The Taxation of Superstars*

Florian Scheuer
Stanford University

Iván Werning
MIT

July 2015

How are optimal taxes affected by the presence of superstar phenomena at the top of the earnings distribution? 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 are not sufficient statistics and 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 tax rates at the top 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. This raises the obvious question: should superstars pay superstar taxes?

The goal of this paper is to provide an answer to this question by studying optimal nonlinear taxation of earnings in the presence of superstar effects. Following Mirrlees (1971), a

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

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 superstar effects into the standard optimal taxation 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 tax rates. Perhaps surprisingly, our results are precisely the opposite. Depending on how one poses the exercise, we find that the 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 whether or not the economy features superstar effects?

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

Question I holds the observed distribution of earnings fixed. The motivation for doing so is straightforward, since at any moment it is a datum to those evaluating the taxation of earnings, whether or not superstar effects are present. Indeed, the distribution of earnings 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 of earnings. Hence, it is imperative to consider hypothetical scenarios that turn superstar effects on and off, so to speak, while holding fixed the distribution of earnings. Doing so establishes whether superstar effects matter per se, in addition to their impact on the distribution of earnings. For these reasons, Question I is our favored perspective for thinking about the taxation of superstars, since it directly informs current policy choices, which already condition on the earnings distribution, by isolating the impact of superstar effects.

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 typically rises when superstar effects emerge (after all, this is what the superstars literature is all about). It is natural to explore such a comparative static exercise and it is of obvious theoretical appeal, but perhaps less relevant for current policy choices, which condition on the observed earnings distribution. The comparative static becomes policy relevant, however, 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. Our first two results are geared towards Question I, that is, taking the distribution of earnings as given we ask whether superstar effects matter. Our first result is a neutrality result of sorts.

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 first 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 two sufficient statistics that emerge in the standard Mirrlees setting, where superstar effects are absent. In fact, not only are the same sufficient statistics involved, but the conditions for efficient taxes and the formulas for optimal taxes are identical.

How should one interpret this result? Conceptually, according to this neutrality result, superstars do not inherently create a force for higher or lower tax rates—once we condition on the correct sufficient statistics optimal tax formulas are unchanged. In particular, it does not matter if superstars contribute towards inequality, as long as we take this greater inequality into account, by conditioning on the observed distribution of earnings. Similarly, it does not matter if superstars affect the elasticity of earnings, as long as we condition on the appropriate values for these elasticities. In this sense, the standard Mirrlees model's conditions for optimal taxes and insights carry over and are robust to the introduction of superstars.

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—a situation that seems rather unlikely—then one must recognize that the required earnings elasticities are affected by the presence of superstar effects.

Result 2. Superstar effects *increase* earnings elasticities, given underlying preferences or common measures of elasticities, which omit superstar effects; this therefore 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.[1]

Why is it that superstar effects push individual earnings elasticities up? There are several forces at work. Indeed, the required adjustments depend on the type of elasticity one has in hand to start with. We discuss three possible elasticity concepts and the forces leading to adjustments in each case.

The most basic force has to do with 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.[2] Thus, earnings elasticities that do not correctly capture this reallocation, which we call "fixed-assignment" elasticities, are too low. In practice, this is especially likely when elasticities are measured off of temporary and unexpected tax changes.

Matters are only made worse if one starts with earnings elasticities for non-superstar workers, for whom earnings are proportional to effort. Empirically, this may correspond to workers that are not near the top of the distribution. This then identifies what we call "effort" elasticities. Suppose we are willing to extrapolate such effort elasticities to workers in superstar positions. However, before doing so, we must take into account that for superstar workers earnings are a convex function of effort that goes through the origin, so that any percentage change in effort induces a greater percentage change in earnings. As a result, we must also adjust the earnings elasticity upwards, relative to the effort elasticity. Although this additional mechanical bias is not due to larger behavioral responses, it is still relevant when extrapolating earnings elasticities from effort elasticities.

Finally, an important consideration is that we require elasticities with respect to individual tax changes. In contrast, often one starts with evidence from economy-wide tax changes, or "macro" elasticities. Economy-wide changes in tax rates induce general equilibrium effects. We show that for progressive tax schedules these general equilibrium effects also create a further bias, relative to elasticities based on a fixed assignment. In our benchmark

[1] Of course, this does not imply that taxes should be low in absolute terms, or lower than current levels. It establishes a robust downward force on optimal taxes, coming from superstar effects, compared to a standard economy with no superstars.

