduction* of the income effect. In view of Proposition 1, both of these new adjustments push for *lower* taxes for a given earnings distribution. We will gauge the quantitative significance of these effects in Section 5.

### 3.2.4 Macro Elasticities

The earnings elasticities required for Proposition 1 measure *partial equilibrium* behavioral responses, holding the equilibrium earnings schedule fixed when varying taxes. The third scenario is one where the estimated elasticities are *macro elasticities*, which capture the general equilibrium effects of tax changes. This occurs when elasticities are estimated using tax reforms with aggregate effects or using cross-country comparisons.<sup>15</sup> These measures include shifts in the equilibrium earnings schedule  $W(y)$  in response to the reform, leading to an additional earnings response even if an individual were to keep her effort  $y$  unchanged.

To fix ideas, consider again a superstar equilibrium for a given tax schedule  $T$ , with earnings schedule  $W(y)$ , and pick some earnings level  $w_0$  with associated effort  $y_0 = W^{-1}(w_0)$  and type  $\theta_0 = \Gamma(y_0)$ . Suppose we increase marginal tax rates by  $\tau$  for everyone. We ask what is the macro elasticity of earnings at  $w_0$ , incorporating all equilibrium responses, and how it compares to the correct, partial equilibrium elasticities from Section 3.1.1.

**Macro Earnings Schedules.** We make two key observations: First, for any value of  $\tau$ , the equilibrium assignment will be the same; i.e., type  $\theta_0$  will stay matched with the same firm. The macro elasticity will therefore be akin to the fixed-assignment elasticity from Section

---

<sup>15</sup>The former approach includes studies based on time series of aggregate income shares in affected tax brackets (e.g. Feenberg and Poterba, 1993; Slemrod, 1996; and Saez, 2004) or studies based on panel data (e.g. Feldstein, 1995; Auten and Carroll, 1999; Gruber and Saez, 2002; and Kopczuk, 2005). For the latter approach, see e.g. Prescott (2004) and Davis and Henrekson (2005).

3.2.2, where earnings follow the equilibrium output function  $B(\theta_0|y)$ . However, the difference is that the macro elasticity also incorporates the *shift* in the earnings schedule, so even when individual  $\theta_0$  sticks to effort  $y_0$ , her earnings will not remain at  $w_0$ . This shift is due to the shift in *profits*  $\pi(\theta_0|\tau)$  of the firm that  $\theta_0$  is matched with in response to the change in  $\tau$ . As a result, the macro earnings response at  $w_0$  is determined by the macro earnings schedule

$$\bar{W}_0(y|\tau) = B(\theta_0, y) - \pi(\theta_0|\tau). \quad (26)$$

To keep the underlying general equilibrium effects tractable, we make two simplifying assumptions for the purpose of this subsection: (i) We again assume  $A$  (and hence  $B$ ) to be linear in  $y$ ; and (ii) we abstract from income effects, so preferences are quasilinear in consumption. Then individual  $\theta_0$  chooses her earnings as follows:

$$\bar{w}_0(\tau) \in \arg \max_w \left\{ (1 - \tau)w - T(w) - \phi \left( \bar{W}_0^{-1}(w|\tau), \theta_0 \right) \right\} \quad (27)$$

for some disutility of effort function  $\phi(y, \theta)$  that is increasing and convex in  $y$  and super-modular in  $(y, \theta)$  (to ensure single-crossing). The compensated and uncompensated macro earnings elasticities then coincide and are defined as

$$\bar{\epsilon}(w_0) = \left. \frac{d\bar{w}_0}{d(1 - \tau)} \right|_{\tau=0} \frac{1 - T'(w_0)}{w_0}. \quad (28)$$

**Adjusting Macro Elasticities.** This leads to the following result (proven in Appendix A.6):

**Proposition 4.** 1. Consider a superstar equilibrium and an individual with earnings  $w_0$  and effort  $y_0 = W^{-1}(w_0)$ . Let the macro earnings elasticity  $\bar{\epsilon}(w_0)$  be as defined in (28), the fixed-assignment elasticity  $\hat{\epsilon}(w_0)$  as in Section 3.2.2, and the effort elasticity  $\bar{\epsilon}(y_0)$  as in (20). Under conditions (i) and (ii), we have

$$\hat{\epsilon}(w_0) = \frac{\bar{\epsilon}(w_0)}{\bar{\Phi}(w_0)} \quad (29)$$

where

$$\bar{\Phi}(w_0) = 1 + \frac{\chi(w_0)}{\bar{\epsilon}(y_0)} \frac{1}{\frac{W'(y_0)y_0}{W(y_0)}} \quad \text{and} \quad \chi(w_0) = \left. \frac{\partial \bar{W}_0(y_0|\tau)}{\partial(1 - \tau)} \right|_{\tau=0} \frac{1 - T'(w_0)}{w_0}. \quad (30)$$

2. If the tax schedule is weakly progressive ( $T''(w) \geq 0 \forall w$ ), then  $\chi(w_0) \leq 0$ , so the macro elasticity is always smaller than the fixed-assignment elasticity.

As the first part of the Proposition shows, the macro elasticity is closely related to the fixed-assignment elasticity, with an adjustment  $\chi$  that captures precisely the shift in the equilibrium earnings schedule due to general equilibrium effects. The second part shows

that this adjustment *lowers* the macro elasticity under a progressive tax schedule. Recall that Proposition 2 showed that the fixed-assignment elasticity is smaller than the correct earnings elasticity. A fortiori, any estimate based on the macro elasticity will also *underestimate* the correct earnings elasticity when there are superstar effects. We will again illustrate this for a simple parametric example in Section 5.

**Intuition.** When we increase  $1 - \tau$ , everyone increases effort in the absence of income effects. Due to complementarities at the firm level, this leads to an increase in profits for each firm, and hence a reduction in earnings holding effort fixed. The assumption of a progressive tax schedule guarantees that this downward shift of the earnings schedule leads to a further increase in effort, and hence profits, because marginal tax rates decrease as we move down the earnings distribution. The general equilibrium effects therefore eventually imply an additional *negative* effect on earnings, reducing the macro earnings elasticity.

### 3.3 Result 3: Comparative Statics

So far, we have been concerned with how of superstars affect the set of Pareto efficient taxes conditional on a given, observed earnings distribution. In this section, we explore the comparative static effects of introducing superstar technology *holding other fundamentals fixed*. These primitives include the parameters of the utility function  $U$  and the skill distribution  $F(\theta)$ . We ask whether, for a given skill distribution and preferences, introducing superstar effects leads to higher or lower taxes.

As in Section 3.1, we begin with a test for the Pareto efficiency of a tax schedule, now however expressed in terms of fundamentals. We denote by  $\tau(\theta) \equiv T'(W(y(\theta)))$  the marginal tax rate faced by type  $\theta$  and focus on the canonical Mirrleesian specification  $U(c, y, \theta) = u(c, y/\theta)$  (see Appendix A.7 for the general case).

**Proposition 5.** *Any Pareto efficient earnings tax  $T$  is such that*

$$\frac{\tau(\theta)}{1 - \tau(\theta)} \frac{\tilde{\varepsilon}^c(\theta)}{1 + \tilde{\varepsilon}^u(\theta)} \left[ p(\theta) - \frac{d \log \left( \frac{\tau(\theta)}{1 - \tau(\theta)} \frac{\tilde{\varepsilon}^c(\theta)}{1 + \tilde{\varepsilon}^u(\theta)} \right)}{d \log \theta} - \frac{\tilde{\eta}(\theta)}{\tilde{\varepsilon}^c(\theta)} \frac{y'(\theta)\theta}{y(\theta)} \right] \leq 1. \quad (31)$$

for all  $\theta$ , where

$$p(\theta) = - \left( 1 + \frac{d \log f(\theta)}{d \log \theta} \right)$$

is the local Pareto parameter of the skill distribution.

Proposition 5 reveals three insights. First, the *formula* for the efficiency test (31) does not feature any parameters related to the superstar technology. This is similar to Proposition 1,

and so is the intuition for the terms in the test condition. Second, however, it shows that *effort* elasticities need to be used when taking the underlying *skill* distribution as an input. This is in contrast to our results in Section 2.3, where *earnings* elasticities were the key sufficient statistic when using the earnings distribution as an input. Third, the test inequality (31) is in fact *identical* to the one that would obtain in a standard Mirrlees model (see [Werning, 2007](#)). Of course, the effort elasticities and effort schedule  $y$  are endogenous in general, so superstar effects can affect the equilibrium at which they are evaluated, but we will next present a natural benchmark case where not even this matters.

**An Example.** Suppose preferences are quasilinear and iso-elastic, so that  $\tilde{\eta} = 0$  and  $\tilde{\varepsilon}^c = \tilde{\varepsilon}^u = \tilde{\varepsilon}$  is constant, and the tax schedule is locally linear with marginal tax rate  $\tau$  (as is the case in the top bracket, for example). Then (31) simplifies to

$$\frac{\tau}{1-\tau} \frac{1}{1+1/\tilde{\varepsilon}} p(\theta) \leq 1. \quad (32)$$

The condition puts an upper bound on the marginal tax rate that is decreasing in the *effort* elasticity  $\tilde{\varepsilon}$  and the local Pareto parameter  $p(\theta)$  of the skill distribution. The bound is independent of the equilibrium earnings schedule  $W(y)$  and the allocation  $c(\theta), y(\theta)$ . In other words, superstar effects are completely *neutral*: The shape of the output function  $A$  does *not* affect the set of Pareto efficient top marginal tax rates given the other fundamentals.

The intuition, which we formalize in Section 5, can be grasped by noting the equivalence of condition (32) with the efficiency condition (17) expressed in terms of the Pareto parameter for *earnings*  $\rho$  and the *earnings* elasticity  $\varepsilon$ . Superstar effects change both  $\rho$  and  $\varepsilon$ . On the one hand, they lead to a fatter tail in the earnings distribution and thus a *lower* Pareto parameter  $\rho$  at the top; on the other hand, they *increase* the earnings elasticity  $\varepsilon$ . Neutrality obtains because both changes precisely cancel out for given primitive parameters  $\tilde{\varepsilon}$  and  $p$ . Hence, despite the potentially extreme effects of superstar technology on earnings inequality, the efficiency test does not change for a given skill distribution and utility function.

**Optimal Tax Schedules in Terms of Primitives** Rather than testing the efficiency of a given tax schedule, it is also possible to characterize *optimal* taxes in terms fundamentals, for any given Pareto weights. This allows us to characterize the comparative static effects of superstars on optimal tax schedules. For example, we can ask whether, for given preferences and skills, the increased role of superstar effects should lead to an increase or decrease in the top marginal tax rate, for given redistributive preferences. The following result provides a formula for the case  $U(c, y, \theta) = u(c, y/\theta)$  (see Appendix A.8 for the general case):

**Proposition 6.** *The optimal tax schedule for any distribution of Pareto weights  $\Lambda$  satisfies*

$$\frac{\tau(\theta)}{1 - \tau(\theta)} = \frac{1 + \tilde{\varepsilon}^u(\theta)}{\theta f(\theta) \tilde{\varepsilon}^c(\theta)} \int_{\theta}^{\infty} \left( 1 - \frac{\lambda(s) U_c(s)}{\eta f(s)} \right) \exp \left( \int_{\theta}^s \frac{\tilde{\eta}(t)}{\tilde{\varepsilon}^c(t)} \frac{dy(t)}{y(t)} \right) dF(s), \quad (33)$$

where  $\eta$  is the multiplier on (12) and  $\lambda$  is the density corresponding to  $\Lambda$ .

Condition (33) is exactly the same as in a standard Mirrlees model. Thus, our neutrality result extends to *optimal* tax schedules given fundamentals, in the sense that superstars do not change the *formula* for optimal taxes.<sup>16</sup> In a linear tax framework, [Diamond and Mirrlees \(1971b\)](#) derived optimality conditions for tax rates, providing a system of equations expressed in terms of taxes and own- and cross-price elasticities for a finite set of goods. Notably, they observed that these conditions do not depend on the curvature of technology (second derivatives of the production function). However, since our non-linear tax framework is different, our Result 3 is not a direct implication of their result.<sup>17</sup>

## 4 Richer Models of Superstars

In this section, we demonstrate that our results are not specific to the simple one-to-one matching model considered so far. They extend to more general settings with superstar effects, including the first- and second-generation span-of-control models we turn to next.