[2] We adopt a broad notion of effort here, one that is certainly not limited to hours of work for superstar earners who have "made it," but also incorporates their entire lifetime choices, such as human capital accumulation or anything else it took them to obtain their job.

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 the neutrality implication in Result 1 demands a number of stringent 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. Whenever these conditions are not met, Result 2 provides the necessary upwards adjustment for each of the three elasticity measures discussed above.

An important distinguishing feature of our perspective is that we express the optimality conditions in terms of sufficient statistics, namely, the distribution of earnings and the individual elasticities of earnings. This allows us to relate the required elasticities in the presence of superstars to those that are likely to be estimated and to provide the adjustment factors necessary. Indeed, we use the parametrization of Gabaix and Landier (2008), who provide an application to CEOs, to inform the required adjustments, which turn out significant. For instance, for an elasticity of .25 that ignores job reassignment, the correct earnings elasticity is 1. This would imply an upper bound to the top marginal tax rate of 25% instead of 57%. Based on the previous discussion, even greater adjustments are required for effort and macro elasticities.

We said earlier that it may seem natural to expect superstar effects to push for higher taxes. After all, superstars exacerbate inequality, leading to earnings that are out of proportion with underlying skills. On first pass, one may expect taxes to be higher, to regain a balance between equality and efficiency. Yet, we just arrived at the opposite conclusion, showing that superstar effects provide a force for *lower* tax rates. This obtains for two intuitive reasons. First, we are conditioning on the earnings distribution, which soaks up any effects that superstars may have on inequality. Indeed, in line with the spirit of Question I, it would not make sense to do otherwise, since the earnings distribution is directly observable and is already an input into standard optimal tax analysis. Second, as we discussed, superstar effects increase the elasticity of earnings to tax changes. When taken together, we are in a situation with higher elasticity for given inequality, so the tradeoff between efficiency and equality is disturbed by superstars, but tilted towards efficiency, rather than equality. This provides the force for lower tax rates.

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 provides a very different neutrality result. Instead of holding the earnings distribution fixed, it considers a change in technology, such as the introduction of superstar effects, which generally increases inequality. This 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 a comparative static of this sort 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 also be affected by where they are evaluated. Nevertheless, when preferences have constant elasticity and when marginal social welfare weights are left unchanged, then one can conclude that marginal tax rates are unchanged. In particular, this is the case at the top of the distribution, when one applies a zero weight at the top, e.g. utilitarian or Rawlsian social welfare functions.

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 the introduction of 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 positive literature on assignment models that are able to generate superstar effects, surveyed in Sattinger (1993) and more recently Edmans and Gabaix (2015) and Garicano and Rossi-Hansberg (2015). This literature has been interested in how these models can explain observed patterns of occupational choice and income inequality. Our contribution is to 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-on-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 later 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 knowledge-based hierarchy models in Garicano (2000) and Garicano and Rossi-Hansberg (2015).

Our paper incorporates a richer labor market into the canonical Mirrlees model. The standard Mirrlees model proceeds with the most basic of labor markets, where exogenous wage differences simply reflect exogenous skill differences. In contrast, labor economists have developed much richer general equilibrium models of the labor market. Much research has enriched the standard Mirrlees model in various of these directions, for example, by considering labor indivisibilities, human capital accumulation, uncertainty, or sorting across different occupations, as in the Roy model, to name just a few.

Recent contributions focus on models with worker sorting across jobs and their general equilibrium effects. In particular, Rothschild and Scheuer (2013) and Ales et al. (2014) show how endogenous sectoral wages and assignments affect optimal income tax schedules. 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 the models by Scheuer (2014) and Ales et al. (2015), which 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, as we point out, our adaptation is different in crucial aspects.

While the present paper is part of this greater agenda extending the standard optimal tax model to incorporate important features of real-world labor markets, it is also 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 often taken to be 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 sources of inefficiencies, such as unproductive rent-seeking or positive spillovers from innovation. For example, some top incomes may be the result of contest-like tournaments with winner-takes-all compensation (Rothschild and Scheuer, 2014a). Likewise, some CEOs may capture the board, due to poor corporate governance, and set their own pay at excessive levels, effectively stealing from stockholders (Piketty et al., 2014). On the other hand, some top earners, such as innovators, may be unable to capture their entire 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, which are our focus.