### 4.1 Span-of-Control with Identical Workers

**Managers and Workers.** We begin with the canonical span-of-control model going back to [Lucas \(1978\)](#) and [Rosen \(1982\)](#), adding an intensive effort margin for managers. There is a unit mass of managers with skill  $\theta \sim F$  who provide effective effort  $y$  as before. Rather than being matched with an exogenous set of firms, managers own the firms and hire a homogeneous labor input to produce output. There is a unit mass of identical workers who each supply one unit of labor inelastically at zero cost. A manager who exerts  $y$  units of effective effort and who hires  $L$  units of labor produces output according to  $A(L, y)$ , which we again assume to be supermodular, so  $A_{Ly} > 0$ .

---

<sup>16</sup>In particular, for the case without income effects, the revenue-maximizing top marginal tax rate (when the social welfare weight put on top earners is zero) is  $\tau = \left( 1 + \frac{1}{1+\tilde{\varepsilon}} p \right)^{-1}$ , independent of superstar effects. See also [Ales and Sleet \(2015\)](#) for the case where the planner puts the same weight on profits as on tax revenue. One might view this result as consistent with the fact that top tax rates have remained relatively constant over the past 30 years compared to the considerable increase in income inequality.

<sup>17</sup>See [Scheuer and Werning \(2015\)](#) for a connection between the [Mirrlees \(1971\)](#) and the [Diamond and Mirrlees \(1971b\)](#) optimal tax formulas.



Denoting the wage per unit of labor by  $\omega$ , a manager's earnings (profits) are  $A(L, y) - \omega L$ . The government imposes a nonlinear earnings tax  $T$  on these profits as well as a (lump-sum) tax on workers' incomes  $\omega$ . Managers hire labor to maximize profits taking the wage as given, so their earnings are

$$W(y|\omega) = \max_L A(L, y) - \omega L \quad (34)$$

with optimal labor demand  $L(y|\omega)$  conditional on any given wage  $\omega$ .

**Superstar Effects.** Superstar effects arise here because more talented managers, who exert higher effort  $y$ , hire more workers, thereby increasing their scale.<sup>18</sup> Indeed, by the envelope theorem,

$$W'(y|\omega) = A_y(L(y|\omega), y) > 0$$

and

$$W''(y|\omega) = A_{yL}(L(y|\omega), y)L'(y|\omega) + A_{yy}(L(y|\omega), y).$$

The first term is positive by supermodularity of  $A$  and because  $L'(y|\omega) > 0$  by Topkis's theorem, providing a force for the managers' earnings schedule to again be convex.

Ales et al. (2015) consider a similar model, but with the difference that output is given by  $\theta^\gamma A(L, y)$  and focusing on the case where  $A$  exhibits constant returns to scale. Then, because of the latter property, earnings are linear in effort for any given manager and, in this sense, there are no superstar effects. On the other hand, the direct "scale of operations" effect of manager skill on output, governed by the parameter  $\gamma$ , introduces a direct role for talent, separate from effective effort, which we have abstracted from here.<sup>19</sup>

**Pareto Problem.** For the purpose of this section, it is convenient to index managers by the quantiles of the skill distribution  $t = F(\theta)$ , which is without loss of generality. Manager  $t$ 's preferences are simply  $U(c, y, F^{-1}(t))$ . Let  $c^w$  denote the consumption of workers and  $\Psi(\{y\})$  the equilibrium aggregate output in the economy for a given effort schedule  $y(t)$  for managers. We characterize this functional in Appendix B.1 and prove the following result:

**Lemma 3.** *For any given  $c^w$ , the Pareto problem for managers in this economy is the same as our original Pareto problem (11) subject to (8) and (12) from Section 2.3 when we replace  $\int B(t, y(t))dt$  by  $\Psi(\{y\})$ . Moreover,  $\Psi(\{y\})$  satisfies, for any  $\omega$  and  $t$ ,*

$$\frac{\partial \Psi(\{y\})}{\partial y(t)} = W'(y(t)|\omega).$$

<sup>18</sup>Similarly, in Rosen (1981), individuals can directly increase the scale of their production.

<sup>19</sup>The constant returns to scale assumption is not required for all of their results and they also explore taxes that may distort firm sizes (such as nonlinear taxes on labor input), which we rule out here.

Since all our results are based on the necessary first-order conditions of the Pareto problem and the partial derivative of aggregate output with respect to effort coincides with the marginal earnings in equilibrium at any effort level by Lemma 3, Propositions 1 to 6 immediately extend to the earnings tax for managers in this model.

## 4.2 Span-of-Control with Heterogeneous Workers

We next demonstrate how our results extend to modern, richer models of organizational hierarchies, including those developed by Garicano and Rossi-Hansberg (2004, 2006) and, most recently, Fuchs et al. (2015).<sup>20</sup> Higher-skilled managers not only leverage their talent by hiring more workers, they are also matched with teams of higher-skilled workers, generating yet an additional source of superstar effects. Moreover, these models allow for endogenous sorting of individuals into hierarchical layers.

**Knowledge-based Hierarchies.** As in Garicano and Rossi-Hansberg (2004) and Fuchs et al. (2015), we consider the simplest possible setting in which hierarchies consist of two layers, managers and workers (or producers).<sup>21</sup> In fact, we closely follow the setup in Fuchs et al. (2015) where individuals use time and knowledge to solve problems, with the only addition that we incorporate an intensive effort margin to generate a role for taxes.

There is free entry of risk-neutral firms. Each firm hires one manager and  $n$  workers, where  $n$  may be non-integer. There is also a continuum of heterogeneous individuals of unit mass indexed by their talent  $t$  who are each endowed with one unit of time and can become either managers or workers. As in the preceding subsection, we associate  $t$  with the quantiles of the skill distribution  $F$ , so  $t \sim U(0, 1)$ . Individuals provide effective effort  $y$  and have preferences  $U(c, y, F^{-1}(t))$ . For the purposes of this subsection, w.l.o.g. we also normalize effective effort such that  $y \in [0, 1]$ .

**Workers and Managers.** Each worker is faced with a problem of random difficulty  $x \sim U(0, 1)$ , distributed independently across problems. She manages to solve the problem on her own if her effective effort exceeds the problem's difficulty level. Formally, if her effective effort is  $y_p$ , she can solve the problem whenever  $x < y_p$ , thus with probability  $y_p$ . If she fails, she can seek help from a manager. It costs a manager  $h$  units of her time to help a worker with an unsolved problem. Hence, if a manager is matched with workers who provide effort

<sup>20</sup>See also Garicano and Rossi-Hansberg (2015) for a recent survey of this literature.

<sup>21</sup>Extending our results to more layers, such as in Garicano and Rossi-Hansberg (2006), is straightforward.

$y_p$ , her time constraint implies that she can help  $n$  of them, where  $n$  is such that

$$1 = hn(1 - y_p) \Rightarrow n = \frac{1}{h(1 - y_p)}. \quad (35)$$

Due to this time constraint, a manager can increase his team size only by matching with more talented workers, who put more effort and solve more problems on their own.<sup>22</sup>

A manager with effort  $y_m$  can solve a problem for which a worker asks for help if  $x < y_m$ . A solved problem generates output 1 and an unsolved problem produces 0. Hence, if a manager provides effective effort  $y_m$  and helps one worker of effort  $y_p$  with an unsolved problem, expected output is

$$\frac{y_m - y_p}{1 - y_p}.$$

In turn, if a manager of effort  $y_m$  is matched with  $n$  workers of effort  $y_p$ , expected output is

$$A(y_m, y_p) = ny_p + n(1 - y_p)\frac{y_m - y_p}{1 - y_p} = ny_m = \frac{y_m}{h(1 - y_p)},$$

where the last step uses the time constraint (35). Observe that  $A$  is strictly supermodular and weakly convex in both arguments.

**Equilibrium.** As before, equilibrium will involve an earnings schedule  $W(y)$  that is taken as given by both firms and individuals. For concreteness, denote by  $W_m(y_m)$  the earnings of managers who exert effort  $y_m$  and by  $w_p(y_p)$  those of workers (producers) with effort  $y_p$ . Expected profits for a firm with a manager of effort  $y_m$  and workers of effort  $y_p$  are

$$\Pi = n(A(y_m, y_p) - w_p(y_p)) - W_m(y_m) = \frac{y_m - w_p(y_p)}{h(1 - y_p)} - W_m(y_m) \quad (36)$$

and firms choose  $y_m$  and  $y_p$  to maximize  $\Pi$  given the earnings schedule and thus  $w_p(y_p)$  and  $W_m(y_m)$ . The government imposes a nonlinear earnings tax  $T(w)$  paid by all individuals.

In this model, it is straightforward to see that the equilibrium for any given  $T$  and hence effort schedule  $y$  that is monotone increasing is exactly as described in Lemma 1 and Corollary 1 in [Fuchs et al. \(2015\)](#): there is a partition of the talent space into three intervals  $[0, t_1]$ ,  $(t_1, t_2)$ , and  $[t_2, 1]$ , such that (i) individuals in  $[0, t_1]$  become workers matched with managers, (ii) individuals in  $(t_1, t_2)$  are unmatched and work on problems on their own as self-employed producers, (iii) individuals in  $[t_2, 1]$  become managers and are matched with

---

<sup>22</sup>By single-crossing preferences, the effort schedule  $y$  is again increasing in talent  $t$  in equilibrium. Moreover, note that we are restricting attention to assignments where all workers who are matched with a given manager have the same talent. This is without loss as there will be positive assortative matching in equilibrium.

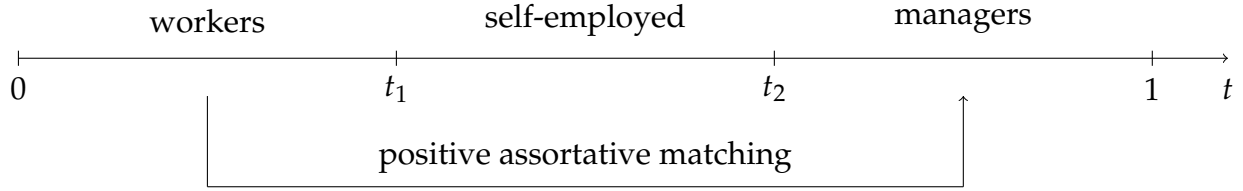


Figure 4: Equilibrium sorting and matching.

workers, and (iv) there is strictly positive assortative matching between workers in  $[0, t_1]$  and managers in  $[t_2, 1]$ . Figure 4 illustrates this pattern.

In other words, the lowest skill types become workers, trying to solve the easiest problems. The highest skill types become managers, providing help with the unsolved problems left by workers. Those with intermediate skill may remain unmatched, working as self-employed on problems without help from managers.

We show in Appendix B.1 that, in contrast to the models we have considered so far, the effort schedule  $y$ , and hence the tax schedule  $T$ , now affects the equilibrium assignment of workers to managers. When workers matched with a manager put more effort, they leave fewer problems unsolved and the manager runs out of time less quickly. This manager can therefore help more workers, and the rate at which we move to lower skilled workers assigned to marginally worse managers increases.

**Superstar Effects.** Superstar effects are different in this framework compared to the previous subsection: Here, better managers get matched with *more* workers, even though all managers face the same time constraint, precisely because they get to supervise *better* workers, who put more effort, and who each require less time to get help. In addition, there are now also superstar effects for workers, as better workers get matched with better managers, who can help them with more problems, boosting their productivity. Formally, we show in Appendix B.1 that both  $W_m(y_m)$  and  $w_p(y_p)$  are convex in equilibrium.