Perhaps equally importantly, we also take a fundamentally different approach in deriving our results. The literature so far 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. occupational patterns or the firm size distribution. 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 are likely to be estimated in standard settings. We can then show how these elasticities are affected by the presence of superstar effects. The effects of superstars on taxes work exclusively through the sufficient 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 any given tax schedule in place 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 observed earnings distribution fixed when characterizing the effects of superstars. In contrast, computing optimal taxes for given redistributive motives requires taking into account that the equilibrium income 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 earnings 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 sets up the planning problem for Pareto efficient taxes. Sections 4, 5 and 6 are devoted to Results 1, 2 and 3, respectively. Section 7 discusses extensions of our baseline model and Section 8 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).[3] In Section 7, we discuss other settings, including first- and second-generation span-of-control models. All our results go through in these more general settings.

[3] 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 later.

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 θ. The utility function for type θ 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. We assume standard conditions on the utility function, including differentiability with $U_c > 0$, $U_y < 0$ and the 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 θ, so that more able agents (higher θ) 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 the cumulative distribution function $F(\theta)$ with density $f(\theta)$ and support Θ.

Technology. There is a unit measure of firms indexed by x, distributed according to the cumulative distribution function $G(x)$.[4] There is a single final consumption good. Production of this final good takes place through one-to-one matching between individuals and firms. If an individual that provides y units of effort is matched with a firm of type x, then the output of this match equals

$$A(x, y),$$

where the function A is non-negative, increasing, twice differentiable and strictly supermodular: $A_{xy}(x, y) > 0$.

An allocation specifies $c(\theta), y(\theta)$ and a measure-preserving assignment rule, mapping

[4]Without loss of generality, one may normalize the distribution of x to be uniform on $[0, 1]$, so that x can be interpreted as the quantiles of the distribution of some underlying firm characteristic.

worker θ to firm $x = \sigma(\theta)$.[5] The resource constraint is then

$$\int c(\theta)dF(\theta) \leq \int A(\sigma(\theta), y(\theta))dF(\theta) + E, \tag{1}$$

where $E \in \mathbb{R}$ is an endowment net of government consumption.

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 $\gamma(y) = y$. Our examples fall into this class.

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

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

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

Firm x maximizes profits, solving

$$\max_{y} \{A(x,y) - W(y)\}. \tag{3}$$

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

$$A_y(x,y) = W'(y). \tag{4}$$

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

[5] A measure-preserving transformation σ is such that $\mu_F(\sigma^{-1}(X)) = \mu_G(X)$ for any subset of firms X, where $\sigma^{-1}(X)$ denotes the set of workers θ with $\sigma(\theta) \in X$, and μ_F, μ_G are the measures induced by the cumulative distributions F and G, respectively.

[6] Profits are possible because we have assumed a fixed distribution of firms. If instead we modeled endogenous entry of firms, this would dissipate ex ante or average profits, negating their distributional effects. Such an extension would take us away from the standard existing assignment models and is beyond the scope of the present paper.

Positive Assortative Matching. As we have noted, on the worker side y and θ 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 θ, given by an increasing assignment function $\sigma(\theta)$. We incorporate this fact in the definition of an equilibrium below.

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:

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

The last condition (iv) ensures that the match between worker θ and firm $x = \sigma(\theta)$ is consistent with their respective supply and demand choices, $y(\theta) = Y(\sigma(\theta))$, and that the measure of firms demanding $Y(x)$ or less, which equals $G(x)$, coincides with the measure of workers supplying $y(\theta)$ or less, which equals $F(\theta)$. Equivalently, we can write the equality of supply and demand for each skill level in differential form as $f(\theta) = g(\sigma(\theta))\sigma'(\theta)$. This is a common way of writing the market-clearing condition in hedonic markets.

By Walras' law, the resource constraint implies that the government budget constraint

$$\int T(W(y(\theta)))dF(\theta) + E + \Pi = 0$$

holds, where $\Pi \equiv \int (A(x, Y(x)) - W(Y(x)))dG(x)$ are aggregate profits.

2.2 Superstar Effects

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

$$W'(y) = A_y(\sigma(\Gamma(y)), y). \tag{5}$$

For a given effort schedule $y(\theta)$, this condition pins down the earnings schedule $W(y)$ up to a constant.[7] This well-known marginal condition, which obtains in a wide set of assign-

[7] 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 $W(0) = 0$.

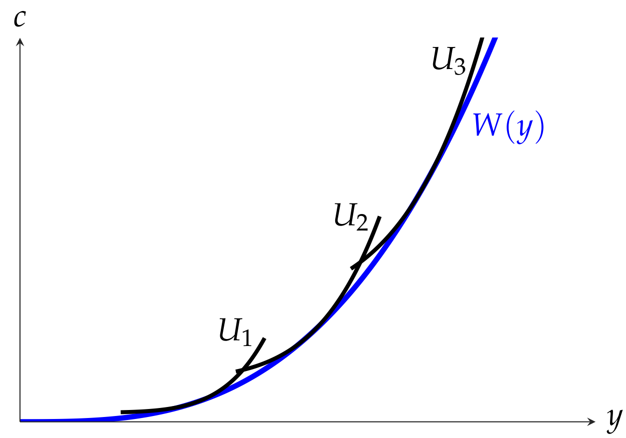

Figure 1: Convex equilibrium earnings schedule.

ment models, is fundamental to the superstar effects on the earnings schedule. To see this, differentiate to obtain

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

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 Two Examples based on Gabaix-Landier

Before turning to optimal taxation, it is useful to illustrate superstar equilibria in our model. We do so by adapting the CEO application in Gabaix and Landier (2008). Given their focus, their framework takes the distribution of effort y as given ("talent" in their terminology). We provide two specific examples with endogenous effort that deliver this distribution in equilibrium. We return to Gabaix-Landier's specification and our two examples repeatedly.

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 cdf), we have near the top[8]

$$w(n) = \kappa n^{-1/\rho}, \tag{7}$$

where ρ is the Pareto parameter.

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,

$$y(n) = b - \frac{\kappa}{\beta\rho - 1} n^{\beta - \frac{1}{\rho}}. \tag{8}$$

When $\beta\rho > 1$, effort is bounded above by b.[9] Otherwise, effort is unbounded above. Combining equations (7) with (8) gives the increasing and convex earnings schedule

$$W(y) = \begin{cases} \tilde{\kappa}\,(b-y)^{\frac{-1}{\beta\rho-1}} & \text{if } \beta\rho > 1 \\ \tilde{\kappa}\,(y-b)^{\frac{1}{1-\beta\rho}} & \text{if } \beta\rho < 1 \end{cases} \tag{9}$$

for some constant $\tilde{\kappa}$,[10] in the range of equilibrium values $y \in [y(1), y(0)]$.

For some purposes, the Gabaix-Landier specification (7)–(9) 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 (7)–(9). 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. Thus, we find it helpful to span both cases: our first example is tailored to the case $\beta\rho < 1$, while our second captures the case $\beta\rho > 1$.[11]

[8]That is, for n near zero. The four relations that follow may be understood as approximations, i.e. $w(n) \approx \kappa n^{-1/\rho}$ for small n. However, to simplify and avoid confusion, it is best to impose them exactly for all $n \in [0,1]$.

[9]b is set such that profits of firm $x = 0$ are zero: $A(0, y(1)) = y(1) = w(1) = \kappa$ and hence $b = \kappa\beta\rho/(\beta\rho - 1)$.

[10]Namely, $\tilde{\kappa} = \kappa^{\frac{\beta\rho}{\beta\rho-1}} |\beta\rho - 1|^{\frac{-1}{\beta\rho-1}} > 0$.

[11]To simplify, we ignore the intermediate case where $\beta\rho = 1$. As Gabaix-Landier discuss, there are three families of distributions, from extreme value theory, that can describe the tail behavior of the effort distribution: bounded for $\beta\rho > 1$, fat-tailed unbounded for $\beta\rho < 1$, and thin-tailed unbounded for $\beta\rho = 1$. Given the preferences in our examples below, the same is true for the tail of the distribution of skills.

Example Economy A ($\beta\rho < 1$). Our first example is most suitable for $\beta\rho < 1$ and 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 linear tax schedule with constant marginal tax rate τ (or the top bracket of a nonlinear schedule). Appendix A shows that (7)–(9) hold for a skill distribution with a Pareto tail, so that $F(\theta) \approx 1 - (\underline{\theta}/\theta)^{\alpha}$ for high θ, where $\alpha > 0$ is the Pareto parameter. In addition,

$$\rho = \frac{\gamma - 1}{\gamma} \frac{\alpha}{\alpha\beta + 1}, \tag{10}$$

so that the Pareto parameter ρ is *decreasing* in β for fixed γ and α, which illustrates that superstar effects induce a *fatter* tail and hence a more unequal earnings distribution in this sense.[12] Note that equation (10) implies $\beta\rho < 1$.

An alternative to backing out the skill distribution that implies (7)–(9) exactly is to assume a Pareto distribution throughout, i.e. $F(\theta) = 1 - (\underline{\theta}/\theta)^{\alpha}$; then, as shown in the appendix, (7)–(9) hold approximately at the top.