**Pareto Problem.** Let  $\Psi(\{y\})$  again denote the equilibrium aggregate output in the economy under the equilibrium assignment for any given (monotone) effort schedule, which we define formally in Appendix B.1. Then we can prove the following result:

**Lemma 4.** *The Pareto problem for this economy is the same as our original Pareto problem (11) subject to (8) and (12) from Section 2.3 when we replace  $\int B(t, y(t))dt$  by  $\Psi(\{y\})$ . Moreover,  $\Psi(\{y\})$  satisfies, for all  $t$ ,*

$$\frac{\partial \Psi(\{y\})}{\partial y(t)} = W'(y(t)).$$

By the same arguments as in the previous subsection, and despite the richer effects of taxes on equilibrium matching in this model, Lemma 4 implies that our results again extend to the earnings tax for both workers and managers in this model.

### 4.3 A General Production Function Approach

Both of our previous extensions concluded that, in equilibrium, there exists an aggregate production function  $\Psi(\{y\})$  that maximizes total output, given effort choices  $\{y\}$ , and the property that marginal earnings  $W'(y(t))$  equal the derivative of  $\Psi$  with respect to  $y(t)$ . This suggests starting at the end of this line of reasoning, in a more general and abstract manner.

To pursue this, we simply postulate that there exists an aggregate production function  $\Phi(M)$ , where  $M$  denotes a positive measure over the set of possible effort levels  $Y$ . Typically,  $\Phi(M)$  will be defined by maximizing output over all feasible allocations, given the distribution of effort  $M$ . We also assume that the equilibrium measure  $M$  maximizes aggregate profits

$$\Phi(M) = \int W(y) dM.$$

This may be taken as a defining property of competitive equilibria.

Starting with a distribution of types  $t \in [0, 1]$ , an allocation  $y : [0, 1] \rightarrow Y$  induces a particular measure  $M_y$ . Define  $\Psi(\{y\}) \equiv \Phi(M_y)$ . It then follows that in equilibrium  $\{y\}$  maximizes

$$\Psi(\{y\}) = \int W(y(t)) dt.$$

A necessary condition is that

$$\frac{\partial \Psi(\{y\})}{\partial y(t)} = W'(y(t)), \tag{37}$$

the desired marginal condition. Our results then follow for this more general formulation.<sup>23</sup>

It is worth pointing out that this approach allows for arbitrary effects of taxes on the equilibrium assignment. This is because taxes can distort the effort schedule  $y(t)$ , which in turn affects matching. However, the marginal condition (37) implies a form of efficiency conditional on a given effort schedule, as (private) marginal earnings coincide with the (social) marginal product. In other words, taxes distort the assignment through their effect on effort, but not conditional on effort. Distortions conditional on effort could arise, for instance, when part of the matching surplus is untaxed, as in [Jaffe and Kominers \(2014\)](#). These issues are not inherently related to superstar effects. They would emerge, more broadly, in any Roy model with untaxed activities, such as household production or an informal sector.

---

<sup>23</sup>As this extension makes clear, our results are not limited to cases where superstar workers are managers as in the previous two subsections.

## 5 Quantitative Illustration

In this section, we illustrate the quantitative importance of superstar effects for taxes and the elasticity adjustments we provide. We do so by adapting the CEO application in [Gabaix and Landier \(2008\)](#). Their framework is based on the one-to-one matching model of Section 2, but takes the distribution of effort  $y$  as given (“talent” in their terminology). We begin with providing two examples with endogenous effort that deliver this distribution in equilibrium.

### 5.1 Two Examples Based on Gabaix-Landier

Individual CEOs produce  $CS^\beta y$  when matched with a firm of size  $S$  and supplying effort  $y$ , for some parameter  $\beta > 0$ . Firm size is a function of the rank,  $1 - x$ , according to Zipf’s law:  $S = \tilde{C}(1 - x)^{-1}$ ; the distribution of  $x$  is uniform on  $[0, 1]$ . It follows that  $A(x, y) = (1 - x)^{-\beta} y$  for some normalization of  $C$ . Gabaix-Landier find that the CEO earnings distribution features a Pareto tail: Letting  $n$  denote the rank of a CEO in the earnings distribution (i.e.,  $n$  equals 1 minus the c.d.f.) and letting  $\rho$  denote the Pareto parameter, we have near the top

$$w(n) = \kappa n^{-1/\rho}. \quad (38)$$

We can restate problem (3) for firm  $x$  as maximizing  $(1 - x)^{-\beta} y(n) - w(n)$ , implying

$$y'(n) = n^\beta w'(n) = -\frac{\kappa}{\rho} n^{\beta - \frac{1}{\rho} - 1},$$

where we used that in equilibrium  $n = 1 - x$ . For some constant of integration  $b$ ,

$$y(n) = b - \frac{\kappa}{\beta\rho - 1} n^{\beta - \frac{1}{\rho}}. \quad (39)$$

When  $\beta\rho > 1$ , effort is bounded above by  $b$ . Otherwise, effort is unbounded above. Combining equations (38) with (39) gives the increasing and convex earnings schedule

$$W(y) = \begin{cases} \tilde{\kappa} (b - y)^{\frac{-1}{\beta\rho - 1}} & \text{if } \beta\rho > 1 \\ \tilde{\kappa} (y - b)^{\frac{1}{1 - \beta\rho}} & \text{if } \beta\rho < 1 \end{cases} \quad (40)$$

for some constant  $\tilde{\kappa} > 0$ , in the range of equilibrium values  $y \in [y(1), y(0)]$ .

For some purposes, the Gabaix-Landier specification (38)–(40) is all we need. Indeed, given our sufficient statistic approach, many of results do not require all primitives. However, to describe a full economy, it remains to specify preferences and a skill distribution consistent with (38)–(40). Our next two examples do precisely this, by picking preferences

and then backing out the distribution of skills.

On the one hand, Gabaix-Landier provide empirical support for  $\beta = 1$  and  $\rho = 3$ , implying  $\beta\rho > 1$  and a bounded effort distribution. On the other hand, the standard Mirrlees case has  $\beta = 0$ , implying  $\beta\rho < 1$  in the neighborhood of this benchmark model. We therefore span both cases: our first example is tailored to  $\beta\rho < 1$  and the second to  $\beta\rho > 1$ .

**Example Economy A ( $\beta\rho < 1$ ).** Our first example adopts a standard parametrization of the Mirrlees model. Utility is quasilinear and iso-elastic,

$$U(c, y, \theta) = c - \frac{1}{\gamma} \left( \frac{y}{\theta} \right)^\gamma,$$

with  $\gamma > 1$ . The compensated and uncompensated elasticity of effort is  $1/(\gamma - 1)$ .

Consider the equilibrium under a tax schedule with constant marginal tax rate  $\tau$  (or the top bracket of a nonlinear schedule). Appendix B.2 shows that (38)–(40) hold for a skill distribution with a Pareto tail, so that  $F(\theta) \approx 1 - (\underline{\theta}/\theta)^\alpha$  for high  $\theta$ , where  $\alpha > 0$  is the Pareto parameter. In addition,

$$\rho = \frac{\gamma - 1}{\gamma} \frac{\alpha}{\alpha\beta + 1}, \quad (41)$$

so that the Pareto parameter  $\rho$  is *decreasing* in  $\beta$  for fixed  $\gamma$  and  $\alpha$ , which illustrates that superstar effects induce a *fatter* tail and hence a more unequal earnings distribution in this sense. Note that equation (41) implies  $\beta\rho < 1$ .<sup>24</sup>

**Example Economy B ( $\beta\rho > 1$ ).** This second example is most suitable for the benchmark parameterization in Gabaix-Landier. Utility is given by

$$U(c, y, \theta) = c - \frac{1}{\gamma} \left( \frac{a - \theta}{b - y} \right)^\gamma,$$

with  $a, b, \gamma > 0$ . This utility satisfies standard assumptions: it is decreasing and concave in  $y \leq b$  and satisfies the single-crossing property. We will see that it also has other, necessary features within the case  $\beta\rho > 1$ . Notably, disutility of effort goes to infinity as  $y \rightarrow b$ , which is crucial (in view of Lemma 2) to rationalize bounded effort with unbounded earnings.

Consider the equilibrium under a constant marginal tax rate  $\tau$ . Appendix B.2 shows that (38)–(40) hold for skill distribution  $F(\theta) = 1 - (a - \theta)^\alpha$  for  $\theta \leq a$ , with  $\alpha\beta > 1$ . In addition,

$$\rho = \frac{\gamma + 1}{\gamma} \frac{\alpha}{\alpha\beta - 1}, \quad (42)$$

---

<sup>24</sup>An alternative to backing out the skill distribution that implies (38)–(40) exactly is to assume a Pareto distribution throughout; then, as shown in the appendix, (38)–(40) hold approximately at the top.

which is decreasing in  $\beta$ , so that earnings inequality is again increasing in  $\beta$ . Note that (42) implies  $\beta\rho > 1$ . In particular, we can always find  $\alpha$  and  $\gamma$  so that  $\rho = 3$  and  $\beta = 1$ , consistent with Gabaix-Landier's favored parameterization.

## 5.2 Illustrating Results 1 and 2

We now use the specifications of the earnings and effort distribution (38) and (39) and earnings schedule (40) implied by Gabaix-Landier, as well as our examples that generate them from primitives, to inform the elasticity adjustments in Section 3.2 and their tax implications.

**Fixed-Assignment Elasticities.** By (40), the elasticities of the earnings schedule and the marginal earnings schedule are

$$\frac{W'(y)y}{W(y)} = \frac{1}{\beta\rho - 1} \frac{y}{b - y} \quad \text{and} \quad \frac{W''(y)y}{W'(y)} = \frac{\beta\rho}{\beta\rho - 1} \frac{y}{b - y}.$$

Substituting this in the adjustment factor (19) for the fixed-assignment elasticity gives

$$\varepsilon^c(w) = \frac{\hat{\varepsilon}^c(w)}{1 - \beta\rho\hat{\varepsilon}^c(w)}. \quad (43)$$

By Proposition 2, in equilibrium  $\hat{\varepsilon}^c(w)$  is restricted so that the denominator is positive.

Recall that Gabaix-Landier's preferred parameterization has  $\beta = 1$  and  $\rho = 3$ ; this requires  $\hat{\varepsilon}^c \leq \frac{1}{\beta\rho} = \frac{1}{3}$  to be consistent with an equilibrium. Equation (43) implies potentially considerable upward adjustments in fixed-assignment elasticities. For instance, if  $\hat{\varepsilon}^c = 1/4$ , then the correct earnings elasticity is in fact  $\varepsilon^c = 1$ . The blue line in the left panel of Figure 5 displays  $\varepsilon^c$  as a function of  $\hat{\varepsilon}^c \in [0, 1/3)$ . As a robustness check, the red line depicts the same relationship for the case  $\beta = 1/2$ , i.e. half the size of the superstar effects found by Gabaix-Landier. In this case, equilibrium requires  $\hat{\varepsilon}^c \leq 2/3$  and the adjustments remain significant. For example, if  $\hat{\varepsilon}^c = 1/3$ , the correct earnings elasticity is  $\varepsilon^c = 2/3$ , twice as high.