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

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

for some $a, b, \gamma > 0$. This utility satisfies standard assumptions: it is decreasing and concave in $y \leq b$ and satisfies the single-crossing property. We will see later that it also has other desirable (in fact, necessary) features, within the case $\beta\rho > 1$. In particular, disutility from effort goes to infinity as $y \to b$, which is crucial to rationalize bounded effort choices with unbounded earnings.

Consider the equilibrium under a constant marginal tax rate τ. Appendix A shows that (7)–(9) hold for skill distribution $F(\theta) = 1 - (a - \theta)^{\alpha}$ for $\theta \leq a$, with $\alpha\beta > 1$. In addition,

$$\rho = \frac{\gamma + 1}{\gamma} \frac{\alpha}{\alpha\beta - 1}, \tag{11}$$

[12] The distribution of effort y is also Pareto in equilibrium, but with parameter $(\gamma - 1)\alpha/(\alpha\beta + \gamma)$, which exceeds the Pareto parameter of the earnings distribution when $\beta > 0$, implying less inequality in effort than in earnings.

which is decreasing in β, so that earnings inequality is again increasing in β. Note that (11) implies $\beta\rho > 1$. In particular, we can always find α and γ so that $\rho = 3$ and $\beta = 1$, consistent with Gabaix-Landier's favored parameterization.

3 Efficient Taxes and Planning Problem

We now turn to the normative study of optimal or efficient taxes T. We first develop a planning problem that can be used to characterize efficient taxes. Our results in later sections exploit the optimality conditions of this planning problem.

3.1 Planning Problem

We first establish a connection between incentive compatible allocations and allocations that are part of an equilibrium with some taxes T. This implementation result allows us to formulate a planning problem in terms of allocations directly.

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

$$V(\theta) = U(c(\theta), y(\theta), \theta). \tag{12}$$

An allocation is *incentive compatible* when

$$V(\theta) = \max_{\theta'} U(c(\theta'), y(\theta'), \theta) \qquad \forall\theta. \tag{13}$$

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 (12)–(13) and $\sigma(\theta) = G^{-1}(F(\theta))$.*

The proof is contained in Appendix B.1. The key to the 'only if' part is that, although the earnings schedule is determined by the labor market and the government cannot observe or 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 (12) and (13) are equivalent to the local constraints

$$V'(\theta) = U_\theta(c(\theta), y(\theta), \theta) \ \ \forall\theta \tag{14}$$

and the requirement

$$y(\theta) \text{ is non-decreasing.} \tag{15}$$

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

$$\max_{c,y,V} \int V(\theta)\, d\Lambda(\theta) \tag{16}$$

subject to (12), (14), (15) and

$$\int (B(\theta, y(\theta)) - c(\theta))\, dF(\theta) + E \geq 0 \tag{17}$$

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 θ 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 refer to an earnings tax T as Pareto efficient if it induces a Pareto efficient allocation. We will consider cases where constraint (15) 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 additional output function B in the resource constraint, which summarizes production by agent θ *in equilibrium.* In other words, 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 model the equilibrium assignment is unaffected by the rest of the allocation $c(\theta), y(\theta)$.[13]

However, this does not imply that assignment plays no role. Indeed, *off the equilibrium* a single agent θ 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 θ. 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.

Explicit and Implicit Marginal Tax Rates. Using (5), the necessary condition for (2) implies

$$B_y(\theta, y(\theta))(1 - T'(W(y(\theta)))) = MRS(c(\theta), y(\theta), \theta).$$

[13]This is not crucial for our analysis. As we show in Section 7, our results extend to settings where the equilibrium assignment is endogenous to taxes.

This condition allows us to back out the implied marginal earnings taxes from any incentive compatible allocation:

$$\tau(\theta) \equiv T'(W(y(\theta))) = 1 - \frac{MRS(c(\theta), y(\theta), \theta)}{B_y(\theta, y(\theta))}. \tag{18}$$

4 Result 1: Neutrality

We now investigate the implications of superstar effects for taxes. We use the first-order conditions of the Pareto problem (16) 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.[14] Our main interest is in how this test differs from a standard Mirrlees economy, and how the additional parameters in our model, such as the properties of technology that govern superstar effects, enter it.

4.1 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 uncompensated earnings function by

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

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

$$\varepsilon^u(w) = \left.\frac{\partial w}{\partial(1-\tau)}\right|_{\tau=I=0} \frac{1-T'(w)}{w}. \tag{19}$$

Income effects on earnings are captured by the parameter

$$\eta(w) = -\left.\frac{\partial w}{\partial I}\right|_{\tau=I=0} (1 - T'(w)), \tag{20}$$

[14] Under additional assumptions, such as when preferences take the separable form $U(c, y, \theta) = u(c) - \phi(y, \theta)$ as in Examples 1 and 2, a transformed version of the Pareto problem can be shown to be convex, so the first-order conditions are sufficient and any allocation that passes the test is Pareto efficient.

and the *compensated earnings elasticity* leaves utility constant with

$$\varepsilon^c(w) = \varepsilon^u(w) + \eta(w), \tag{21}$$

by the Slutsky equation.

Three observations are in order. First, these are elasticities for *earnings*, so they take into account the shape of the equilibrium earnings schedule $W(y)$.[15] 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 5 and contrast them with other elasticity concepts.

4.2 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.

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

$$\frac{T'(w)}{1-T'(w)}\varepsilon^c(w)\left(\rho(w) - \frac{d\log\left(\frac{T'(w)}{1-T'(w)}\varepsilon^c(w)\right)}{d\log w} - \frac{\eta(w)}{\varepsilon^c(w)}\right) \leq 1 \tag{22}$$

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 B.2, reveals three, perhaps surprising, insights. First, for a given tax schedule T, the observed distribution of earnings h as well as the earnings elasticities ε^c and ε^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, although, as we argue below, identifying it is challenging.

[15] This is in contrast to pure *effort* elasticities. Our earnings elasticities are therefore more closely related to the empirical literature on *taxable income elasticities* than, for instance, traditional labor supply elasticities as measured by hours. Keane (2011), Chetty (2012), Keane and Rogerson (2012) and Saez et al. (2012) provide recent surveys. See Section 5 for more details.

Second, condition (22) 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 completely 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. Indeed, this is the key channel through which superstar effects enter the test formula. Therefore, a crucial question that we address in the next section is how superstar effects affect these earnings elasticities and how they may compare with the elasticities typically estimated in the empirical literature.

Intuition. Condition (22) provides an *upper bound* for Pareto efficient marginal tax rates at w. The intuition can be grasped by imagining a reduction in taxes locally around a given earnings level 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 can be done by comparing the (positive) effect on revenue of those who initially earn less than w and thus 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 (22). Similar to a high earnings elasticity (in front of the brackets), this pushes for a lower 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 the second term in brackets is negative and large in absolute value, and 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. The same is true if income effects are large (so η is positive and large) for the reasons emphasized in Saez (2001).

Rawlsian Optimum. In fact, there is a tight connection between the efficiency test and the *optimal* tax schedule under a Rawlsian criterion, which effectively only puts positive Pareto weight on the lowest θ-type in (16). The optimum then sets the tax rate at the *maximal* level consistent with (22). This leads to the following corollary.

Corollary 1. *The optimal Rawlsian tax schedule satisfies*

$$\frac{T'(w)}{1-T'(w)} = \frac{1}{\varepsilon^c(w)wh(w)} \int_w^\infty \exp\left[\int_w^s \frac{\eta(t)}{\varepsilon^c(t)} \frac{dt}{t}\right] dH(s) \;\; \forall w.$$

This corresponds to the integral version of (22) evaluated as an equality everywhere.

The role of earnings elasticities and the earnings distribution. We illustrate the usefulness of Proposition 1 in the simpler case when there are no income effects on earnings, so ε^c and ε^u coincide, and the tax schedule is locally linear at w, with marginal tax rate τ. Hence, (22) reduces to

$$\frac{\tau}{1-\tau}\varepsilon(w)\left[\rho(w) - \frac{d\log\varepsilon(w)}{d\log w}\right] \leq 1. \tag{23}$$

The upper bound on τ now only depends on the shape of the equilibrium earnings distribution $h(w)$ (as captured by the local Pareto parameter $\rho(w)$) and the earnings elasticity $\varepsilon(w)$ (and its change) at w. The same holds, by Corollary 1, for the optimal Rawlsian tax schedule, which now reduces to

$$\frac{T'(w)}{1-T'(w)} = \left(\varepsilon(w)\frac{wh(w)}{1-H(w)}\right)^{-1} \quad \forall w, \tag{24}$$

the inverse of the earnings elasticity times a measure of the tail thickness of the earnings distribution.[16]

This motivates two questions that we will investigate in the following. First, suppose we have a tax T in place and observe the empirical earnings distribution in the resulting equilibrium. Given this earnings distribution, does the presence of superstar effects make it more or less likely that the efficiency test (23) is passed, i.e., does it push for lower or higher taxes? We see that this crucially depends on how superstars affect earnings elasticities, which we will explore in detail in the next Section 5. Second, for given fundamentals (in particular a given skill distribution and preferences), how do superstar effects change the set of Pareto efficient taxes? For instance, how do superstars affect revenue-maximizing marginal tax rates through their effect, in view of (24), on both the earnings distribution and earnings elasticities? We turn to this comparative static exercise in Section 6.