For the case of no income effects (as in our examples), the right panel in Figure 5 plots the upper bound to the top marginal tax rate implied by the efficiency condition (17) as a function of the elasticity  $\hat{\varepsilon}$ . The black line is based on (erroneously) using the unadjusted elasticity  $\hat{\varepsilon}$  instead of  $\varepsilon$  in (17), whereas the blue line uses the (correct) adjustment (43) for  $\beta = 1$  and the red line for  $\beta = 1/2$  (and  $\rho = 3$  in both cases). For example, if  $\hat{\varepsilon} = 1/4$ , then  $\tau \leq 57\%$  based on the unadjusted elasticity, but  $\tau \leq 45\%$  based on the correct adjustment for  $\beta = 1/2$  and  $\tau \leq 25\%$  for  $\beta = 1$ .<sup>25</sup>

<sup>25</sup>Diamond and Saez (2011) argue that the Pareto coefficient for the general population converges to  $\rho = 1.5$  at the top, rather than  $\rho = 3$  found by Gabaix and Landier (2008) for CEOs (the latter value is also found by Ales



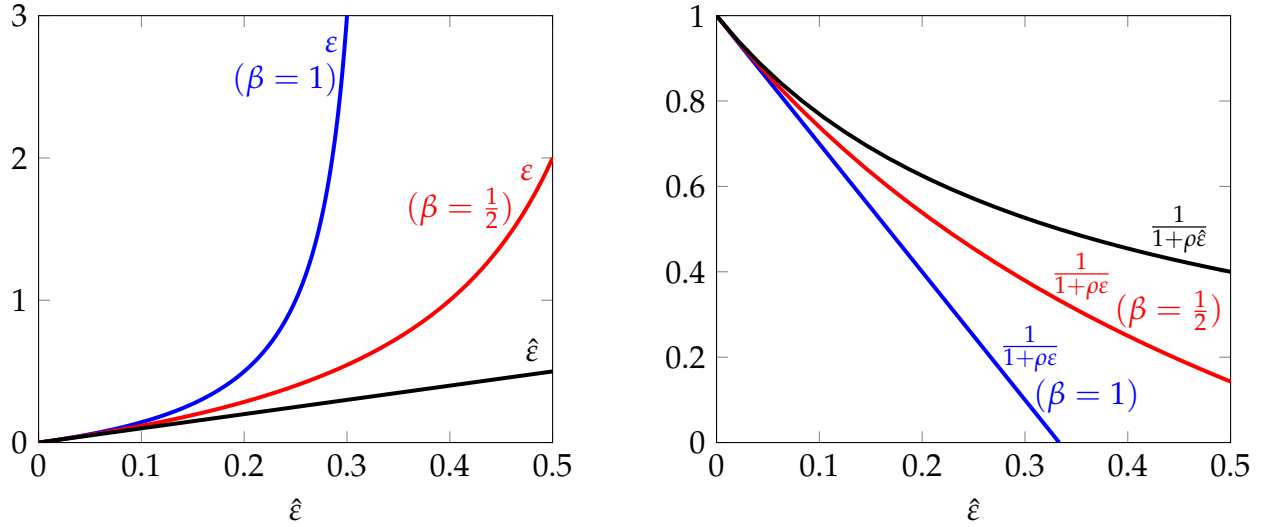


Figure 5: Unadjusted and adjusted earnings elasticities (left panel) and upper bound on top tax rate using unadjusted and adjusted earnings elasticities (right panel).

**Effort Elasticities.** Using (39) and (40), the required adjustment (23) and (25) for the effort elasticity at any point in the distribution simplifies to

$$\varepsilon^c = \frac{\tilde{\varepsilon}^c}{\frac{\beta\rho-1}{\beta\rho n^{\frac{1}{\rho}-\beta}-1} - \beta\rho\tilde{\varepsilon}^c}. \quad (44)$$

The adjustment now depends on the rank  $n$  (and we omitted the dependence of  $\varepsilon^c$  and  $\tilde{\varepsilon}^c$  on  $n$  to simplify notation). Note that for the bottom ( $n = 1$ ) the effort elasticity adjustment is the same as the fixed-assignment elasticity adjustment (43); this reflects the fact that  $\frac{W'(y)y}{W(y)} = 1$  for the lowest worker. The adjustment is larger, for a given  $\tilde{\varepsilon}^c$ , for lower  $n$  as we move up the distribution towards the top; this is because  $\frac{W'(y)y}{W(y)}$  increases with  $y$ .

If  $\beta\rho < 1$ , then at the top, as  $n \rightarrow 0$ ,

$$\varepsilon^c = \frac{\tilde{\varepsilon}^c}{1 - \beta\rho - \beta\rho\tilde{\varepsilon}^c} > \tilde{\varepsilon}^c, \quad (45)$$

since Lemma 2 ensures that the denominator is positive. We see that the adjustment is increasing in the superstar parameter  $\beta$  and vanishes if  $\beta = 0$ . For a given earnings distribution and hence parameter  $\rho$ , this implies a lower upper bound on the set of Pareto efficient top marginal tax rates  $\tau$  (based on, for instance, (17) in the absence of income effects).

If  $\beta\rho > 1$ , as supported by Gabaix-Landier, then, since  $\beta\rho n^{\frac{1}{\rho}-\beta} \rightarrow \infty$  as  $n \rightarrow 0$ , the

---

and Sleet, 2015, in their empirical analysis). In that case, and using their preferred elasticity of .25, Diamond and Saez (2011) find a top tax rate of 73% to be revenue-maximizing. If this elasticity ignores superstar effects and  $\beta = 1$ , however, the correct elasticity is .4 and the revenue-maximizing top tax rate falls to 62%.

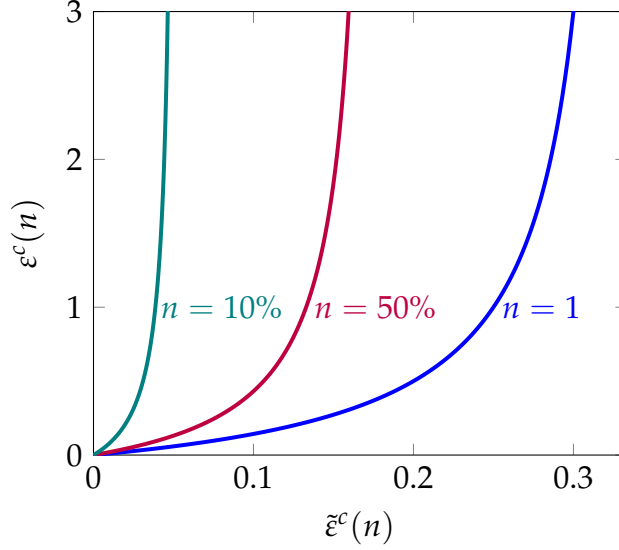


Figure 6: Adjusted earnings elasticities for effort elasticities at bottom, median and top 10%.

adjustment becomes unbounded for any given  $\tilde{\varepsilon}^c$ . Erroneously using such an unbounded earnings elasticity in (17) would appear to imply that only a zero top tax rate can be Pareto efficient. Figure 6 illustrates this. It shows the relationship between the earnings elasticity and the effort elasticity (44) at a given rank  $n$ , for the bottom, median and top 10% and  $\beta = 1, \rho = 3$ . But the adjustment explodes as we move to effort elasticities of higher-ranked individuals; this reflects the fact that  $\frac{W''(y)y}{W'(y)}$  is unbounded. Conversely, any given earnings elasticity requires lower and lower effort elasticities as we move up the earnings distribution. Indeed, this is illustrating Lemma 2. Recall that, when  $\beta\rho > 1$ , holding  $\tilde{\varepsilon}^c$  fixed as  $n \rightarrow 0$  is incompatible with an equilibrium. Instead, we require

$$\tilde{\varepsilon}^c(y) \leq \frac{1}{\frac{W''(y)y}{W'(y)}} = \frac{\beta\rho - 1}{\beta\rho} \frac{b - y}{y},$$

so  $\tilde{\varepsilon}^c(y)$  must vanish towards the top at least at the rate of  $(b - y)/y$ . The specification in Example B features precisely this property and

$$\tilde{\varepsilon}(y) = \frac{1}{\gamma + 1} \frac{b - y}{y} \rightarrow 0 \quad \text{as } y \rightarrow b.$$

Naively using this effort elasticity in the efficiency test (17), we would erroneously conclude that the upper bound to the set of Pareto efficient top marginal tax rates is 100%. However, substituting in (44) reveals that the earnings elasticity is in fact constant and equals

$$\varepsilon = \frac{1}{\gamma(\beta\rho - 1) - 1} > 0. \quad (46)$$

Using this correct elasticity, the efficiency test for the top marginal tax rate (17) implies a well-defined upper bound on  $\tau$  strictly less than 1 greater than 0.

**Macro Elasticities.** Our examples also deliver simple expressions for the macro elasticity adjustments from Proposition 4. For the iso-elastic preferences in Example A with a Pareto skill distribution, Appendix B.2 shows that, at the top, the general equilibrium correction in (30) converges to

$$\chi = -\frac{\beta\rho}{(\gamma-1)(1-\beta\rho)},$$

which is negative as predicted by the second part of Proposition 4. Combining this with (29), (30), (43) and (45) allows us to derive the macro elasticity at the top

$$\bar{\varepsilon} = \hat{\varepsilon} \left( 1 + \frac{\chi}{\hat{\varepsilon}} (1 - \beta\rho) \right) = \frac{1}{\gamma - 1}.$$

Thus, the macro elasticity at the top coincides with the effort elasticity. In sum, the relationships between the four elasticity concepts are therefore  $\tilde{\varepsilon} = \bar{\varepsilon} < \hat{\varepsilon} < \varepsilon$ . Similar results can be shown to apply for Example B, where at the top one finds that  $\bar{\varepsilon} < \tilde{\varepsilon} = 0 < \hat{\varepsilon} < \varepsilon$ .

### 5.3 Illustrating Result 3

We finally illustrate Result 3 using our examples. To do so, we return to the efficiency condition (17), expressed in terms of the Pareto parameter for earnings  $\rho$  and the earnings elasticity  $\varepsilon$ . Writing both parameters in terms of primitives, we obtain in Example A

$$\rho = \frac{\gamma-1}{\gamma} \frac{\alpha}{\alpha\beta+1} \quad \text{and} \quad \varepsilon = \frac{1+\alpha\beta}{\gamma-1};$$

while in Example B

$$\rho = \frac{\gamma+1}{\gamma} \frac{\alpha}{\alpha\beta-1} \quad \text{and} \quad \varepsilon = \frac{\alpha\beta-1}{\gamma+1}.$$

Superstar effects change both  $\rho$  and  $\varepsilon$ . The Pareto parameter  $\rho$  of the *earnings* distribution is *decreasing* in  $\beta$ , i.e. superstar effects lead to fatter tails in the earnings distribution; the earnings elasticity is *increasing* in  $\beta$ , given the primitive parameters  $\alpha$  and  $\gamma$ . The neutrality property in Result 3 obtains because both these changes precisely cancel out: in condition (17), the term  $\varepsilon\rho = \alpha/\gamma$  is independent of  $\beta$ , and we obtain the equivalent condition (32).

## 6 Conclusion

This paper extends the Mirrlees optimal taxation model by incorporating an assignment problem between workers and firms. As first shown by [Rosen \(1981\)](#), assignment models are able to produce superstar effects, where innate skill differences are greatly magnified in terms of earnings. Despite their potential to dramatically increase inequality, our results show that these effects do not provide a basis for higher taxation. Depending on the interpretation, they are either neutral or provide a force for lower taxes.

It is worth clarifying that this force does not imply that taxes should be low or lower than current tax rates. Other forces or considerations, unrelated to superstars, may be at play shaping the optimal level of taxes. These include the “Stiglitz effects” and the inefficiencies, for example from rent-seeking, discussed in the introduction. Some top incomes may be due to luck, providing insurance motives for taxes. All these issues arise even in standard Mirrlees models without superstar effects. To isolate the role of superstars, we have, therefore, ignored them here.

One may argue that superstars at the top of the earnings distribution have very low labor supply elasticities, because their pay and effort are so high or because they are intrinsically motivated. Of course, this argument for higher taxes at the top is again unrelated to the presence of superstar effects and is also present in a standard Mirrleesian model. Whatever the effort elasticity of top earners, our Result 2 shows that this elasticity needs to be adjusted upwards when accounting for superstar effects. For instance, our Example B in Section 5 features an effort elasticity that converges to zero at the top, yet the correct earnings elasticity is a positive constant.