[16]In the special case where H is a Pareto distribution with parameter ρ: $\frac{wh(w)}{1-H(w)} = -\left(1+\frac{d\log h(w)}{d\log w}\right) = \rho$. This also is true for any distribution at the top: if either $\frac{wh(w)}{1-H(w)}$ or $\rho(w) = -\left(1+\frac{d\log h(w)}{d\log w}\right)$ converge as $w \to \infty$ then both measures must coincide.

5 Result 2: Superstars and Earnings Elasticities

In the previous section, we showed that the test for the efficiency of a given tax schedule only depends on the (directly observable) earnings distribution as well as the earnings elasticities of superstars under this tax. In this section, we explain how to apply the test in practice if available estimates of earnings elasticities in fact do not appropriately account for superstar phenomena. A closely related question we also address is how the efficiency test would change if we *believed* that the observed earnings distribution was generated by a standard economy without superstars.

5.1 Estimated versus Required Elasticities

To apply the efficiency test, we need the correct earnings elasticities for superstars, as defined in Section 4.1. 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 their effort at their current match. This is likely to be the case if the estimates are based on temporary and/or unexpected tax changes.

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 simply are not present. Or even if they are based on samples of high-income individuals, they could be from settings where superstar phenomena only play a limited role.

3. The estimates are *macro elasticities*, which measure earnings responses to tax changes including the general equilibrium effects on the entire earnings schedule $W(y)$. For instance, this is the case whenever identification comes from large tax reforms with aggregate effects. These elasticities are not the required ones, because they account for changes in the equilibrium earnings schedule in response to tax changes, instead of holding $W(y)$ fixed when measuring earnings responses.

In the next 3 subsections, we address these scenarios and show in each case how the estimated elasticities need to be adjusted in order to use them for the efficiency test.

5.2 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 follows:

$$w_0 = B(\theta_0, y_0) - \pi(\theta_0),$$

i.e. as the difference between output and profits at the firm type θ_0 is matched with.

Fixed-Assignment Earnings Schedules. To formalize the idea that the empirically 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

$$\hat{W}_0(y) = B(\theta_0, y) - \pi(\theta_0). \tag{25}$$

$\hat{W}_0$ 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),

$$\hat{W}_0'(y_0) = B_y(\theta_0, y_0) = W'(y_0),$$

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 θ_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), the interpretation is particularly simple: then $\hat{W}_0(y)$ is linear, and given by the line that is 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, $\hat{W}_0$ neatly 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.

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 (19)–(21) when replacing W by $\hat{W}_0$. The following result shows how they differ from the elasticity terms required for the

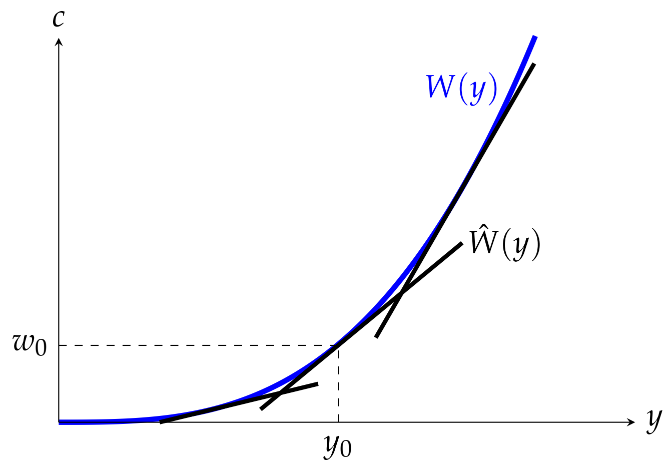

Figure 2: Equilibrium schedule $W(y)$ versus fixed-assignment schedule $\hat{W}_0(y)$.

efficiency test (22):

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 (25). Let y_0 be such that $w_0 = W(y_0) = \hat{W}(y_0)$ and $\theta_0 = \Gamma(y_0)$. Then we have*