## References

- [Ales, Laurence and Christopher Sleet](#), “Taxing Top Incomes,” *Working Paper*, 2015. [1](#), [16](#), [25](#)
- , [Andres Bellofatto](#), and [Jessie Wang](#), “Taxing Atlas: Using Firm Data to Derive Optimal Income Tax Rates,” *Working Paper*, 2015. [1](#), [4.1](#)
- [Ales, Laurence, Musab Kurnaz, and Christopher Sleet](#), “Tasks, Talents, and Taxes,” *American Economic Review*, 2015, 105, 3061–3101. [1](#)
- [Auten, Gerald and Robert Carroll](#), “The Effect of Income Taxes on Household Income,” *Review of Economics and Statistics*, 1999, 81, 681–693. [15](#)
- [Blundell, Richard, Alan Duncan, and Costas Meghir](#), “Estimating Labor Supply Responses Using Tax Reforms,” *Econometrica*, 1998, 66, 827–862. [3.2.3](#)

- Bovenberg, L. and B. Jacobs**, “Redistribution and Education Subsidies are Siamese Twins,” *Journal of Public Economics*, 2005, 89, 2005–2035. 5
- Chetty, R., J. Friedman, T. Olsen, and L. Pistaferri**, “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records,” *Quarterly Journal of Economics*, 2011, 126, 749–804. 14
- Chetty, Raj**, “Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply,” *Econometrica*, 2012, 80 (969-1018). 3.1.1
- Davis, Steven and Magnus Henrekson**, “Tax Effects on Work Activity, Industry Mix and Shadow Economy Size: Evidence From Rich-Country Comparisons,” in Salvador et al., ed., *Labour Supply and Incentives to Work in Europe*, Edward Elgar Press, 2005. 15
- Diamond, Peter and Emmanuel Saez**, “The case for a progressive tax: from basic research to policy recommendations,” *The Journal of Economic Perspectives*, 2011, pp. 165–190. 12, 25
- **and James Mirrlees**, “Optimal Taxation and Public Production: I–Production Efficiency,” *American Economic Review*, March 1971, 61 (1), 8–27. 2.1
- **and –**, “Optimal Taxation and Public Production II: Tax Rules,” *American Economic Review*, 1971, 61, 261–78. 3.3, 17
- Edmans, Alex and Xavier Gabaix**, “Executive Compensation: A Modern Primer,” *Working Paper*, 2015. 1, 1
- Eissa, Nada and Hilary Hoynes**, “The Earned Income Tax Credit and the Labor Supply of Married Couples,” *NBER Working Paper 6856*, 1998. 3.2.3
- Feenberg, Daniel and James Poterba**, “Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns,” in James Poterba, ed., *Tax Policy and the Economy*, Vol. 7, MIT Press, 1993, pp. 145–177. 15
- Feldstein, Martin**, “The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act,” *Journal of Political Economy*, 1995, 103, 551–572. 15
- Fuchs, William, Luis Garicano, and Luis Rayo**, “Optimal Contracting and the Organization of Knowledge,” *Review of Economic Studies*, 2015, forthcoming. 4.2, 4.2, 4.2
- Gabaix, Xavier and Augustin Landier**, “Why Has CEO Pay Increased So Much?,” *Quarterly Journal of Economics*, 2008, 123, 49–100. 1, 1, 2.1, 3.2.2, 3.2.3, 5, 25

- Garicano, Luis**, “Hierarchies and the Organization of Knowledge in Production,” *Journal of Political Economy*, October 2000, 108 (5), 874–904. [1](#)
- **and Esteban Rossi-Hansberg**, “Inequality and the Organization of Knowledge,” *American Economic Review*, 2004, 94, 197–202. [4.2](#), [4.2](#)
- **and –**, “Organization and Inequality in a Knowledge Economy,” *Quarterly Journal of Economics*, 2006, 121, 1383–1435. [4.2](#), [21](#)
- **and –**, “Knowledge-Based Hierarchies: Using Organizations to Understand the Economy,” *Annual Review of Economics*, 2015, 7, 1–30. [1](#), [1](#), [20](#)
- Goolsbee, A.**, “What happens when you tax the rich? Evidence from executive compensation,” *Journal of Political Economy*, 2000, 108, 352–378. [3.2.2](#)
- Gruber, Jon and Emmanuel Saez**, “The Elasticity Of Taxable Income: Evidence And Implications,” *Journal of Public Economics*, 2002, 84, 1–32. [15](#)
- Hausman, Jerry**, “Taxes and Labor Supply,” in Alan Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, Vol. 1, Elsevier, 1985, chapter 4, pp. 213–263. [3.2.3](#)
- Jaffe, Sonia and Scott Kominers**, “Taxation in Matching Markets,” *Working Paper*, 2014. [4.3](#)
- Keane, Michael**, “Labor Supply and Taxes: A Survey,” *Journal of Economic Literature*, 2011, 49, 961–1075. [3.1.1](#)
- Kopczuk, Wojciech**, “Tax Bases, Tax Rates and the Elasticity of Reported Income,” *Journal of Public Economics*, 2005, 89, 2093–2119. [15](#)
- Lucas, Robert**, “On the Size Distribution of Business Firms,” *Bell Journal of Economics*, 1978, 9, 508–523. [1](#), [4.1](#)
- MaCurdy, Thomas**, “An Empirical Model of Labor Supply in a Life-Cycle Setting,” *Journal of Political Economy*, 1981, 89, 1059–1085. [3.2.3](#)
- Mirrlees, James**, “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 1971, 38 (2), 175–208. [1](#), [2.1](#), [17](#)
- Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva**, “Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities,” *American Economic Journal: Economic Policy*, 2014, 6, 230–271. [1](#)
- Prescott, Edward**, “Why Do Americans Work so Much More Than Europeans?,” *Federal Reserve Bank of Minneapolis Quarterly Review*, 2004, 28, 2–13. [15](#)

- Rosen, Sherwin**, “The Economics of Superstars,” *American Economic Review*, 1981, 71, 845–858. [1](#), [2](#), [4](#), [18](#), [6](#)
- , “Authority, Control, and the Distribution of Earnings,” *Bell Journal of Economics*, 1982, 13, 311–323. [1](#), [4.1](#)
- Rothschild, Casey and Florian Scheuer**, “Redistributive Taxation in the Roy Model,” *Quarterly Journal of Economics*, 2013, 128, 623–668. [1](#)
- **and** – , “Optimal Taxation with Rent-Seeking,” *Review of Economic Studies*, 2016, forthcoming. [1](#)
- Saez, Emmanuel**, “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 2001, 68 (1), 205–29. [1](#), [1](#), [3.2.3](#)
- , “Reported Incomes and Marginal Tax Rates, 1960-2000,” in James Poterba, ed., *Tax Policy and the Economy*, Vol. 18, MIT Press, 2004, pp. 117–174. [15](#)
- , “Do Taxpayers Bunch at Kink Points?,” *American Economic Journal: Economic Policy*, 2010, 2, 180–212. [3.2.3](#)
- , **Joel Slemrod**, **and Seth Giertz**, “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Perspectives*, 2012, 50, 3–50. [3.1.1](#), [3.2.2](#)
- Samuelson, Paul**, *The Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press, 1947. [3.2.2](#)
- Sattinger, Michael**, “Assignment Models of the Distribution of Earnings,” *Journal of Economic Literature*, 1993, 91, 831–880. [1](#)
- Scheuer, Florian**, “Entrepreneurial Taxation with Endogenous Entry,” *American Economic Journal: Economic Policy*, 2014, 6, 126–163. [1](#)
- **and Iván Werning**, “Mirrlees Meets Diamond-Mirrlees,” *Working Paper*, 2015. [17](#)
- Slemrod, Joel**, “High-Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response,” in Martin Feldstein and James Poterba, eds., *The Empirical Foundations of Household Taxation*, University of Chicago Press, 1996, pp. 169–189. [3.2.2](#), [15](#)
- Stantcheva, S.**, “Optimal Taxation and Human Capital Policies over the Life Cycle,” *Mimeo, Harvard University*, 2016. [5](#)

**Stiglitz, Joseph**, “Self-Selection and Pareto Efficient Taxation,” *Journal of Public Economics*, 1982, 17, 213–240. **1**

**Terviö, Marko**, “The difference that CEOs make: An assignment model approach,” *The American Economic Review*, 2008, pp. 642–668. **1, 1, 2, 2.1**

**Werning, Iván**, “Pareto Efficient Income Taxation,” *Mimeo, MIT*, 2007. **1, 3.1, 3.1.2, 3.1.2, 3.3**

**Ziliak, James and Thomas Kniesner**, “Estimating Life Cycle Labor Supply Tax Effects,” *Journal of Political Economy*, 1999, 107, 326–359. **3.2.3**

## A Appendix: Omitted Proofs

### A.1 Proof of Lemma 1

Take any allocation satisfying conditions (1), (7) and (8). Take any  $W(y)$  solving (5) and define  $T$  by

$$T(W(y(\theta))) = c(\theta) - W(y(\theta)),$$

for all observed equilibrium earnings levels  $\{W(y(\theta)) : \theta \in \Theta\}$  (set  $T$  sufficiently high otherwise, for all off-equilibrium earning levels).

Conversely, any equilibrium implies a resource-feasible incentive-compatible allocation  $(c(\theta), y(\theta))$  with  $\sigma(\theta) = G^{-1}(F(\theta))$ . In particular, given  $T$  and the equilibrium earnings schedule  $W$ ,  $y(\theta)$  must solve

$$\max_y U(W(y) - T(W(y)), y, \theta).$$

and imply the allocation  $c(\theta) = W(y(\theta)) - T(W(y(\theta)))$ . Thus,

$$U(c(\theta), y(\theta), \theta) = \max_y U(W(y) - T(W(y)), y, \theta) \geq \\ U(W(y(\theta')) - T(W(y(\theta'))), y(\theta'), \theta) = U(c(\theta'), y(\theta'), \theta),$$

which establishes incentive compatibility, (7) and (8).

### A.2 Proof of Proposition 1

**Pareto problem.** The Pareto problem (11) s.t. (7), (9) and (12) can be rewritten as

$$\max_{c, V} \int V(\theta) d\Lambda(\theta)$$

s.t.

$$V'(\theta) = U_\theta(e(V(\theta), y(\theta), \theta), y(\theta), \theta) \quad \forall \theta$$

and

$$\int (B(\theta, y(\theta)) - e(V(\theta), y(\theta), \theta)) dF(\theta) \geq 0,$$

where  $e(V, y, \theta)$  is the inverse function of  $U(c, y, \theta)$  w.r.t. its first argument. The corresponding Lagrangian is, after integrating by parts,



$$\begin{aligned}\mathcal{L} &= \int V(\theta)d\Lambda(\theta) + \eta \int [B(\theta, y(\theta)) - e(V(\theta), y(\theta), \theta)] dF(\theta) \\ &\quad - \int \mu'(\theta)V(\theta)d\theta - \int \mu(\theta)U_\theta(e(V(\theta), y(\theta), \theta), y(\theta), \theta)d\theta,\end{aligned}$$

where  $\mu(\theta)$  are the multipliers on the incentive constraints and  $\eta$  on the resource constraint.

**First-order conditions.** The first-order condition for  $V$  is (using  $e_V = 1/U_c$ )

$$-\mu'U_c - \mu U_{\theta c} = \eta f - U_c \lambda \leq \eta f,$$

where we dropped arguments. Define  $\hat{\mu} \equiv U_c \mu / \eta$ , so

$$\hat{\mu}' = \frac{U_c \mu'}{\eta} + \frac{\mu}{\eta} [U_{c\theta} + U_{cc}c' + U_{cy}y'].$$

Substituting this yields

$$-\hat{\mu}' + \hat{\mu} \frac{U_{cc}c' + U_{cy}y'}{U_c} \leq f.$$

Note that

$$\frac{U_{cc}c' + U_{cy}y'}{U_c} = \frac{U_{cc} \frac{c'}{y'} + U_{cy}}{U_c} y' = \frac{-U_{cc}U_y + U_{cy}U_c}{U_c^2} y' = -\frac{\partial}{\partial c} \left[ -\frac{U_y}{U_c} \right] y' = -MRS_c y'$$

since  $c'/y' = -U_y/U_c = MRS$  by the local incentive constraints. Hence, the first-order condition for  $V$  implies

$$-\hat{\mu}' - \hat{\mu} MRS_c y' \leq f. \quad (47)$$

The first-order condition for  $y$  is (using  $e_y = MRS$ )

$$\frac{B_y - MRS}{MRS} f = \frac{\hat{\mu}}{U_c} \frac{U_{\theta c} MRS + U_{\theta y}}{MRS}.$$

Recall that, using (5), the necessary condition for (2) implies

$$\tau(\theta) \equiv T'(W(y(\theta))) = 1 - \frac{MRS(c(\theta), y(\theta), \theta)}{B_y(\theta, y(\theta))},$$

so this becomes

$$\frac{\tau}{1-\tau} f = \frac{\hat{\mu}}{U_c} \frac{U_{\theta c} MRS + U_{\theta y}}{MRS}.$$

Also,

$$-U_{\theta c} \frac{U_y}{U_c} + U_{\theta y} = U_c \frac{U_{y\theta} U_c - U_{c\theta} U_y}{U_c^2} = -U_c \frac{\partial}{\partial \theta} \left[ -\frac{U_y}{U_c} \right] = -U_c \frac{\partial MRS}{\partial \theta},$$

so the first-order condition for  $y$  finally becomes

$$\frac{\tau}{1-\tau} f = -\hat{\mu} \frac{\partial \log MRS}{\partial \theta}. \quad (48)$$

**Elasticities.** To rewrite these conditions for Pareto efficiency in terms of elasticities, we compute the earnings elasticities defined in Section 3.1.1. Recall the earnings function  $w(1 - \tau, I)$  defined by

$$\max_w U((1 - \tau)w - T(w) + I, W^{-1}(w), \theta)$$

with first-order condition

$$MRS((1 - \tau)w - T(w) + I, W^{-1}(w), \theta) = (1 - \tau - T'(w))W'(W^{-1}(w)).$$

This implies the uncompensated earnings elasticity (dropping arguments)

$$\varepsilon^u = \left. \frac{\partial w}{\partial(1 - \tau)} \right|_{\tau=I=0} \frac{1 - T'}{w} = \frac{W'/w - MRS_c}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1 - T'} W'}. \quad (49)$$

Moreover,

$$\eta \equiv -(1 - T') \left. \frac{\partial w}{\partial I} \right|_{\tau=I=0} = \frac{MRS_c}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1 - T'} W'}$$

and

$$\varepsilon^c = \varepsilon^u + \eta = \frac{W'/w}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1 - T'} W'}. \quad (50)$$

As a result, for later use,

$$\frac{\varepsilon^c - \varepsilon^u}{\varepsilon^c} = \frac{\eta}{\varepsilon^c} = MRS_c \frac{w}{W'} \quad (51)$$

and

$$\frac{1}{w\varepsilon^c} = \frac{1}{W'} \left( MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} \right) + \frac{T''}{1 - T'}. \quad (52)$$

**Identification.** To relate the first-order conditions to the equilibrium earnings distribution, note that, for an equilibrium earnings schedule  $w(\theta) = W(y(\theta))$ , the earnings distribution  $H$  satisfies  $H(w(\theta)) \equiv F(\theta)$  and therefore

$$f(\theta) = h(w(\theta))w'(\theta).$$

The individuals' problem for a given earnings tax  $T(w)$  is

$$\max_w U(w - T(w), W^{-1}(w), \theta)$$

with first-order condition

$$MRS(w - T(w), W^{-1}(w), \theta) = (1 - T'(w))W'(W^{-1}(w)).$$

Differentiating this w.r.t.  $w$  yields

$$w'(\theta) = \frac{-\partial \log MRS / \partial \theta}{\frac{1}{W'} \left( MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} \right) + \frac{T''}{1 - T'}} = -\frac{\partial \log MRS}{\partial \theta} w\varepsilon^c(w) \quad (53)$$

by (52). Next, define  $\hat{\mu}(w) \equiv \mu(w^{-1}(w))$  where  $w^{-1}(w)$  is the inverse of  $w(\theta)$ , so

$$\hat{\mu}'(\theta) = \hat{\mu}'(w(\theta))w'(\theta).$$

Using this and  $w'(\theta) = W'(y(\theta))y'(\theta)$  in (47) yields

$$-\hat{\mu}'(w) - \hat{\mu}(w) \frac{\partial MRS(w)}{\partial c} \frac{1}{W'(W^{-1}(w))} \leq h(w). \quad (54)$$

Similarly, we can rewrite (48) using (53) as

$$\hat{\mu} = \frac{\frac{\tau}{1-\tau} h w'}{-\frac{\partial \log MRS}{\partial \theta}} = \frac{\tau}{1-\tau} h w \varepsilon^c(w)$$

and so

$$\hat{\mu}(w) = \frac{T'(w)}{1-T'(w)} \varepsilon^c(w) w h(w). \quad (55)$$

**Test Inequality.** Taking logs of (55) and differentiating w.r.t.  $\log w$  yields

$$\frac{\hat{\mu}'(w)w}{\hat{\mu}(w)} = \frac{d \log \hat{\mu}(w)}{d \log w} = \frac{d \log \left( \frac{T'(w)}{1-T'(w)} \varepsilon^c(w) \right)}{d \log w} + \frac{d \log h(w)}{d \log w} + 1. \quad (56)$$

Returning to (54), multiply through by  $w/\hat{\mu}(w)$

$$-\frac{\hat{\mu}'(w)w}{\hat{\mu}(w)} - \frac{\partial MRS}{\partial c} \frac{w}{W'} \leq h(w) \frac{w}{\hat{\mu}(w)}.$$

Substitute (55) on the RHS and (56) on the LHS, and use (51) for the income effect, to get

$$\frac{T'(w)}{1-T'(w)} \varepsilon^c(w) \left[ -\frac{d \log \left( \frac{T'(w)}{1-T'(w)} \varepsilon^c(w) \right)}{d \log w} - \frac{d \log h(w)}{d \log w} - 1 - \frac{\eta(w)}{\varepsilon^c(w)} \right] \leq 1.$$

### A.3 Proof of Proposition 2

We obtain the fixed-assignment elasticities as a special case of (49) and (50) replacing  $W$  by  $\hat{W}_0$  and evaluating at  $w_0, y_0$ . Using  $\hat{W}'_0(y_0) = W'(y_0)$ , this yields

$$\hat{\varepsilon}^\mu = \frac{W'/w_0 - MRS_c}{MRS_c + \frac{MRS_y}{MRS} + \frac{T''}{1-T'} W' - \frac{\hat{W}''_0}{W'}}$$

$$\hat{\eta} = \frac{MRS_c}{MRS_c + \frac{MRS_y}{MRS} + \frac{T''}{1-T'} W' - \frac{\hat{W}''_0}{W'}}$$

and

$$\hat{\varepsilon}^c = \hat{\varepsilon}^\mu + \hat{\eta} = \frac{W'/w_0}{MRS_c + \frac{MRS_y}{MRS} + \frac{T''}{1-T'} W' - \frac{\hat{W}''_0}{W'}}. \quad (57)$$

This immediately implies  $\hat{\eta}/\hat{\varepsilon}^c = MRS_c w_0/W' = \eta/\varepsilon^c$  and hence the second part of the proposition. For the first, use the fact that, by (6),

$$W''(y_0) = B_{y\theta}(\theta_0, y_0)\Gamma'(y_0) + B_{yy}(\theta_0, y_0) = B_{y\theta}(\theta_0, y_0)\Gamma'(y_0) + \hat{W}''_0(y_0).$$

Hence, we can write

$$\varepsilon^c = \frac{W'/w_0}{MRS_c + \frac{MRS_y}{MRS} + \frac{T''}{1-T'}W' - \frac{\hat{W}_0''}{W'} - \frac{B_{y\theta}\Gamma'}{W'}} = \left( \frac{1}{\hat{\varepsilon}^c} - \frac{B_{y\theta}\Gamma'w_0}{W'^2} \right)^{-1} = \frac{\hat{\varepsilon}^c}{\hat{\Phi}}.$$

Finally, note that the workers' first-order condition can be written as

$$(W'(1 - T') - MRS) U_c = 0,$$

so the necessary second-order condition for a maximum is

$$MRS_c + \frac{MRS_y}{MRS} + \frac{T''}{1-T'}W' - \frac{W''}{W'} \leq 0.$$

This implies

$$\frac{W'/w_0}{\hat{\varepsilon}^c} \geq \frac{W'' - \hat{W}_0''}{W'} = \frac{B_{y\theta}\Gamma'}{W'} \quad \text{and hence} \quad \hat{\varepsilon}^c \frac{B_{y\theta}\Gamma'w_0}{W'^2} \leq 1 \quad \Leftrightarrow \quad \hat{\Phi} \geq 0.$$

#### A.4 Proof of Lemma 2

The individual's problem is  $\max_y U((1 - \tau)W(y), y, \theta)$ .

$$((1 - \tau)W'(y) - MRS(W(y), y, \theta)) U_c(W(y), y, \theta) = 0.$$

The second-order condition is (after using the first-order condition and dropping arguments)

$$(1 - \tau)W'' - (MRS_c W' + MRS_y) \leq 0$$

or, using (58),

$$\frac{W''y}{W'} \leq MRS_c y + \frac{yMRS_y}{MRS} = \frac{1}{\varepsilon^c}.$$

#### A.5 Proof of Proposition 3

The effort elasticities can be obtained from (49), (50) and (51) as a special case for  $W(y) = y$  and  $T'(w) = \tau$ . Then we have

$$\tilde{\varepsilon}^c = \frac{1/y}{MRS_c + \frac{MRS_y}{MRS}} \quad \text{and} \quad \frac{\tilde{\eta}}{\tilde{\varepsilon}^c} = MRS_c y. \quad (58)$$

Comparing the former with (50) and following the same steps as in the proof of Proposition 2 yields (23) and (25). Comparing the latter with (51) yields (24). The inequalities follow, first, from  $W'' > 0$  since  $W$  is convex by (6) when  $A$  is linear in  $y$ . Second,  $W'y/w < 1$  again by convexity of  $W(y)$  and  $W(0) = 0$  because  $0 \leq W(0) \leq A(x, 0) = 0$ . Finally,  $\hat{\Phi} > 0$  follows from Lemma 2.

## A.6 Proof of Proposition 4

Under the assumption that  $B(\theta, y) = b(\theta)y$ , we have  $\bar{W}_0^{-1}(w|\tau) = (w + \pi(\theta_0|\tau))/b(\theta_0)$ , so the first-order condition corresponding to (27) is

$$(1 - \tau - T'(w))b(\theta_0) = \phi_y \left( \frac{w + \pi(\theta_0|\tau)}{b(\theta_0)}, \theta \right),$$

which yields (using  $b = W'$  and  $\phi_y = MRS$ )

$$\bar{\varepsilon}(w_0) = \frac{d\bar{w}_0}{d(1 - \tau)} \Big|_{\tau=0} \frac{1 - T'(w_0)}{w_0} = \frac{W'/w_0}{\frac{MRS_y}{MRS} + \frac{T''}{1-T'} W'} \left( 1 - \frac{w_0}{W'y} \frac{MRS_y y}{MRS} \frac{\partial \pi(\theta_0|\tau)}{\partial(1 - \tau)} \Big|_{\tau=0} \frac{1 - T'}{w_0} \right).$$

The first result follows from observing that (i) the term in front of the brackets equals  $\hat{\varepsilon}$  when utility is quasilinear in  $c$  and  $B$  is linear in  $y$  (by comparison with (57)), (ii)  $MRS_y y / MRS = 1/\hat{\varepsilon}$  by (58), and (iii), by (26),

$$\frac{\partial \pi(\theta_0|\tau)}{\partial(1 - \tau)} \Big|_{\tau=0} = - \frac{\partial \bar{W}_0(y_0|\tau)}{\partial(1 - \tau)} \Big|_{\tau=0}.$$

For the second part, write the problem of an individual of type  $\theta$  as

$$y(\theta|\tau) \in \arg \max_y (1 - \tau)b(\theta)y - T(b(\theta)y - \pi(\theta|\tau)) - \phi(y, \theta). \quad (59)$$

By Topkis's theorem, we know that (i)  $y$  is increasing in  $1 - \tau$  for each  $\theta$  when holding  $\pi$  fixed, and that (ii)  $y$  is increasing in  $\pi$  for each  $\theta$  holding  $\tau$  fixed when  $T'' \geq 0$ . Next, from the firms' problem (3) and by the envelope theorem,  $\pi'(\theta|\tau) = b'(\theta)y(\theta|\tau) \forall \theta$  and, since profits are zero for the lowest firm,

$$\pi(\theta|\tau) = \int_{\underline{\theta}}^{\theta} b'(s)y(s|\tau)ds. \quad (60)$$

Hence,  $\pi$  is increasing in  $1 - \tau$  for each  $\theta$  if  $y(\theta|\tau)$  is increasing in  $1 - \tau$  for each  $\theta$ . Therefore, the fixed point of (59) and (60), i.e., the schedules  $y(\theta|\tau)$  and  $\pi(\theta|\tau)$  that satisfy (59) and (60) simultaneously given  $\tau$ , involves a profit schedule such that  $\partial \pi(\theta|\tau) / \partial(1 - \tau) \Big|_{\tau=0} > 0$  if  $T'' \geq 0$ . This implies  $\chi < 0$ .

## A.7 Proof of Proposition 5

Recall the first-order conditions for  $V$  and  $y$  from Appendix A.2, given by (47) and (48), and rewrite them as

$$-\hat{\mu} MRS_c y' = -\hat{\mu} \frac{\tilde{\eta}}{\tilde{\varepsilon}^c} \frac{y'}{y} = f - \lambda U_c / \eta + \hat{\mu}' \leq f + \hat{\mu}' \quad (61)$$

and

$$\hat{\mu} = \frac{\tau}{1 - \tau} \theta f \left( -\frac{MRS_{\theta} \theta}{MRS} \right)^{-1} = \frac{\tau}{1 - \tau} \theta f \xi, \quad (62)$$

where we used  $\tilde{\eta} / \tilde{\varepsilon}^c = MRS_c y$  and defined  $\xi(\theta) \equiv (-\partial \log MRS(c(\theta), y(\theta), \theta) / \partial \log \theta)^{-1} \geq 0$ . Take logs of (62) and differentiate w.r.t.  $\log \theta$

$$\frac{\hat{\mu}' \theta}{\hat{\mu}} = \frac{d \log \left( \frac{\tau}{1 - \tau} \right)}{d \log \theta} + 1 + \frac{d \log f}{d \log \theta} + \frac{d \log \xi}{d \log \theta}.$$

Multiplying (61), dropping the Pareto weights, by  $\theta/\hat{\mu}$  yields

$$\frac{\tau}{1-\tau}\zeta \left[ -\frac{d \log \left( \frac{\tau}{1-\tau} \right)}{d \log \theta} - 1 - \frac{d \log f}{d \log \theta} - \frac{d \log \zeta}{d \log \theta} - \frac{\tilde{\eta}}{\tilde{\varepsilon}^c} \frac{y'\theta}{y} \right] \leq 1.$$

If preferences are such that  $U(c, y, \theta) = u(c, y/\theta)$ , then, defining  $l = y/\theta$ ,

$$MRS(c, y, \theta) = -\frac{u_l(c, y/\theta)}{\theta u_c(c, y/\theta)}$$

and it can be shown with some algebra that

$$-\frac{MRS_{\theta\theta}}{MRS} = 1 + y \frac{MRS_y}{MRS}.$$

Using (58), this implies  $\zeta = \tilde{\varepsilon}^c / (1 + \tilde{\varepsilon}^u)$  as claimed in the text.

### A.8 Proof of Proposition 6

For general Pareto weights  $\Lambda$ , (61) is the differential equation

$$-\hat{\mu} \frac{\tilde{\eta}}{\tilde{\varepsilon}^c} \frac{y'}{y} = f - \lambda U_c / \eta + \hat{\mu}',$$

which we can solve for  $\hat{\mu}$ :

$$\hat{\mu}(\theta) = \int_{\theta}^{\infty} \left( 1 - \frac{\lambda(s)U_c(s)}{\eta f(s)} \right) \exp \left( \int_{\theta}^s \frac{\tilde{\eta}(t)}{\tilde{\varepsilon}^c(t)} \frac{dy(t)}{y(t)} \right) dF(s).$$

Substituting this in (62) yields the result.

## B Online Appendix

### B.1 Proofs for Section 4

#### B.1.1 Proof of Lemma 3

By the same arguments as in Section 2.3, the Pareto problem for the nonlinear profit tax on managers and the lump-sum tax on workers is

$$\max_{c(t), y(t), c^w, \omega} \lambda^w u(c^w) + \int \lambda(t) U(c(t), y(t), F^{-1}(t)) dt$$

s.t.

$$U(c(t), y(t), F^{-1}(t)) \geq U(c(t'), y(t'), F^{-1}(t)) \quad \forall t, t' \quad (63)$$

$$\int c(t) dt + c^w \leq \int A(L(y(t)|\omega), y(t)) dt \quad (64)$$

$$\int L(y(t)|\omega) dt = 1, \quad (65)$$

where  $c^w$  is the consumption of workers,  $u$  is their utility of consumption,  $c(t)$  is the consumption allocated to managers of talent  $t$ ,  $\lambda^w$  and  $\lambda(t)$  are the respective Pareto weights, and the last equation (65) is the labor market clearing condition. Observe that the wage  $\omega$  only appears in the last two constraints (64) and (65). This allows us to rewrite the Pareto problem as follows:

$$\max_{c(t), y(t), c^w} \lambda^w u(c^w) + \int \lambda(t) U(c(t), y(t), F^{-1}(t)) dt \quad (66)$$

s.t. (63) and

$$\int c(t) dt + c^w \leq \Psi(\{y\}), \quad (67)$$

where

$$\Psi(\{y\}) \equiv \max_{\omega} \int A(L(y(t)|\omega), y(t)) dt \quad \text{s.t.} \quad \int L(y(t)|\omega) dt = 1. \quad (68)$$

Hence,  $\Psi(\{y\})$  is the equilibrium aggregate output in the economy for any given effort schedule  $y(t)$  for managers. With this decomposition, and for any given  $c^w$ , we see that the Pareto problem (66) subject to (63) and (67) is the same as our original Pareto problem (11) subject to (8) and (12) from Section 2.3 when we replace  $\int B(t, y(t)) dt$  by  $\Psi(\{y\})$ .

The first-order condition corresponding to (68) is

$$\int A_L(L(y(t)|\omega), y(t)) \frac{\partial L(y(t)|\omega)}{\partial \omega} dt - \eta \int \frac{\partial L(y(t)|\omega)}{\partial \omega} dt = 0,$$

where  $\eta$  is the multiplier on the labor market clearing constraint (65). The first-order condition corresponding to (34) is

$$A_L(L(y(t)|\omega), y(t)) = \omega \quad \forall t.$$

Hence  $\eta = w$ . Next,

$$\begin{aligned} \frac{\partial \Psi(\{y\})}{\partial y(t)} &= A_y(L(y(t)|\omega), y(t)) + A_L(L(y(t)|\omega), y(t)) L'(y(t)|\omega) - \eta L'(y(t)|\omega) \\ &= A_y(L(y(t)|\omega), y(t)) = W'(y(t)|\omega). \end{aligned}$$

### B.1.2 Proof of Lemma 4

By property (iv) in Section 4.2, for a given threshold  $t_1$  and effort schedule  $y$ , managers of type  $t_m$  will be matched with workers of type  $t_p = P(t_m)$ , where the matching function  $P(t)$  satisfies

$$h \int_{P(t)}^{t_1} (1 - y(s)) ds = \int_t^1 ds \quad \forall t \leq 1. \quad (69)$$

This constraint requires that, for any manager type  $t \leq 1$ , the total time available to managers in the interval  $[t, 1]$  equals the total time required to help with the problems left unsolved by workers in the interval  $[P(t), t_1]$ . Conditional on  $t_1$  and a  $y$ -schedule, there is a unique function  $P(t)$  that satisfies this constraint, namely, the function that solves the differential equation

$$P'(t) = \frac{1}{h(1 - y(P(t)))} \quad \forall t \leq 1, \quad (70)$$

with initial condition  $P(1) = t_1$ , where (70) follows from differentiating (69) w.r.t.  $t$ . Observe that this then also uniquely pins down the manager cutoff  $t_2$ , namely such that  $P(t_2) = 0$ .

As discussed in Section 4.2, (70) implies that the effort schedule  $y$ , and hence the tax schedule  $T$ , now affects the equilibrium assignment of workers to managers:  $P'(t)$  is increasing in  $y(P(t))$ .

We can also verify that the equilibrium earnings schedule is convex. The firm's first-order condition for  $y_m$  in (36) implies

$$W'_m(y(t)) = \frac{1}{h(1 - y(P(t)))}, \quad (71)$$

which is increasing along the  $y$ -schedule among managers. The first-order condition for  $y_p$  yields

$$w'_p(y_p) = \frac{y_m - w_p(y_p)}{1 - y_p}.$$

Moreover, by free entry, firms make zero profits, so

$$w_p(y_p) = y_m - W_m(y_m)h(1 - y_p).$$

Substituting this and using the equilibrium assignment function  $P$ , we obtain

$$w'_p(y(P(t))) = W_m(y(t))h, \quad (72)$$

which is also increasing along the  $y$ -schedule. Of course, the earnings schedule is linear in the intermediate segment with the self-employed between  $y(t_1)$  and  $y(t_2)$ , if it exists, leading to an overall earnings schedule  $W(y)$  that is weakly convex.<sup>26</sup>

To prove Lemma 4, note that, by the same arguments as in Section 2.3, the Pareto problem for the earnings tax  $T$  can now be written as:

$$\max_{c, y, P, t_1, t_2} \int_0^1 \lambda(t) U(c(t), y(t), F^{-1}(t)) dt$$

<sup>26</sup>Conditional on an effort schedule  $y$  and the thresholds  $t_1$  and  $t_2$ , the overall earnings schedule  $W(y)$  is entirely pinned down by (71) and (72) as well as the indifference conditions  $W_m(y(t_2)) = y(t_2)$  and  $w_p(y(t_1)) = y(t_1)$  at the thresholds.



s.t.

$$\int_0^1 c(t)dt \leq \int_{t_2}^1 \frac{y(t)}{h(1-y(P(t)))} dt + \int_{t_1}^{t_2} y(t)dt, \quad (73)$$

(63), (70) and  $P(1) = t_1, P(t_2) = 0$ . The first integral on the right-hand side of (73) is total output produced by matched agents (integrating over managers), and the second is total output produced by the self-employed.

A key observation is again that  $P, t_1$  and  $t_2$  only enter constraints (70) and (73), which allows us to decompose the Pareto problem as follows:

$$\max_{c,y} \int_0^1 \lambda(t)U(c(t), y(t), F^{-1}(t))dt \quad (74)$$

s.t. (63) and

$$\int_0^1 c(t)dt \leq \Psi(\{y\}), \quad (75)$$

where

$$\Psi(\{y\}) = \max_{P,t_1,t_2} \int_{t_2}^1 \frac{y(t)}{h(1-y(P(t)))} dt + \int_{t_1}^{t_2} y(t)dt \quad (76)$$

s.t. (70) and  $P(1) = t_1, P(t_2) = 0$ . Hence,  $\Psi(\{y\})$  is again aggregate output in the economy under the equilibrium assignment for any given (monotone) effort schedule  $y$ .<sup>27</sup> The Pareto problem (74) s.t. (63) and (75) is therefore again identical to our original Pareto problem (11) subject to (8) and (12) from Section 2.3 when we replace  $\int B(t, y(t))dt$  by  $\Psi(\{y\})$ . Since all our results are based on the first-order conditions of the planning problem, they all go through if the partial derivative of this aggregate output w.r.t.  $y(t)$  coincides with the marginal earnings in equilibrium at that effort level.

To see why this is indeed the case, we write the Lagrangian corresponding to (76), integrating by parts, as

$$\mathcal{L} = \int_{t_2}^1 \frac{y(t)}{h(1-y(P(t)))} dt + \int_{t_1}^{t_2} y(t)dt - \int_{t_2}^1 \mu'(t)P(t)dt - \int_{t_2}^1 \frac{\mu(t)}{h(1-y(P(t)))} dt.$$

For any  $t \in (t_2, 1)$ , we therefore immediately have  $\partial\Psi(\{y\})/\partial y(t) = 1/[h(1-y(P(t)))]$ . Together with (71), this implies  $\partial\Psi(\{y\})/\partial y(t) = W'_m(y(t))$  for the managers as desired.

As for workers, let  $M(t)$  denote the inverse of  $P(t)$ , with

$$M'(t) = h(1-y(t)) \quad \forall t \in [0, t_1] \quad (77)$$

and  $M(t_1) = 1, M(0) = t_2$ . Using this, it is convenient to rewrite the output from matched agents by integrating over workers rather than managers:

$$\int_{t_2}^1 \frac{y(t)}{h(1-y(P(t)))} dt = \int_{P(t_2)}^{P(1)} \frac{y(M(t))}{h(1-y(t))} h(1-y(t))dt = \int_0^{t_1} y(M(t))dt.$$

Hence, we can rewrite  $\Psi$  as

$$\Psi(\{y\}) = \max_{M,t_1,t_2} \int_0^{t_1} y(M(t))dt + \int_{t_1}^{t_2} y(t)dt$$

---

<sup>27</sup>By single-crossing and (63), any incentive compatible effort schedule  $y$  must be monotone, so we will evaluate  $\Psi$  only for monotone schedules.

s.t. (77) and  $M(t_1) = 1, M(0) = t_2$ . The corresponding Lagrangian is

$$\begin{aligned} \mathcal{L} = & \int_0^{t_1} y(M(t))dt + \int_{t_1}^{t_2} y(t)dt + \mu(t_1)M(t_1) - \mu(0)M(0) - \int_0^{t_1} \mu'(t)M(t)dt \\ & - \int_0^{t_1} \mu(t)h(1-y(t))dt + \xi_1(M(t_1) - 1) + \xi_2(M(0) - t_2). \end{aligned}$$

The first-order condition for  $M(t), t \in (0, t_1)$ , is  $\mu'(t) = y'(M(t))$ , the first-order condition for  $M(0)$  is  $\mu(0) = \xi_2$ , and the first-order condition for  $t_2$  is  $y(t_2) = \xi_2$ . Hence,  $\mu(0) = y(t_2)$  and

$$\mu(t) = \int_0^t y'(M(s))ds + y(t_2).$$

For any  $t \in (0, t_1)$  we therefore have

$$\frac{\partial \Psi(\{y\})}{\partial y(t)} = h\mu(t) = h \int_0^t y'(M(s))ds + hy(t_2).$$

In addition, we have  $W_m(y(t_2)) = y(t_2)$ , which pins down  $W_m$  entirely:

$$W_m(y(t)) = \int_{t_2}^t W'_m(y(s))y'(s)ds + y(t_2) = \int_{t_2}^t \frac{y'(s)}{h(1-y(P(s)))}ds + y(t_2)$$

or

$$W_m(y(M(t))) = \int_0^t \frac{y'(M(s))}{h(1-y(s))}h(1-y(s))ds + y(t_2) = \int_0^t y'(M(s))ds + y(t_2).$$

Hence,  $\partial \Psi(\{y\})/\partial y(t) = hW_m(y(M(t)))$ . Together with (72), this shows that  $\partial \Psi(\{y\})/\partial y(t) = w'_p(y(t))$  for workers as desired.

Finally, as for the self-employed, trivially  $\partial \Psi(\{y\})/\partial y(t) = 1 = W'(y(t))$ .

## B.2 Proofs for Section 5

### B.2.1 Example A

**Exact Earnings Schedule and Approximate Pareto Skill Distribution.** We begin with Example A in Section 5 and construct the skill distribution  $F(\theta)$  as claimed in the text. Taking  $W(y)$  as given by (40), individuals solve

$$\max_y (1 - \tau)W(y) - \frac{1}{\gamma} \left( \frac{y}{\theta} \right)^\gamma.$$

The first-order condition is

$$(1 - \tau) \frac{\tilde{\kappa}}{1 - \beta\rho} \left( \frac{\kappa\beta\rho}{1 - \beta\rho} + y \right)^{\frac{1}{1-\beta\rho}-1} = y^{\gamma-1}\theta^{-\gamma},$$

where we used the fact that  $b$  is set such that profits of the lowest firm  $x = 0$  are zero, so  $A(0, y(1)) = y(1) = w(1) = \kappa$  and hence  $b = \kappa\beta\rho/(\beta\rho - 1)$ , and  $\tilde{\kappa} = \kappa^{\frac{\beta\rho}{\beta\rho-1}} |\beta\rho - 1|^{\frac{-1}{\beta\rho-1}}$ . The second-order condition for a maximum is satisfied if  $\gamma(1 - \beta\rho) > 1$ . Substituting  $y(n)$  from (39) for  $y$  yields

$$\theta = (1 - \tau)^{-\frac{1}{\gamma}} \left( \frac{\kappa}{1 - \beta\rho} \right)^{\frac{\gamma-1}{\gamma}} \left( n^{\frac{\beta}{\gamma-1}} \left( n^{-\frac{1-\beta\rho}{\rho}} - \beta\rho \right) \right)^{\frac{\gamma-1}{\gamma}}.$$

This implicitly defines  $n = 1 - F(\theta)$  given  $\theta$  and hence  $F(\theta)$ . Rewriting this as

$$\theta = (1 - \tau)^{-\frac{1}{\gamma}} \left( \frac{\kappa}{1 - \beta\rho} \right)^{\frac{\gamma-1}{\gamma}} \left( n^{-\frac{\gamma(1-\beta\rho)-1}{(\gamma-1)\rho}} - \beta\rho n^{\frac{\beta}{\gamma-1}} \right)^{\frac{\gamma-1}{\gamma}}$$

reveals that, if  $\gamma(1 - \beta\rho) > 1$ , then for sufficiently small for  $n \approx 0$  (i.e., at the top of the skill distribution)

$$\theta \approx (1 - \tau)^{-\frac{1}{\gamma}} \left( \frac{\kappa}{1 - \beta\rho} \right)^{\frac{\gamma-1}{\gamma}} n^{-\frac{\gamma(1-\beta\rho)-1}{\gamma\rho}}.$$

Hence,  $n = 1 - F(\theta) \approx (\underline{\theta}/\theta)^\alpha$  for some  $\underline{\theta}$ , with

$$\alpha = \frac{\gamma\rho}{\gamma(1 - \beta\rho) - 1}.$$

In other words,  $\theta$  has a Pareto tail with parameter  $\alpha > 0$ . Conversely, solving for  $\rho$  yields (41) and, since  $\alpha > 0$ , we indeed must have  $\gamma(1 - \beta\rho) > 1$ .

**Exact Pareto Skill Distribution and Approximate Earnings Schedule.** We next derive the equilibrium earnings schedule for Example A when  $F(\theta) = 1 - (\underline{\theta}/\theta)^\alpha$ , that is,  $F$  is a Pareto distribution for all  $\theta \geq \underline{\theta}$ . We do so by guessing and verifying a schedule of the form  $W(y) = ky^\delta + w_0$  for some parameters  $k, \delta, w_0$ . Taking  $\tau$  and  $W(y)$  as given, individuals solve

$$\max_y (1 - \tau)ky^\delta - \frac{1}{\gamma} \left( \frac{y}{\theta} \right)^\gamma.$$

The first-order condition yields the  $y$ -schedule

$$y(\theta) = \theta^{\frac{\gamma}{\gamma-\delta}} (\delta k(1 - \tau))^{\frac{1}{\gamma-\delta}} \quad (78)$$

as well as its inverse  $\Gamma(y) = y^{\frac{\gamma-\delta}{\gamma}} (\delta k(1 - \tau))^{-\frac{1}{\gamma}}$ . Note that we require  $\delta < \gamma$ , which we will verify below. We know from the assignment condition (5) that  $W'(y) = B_y(\Gamma(y), y)$ , so using  $B(\theta, y) = \underline{\theta}^{-\alpha\beta} \theta^{\alpha\beta} y$ , this becomes

$$\delta ky = y^{\alpha\beta \frac{\gamma-\delta}{\gamma}} (\delta k(1 - \tau))^{-\frac{\alpha\beta}{\gamma}} \underline{\theta}^{-\alpha\beta}.$$

Matching coefficients requires  $\delta k = (\delta k(1 - \tau))^{-\frac{\alpha\beta}{\gamma}} \underline{\theta}^{-\alpha\beta}$  and  $\delta - 1 = \alpha\beta \frac{\gamma-\delta}{\gamma}$ . Solving this yields

$$\delta = \gamma \frac{\alpha\beta + 1}{\alpha\beta + \gamma}$$

as claimed in the text and

$$k = \frac{1}{\gamma} \frac{\alpha\beta + \gamma}{\alpha\beta + 1} (1 - \tau)^{-\frac{\alpha\beta}{\gamma + \alpha\beta}} \underline{\theta}^{-\frac{\alpha\beta\gamma}{\gamma + \alpha\beta}}.$$

Observe that we indeed have  $\delta < \gamma$  because  $\gamma > 1$  by assumption. Using these expressions for  $\delta$  and  $k$  in (78) yields the equilibrium effort schedule  $y(\theta)$ . Finally,  $w_0$  is determined such that profits of the lowest firm  $x = 0$  are zero, so  $B(\underline{\theta}, y(\underline{\theta})) = y(\underline{\theta}) = W(y(\underline{\theta})) = ky(\underline{\theta})^\delta + w_0$ .

Using  $n = 1 - F$  and (78), we can write  $y(n) = Cn^{-\frac{\alpha\beta + \gamma}{\alpha(\gamma-1)}} = Cn^{\beta - \frac{1}{\rho}}$  for some constant  $C$ , where the second step used (41). Hence, (39) holds approximately for small enough  $n$ . Again using (41), we have  $\delta = 1/(1 - \beta\rho)$ , so  $W(y) = ky^{\frac{1}{1-\beta\rho}} + w_0$ . For large enough  $y$ , this approximates (40). This finally implies that (38) also holds approximately.

### B.2.2 Example B

We guess and verify an earnings schedule of the form  $W(y) = k(b - y)^{-\delta}$  for some constants  $k, \delta > 0$ . Taking  $\tau$  and  $W(y)$  as given, individuals solve

$$\max_y (1 - \tau)k(b - y)^{-\delta} - \frac{1}{\gamma} \left( \frac{a - \theta}{b - y} \right)^\gamma$$

with first-order condition

$$\frac{(b - y)^{\delta - \gamma}}{(1 - \tau)\delta k} = (a - \theta)^{-\gamma}. \quad (79)$$

From the equilibrium condition (5) and  $B(\theta, y) = (a - \theta)^{-\alpha\beta}y$ , we have

$$W'(y) = B_y(\Gamma(y), y) = (a - \Gamma(y))^{-\alpha\beta},$$

so

$$k\delta(b - y)^{-\delta - 1} = [(1 - \tau)\delta k]^{-\frac{\alpha\beta}{\gamma}} (b - y)^{\frac{(\delta - \gamma)\alpha\beta}{\gamma}}.$$

Matching coefficients requires  $k\delta = [(1 - \tau)\delta k]^{-\frac{\alpha\beta}{\gamma}}$  and  $\delta + 1 = (\gamma - \delta)\frac{\alpha\beta}{\gamma}$ , which we can solve for  $\delta$ :

$$\delta = \gamma \frac{\alpha\beta - 1}{\alpha\beta + \gamma}$$

and similarly for  $k$ :

$$k = (1 - \tau)^{-\frac{\alpha\beta}{\gamma + \alpha\beta}} \frac{\gamma + \alpha\beta}{\gamma(\alpha\beta - 1)}.$$

Note that, since we assumed  $\alpha\beta > 1$ , we have  $\delta, k > 0$ . Using this in (79), effort is given by

$$y(\theta) = b - (1 - \tau)^{-\frac{1}{\gamma + 1}} (a - \theta)^{\frac{\gamma + \alpha\beta}{\gamma + 1}} = b - (1 - \tau)^{-\frac{1}{\gamma + 1}} n^{\frac{\gamma + \alpha\beta}{\alpha(\gamma + 1)}} = b - (1 - \tau)^{-\frac{1}{\gamma + 1}} n^{\beta - \frac{1}{\rho}},$$

where the last step used (42). Hence, we match (39) for some appropriate  $\kappa$ . The same is true for (40) and hence (38).