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

$$\frac{\eta(w_0)}{\varepsilon^c(w_0)} = \frac{\hat{\eta}(w_0)}{\hat{\varepsilon}^c(w_0)},$$

where

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

Proposition 2, proved in Appendix B.4, 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 η/ε^c. As we illustrate further below, a lower elasticity provides a force for a *lower* upper bound on marginal tax rates. 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. Indeed, from the definition of the adjustment factor

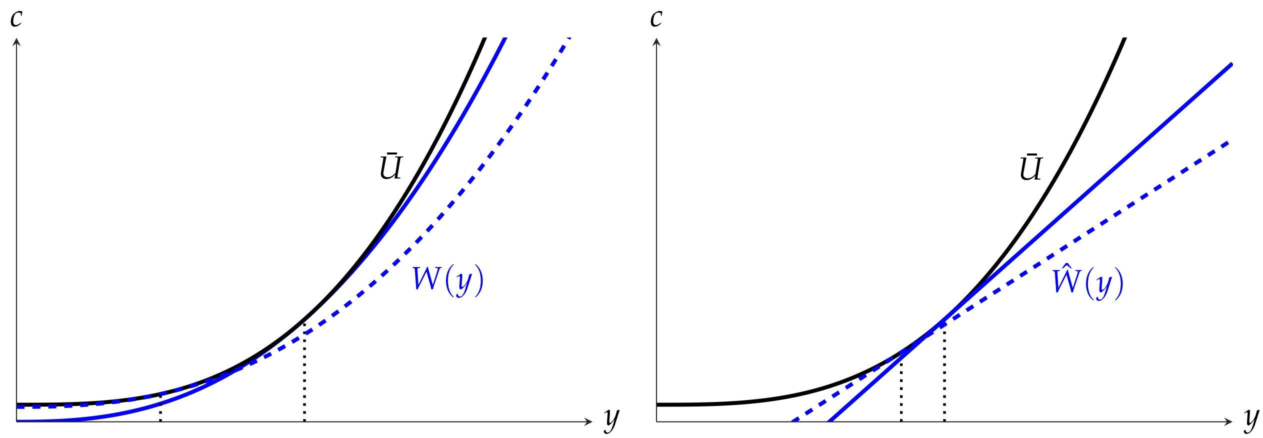

Figure 3: Compensated earnings elasticity with superstar effects (left) and without (right).

$\hat{\Phi}(w_0)$, the additional convexity is due to supermodularity of the output function $B_{y\theta} > 0$; thus, it is driven precisely by superstar effects.

This is particularly transparent when B is linear in y. Then by (6) the adjustment becomes

$$\hat{\Phi}(w_0) = 1 - \hat{\varepsilon}^c(w_0)\frac{\frac{W''(y_0)y_0}{W'(y_0)}}{\frac{W'(y_0)y_0}{W(y_0)}} \in [0,1). \tag{26}$$

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.

Gabaix-Landier Economy. We can illustrate the magnitude of this adjustment using the Gabaix-Landier parameterization from Section 2.3. By (9), the elasticities of the earnings schedule and the marginal earnings schedule are

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

Substituting this in the adjustment factor (26) gives[17]

$$\varepsilon^c(w) = \frac{\hat{\varepsilon}^c(w)}{1 - \beta\rho\hat{\varepsilon}^c(w)}. \tag{27}$$

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

[17]Note that the framework in Gabaix and Landier (2008) indeed features an output function $A(x,y)$ that is linear in y. Hence (26) applies.

As an illustration, 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 (27) provides potentially drastic 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 left panel of Figure 4 displays ε^c as a function of $\hat{\varepsilon}^c \in [0, 1/3)$. The adjustment becomes infinite as $\hat{\varepsilon}^c \to 1/3$.

Tax Implications of Superstar Earnings Elasticities. In general, the test condition in Proposition 1 not only depends on the *level* of the earnings elasticity, but also on how it *changes* along the earnings distribution, as captured by the term $d \log \varepsilon^c(w) / d \log w$ in (22). We therefore obtain particularly clean results in the important case where both marginal tax rates and elasticities are locally constant, such as in the top bracket of earnings:

Corollary 2. *Suppose that, as $w \to \infty$, $T'(w), \rho(w), \varepsilon^c(w)$ and $\eta(w)$ all converge asymptotically to the constants $\tau, \rho, \varepsilon^c$ and η, respectively. Suppose the same holds for the fixed-assignment elasticities $\hat{\varepsilon}^c(w) \to \hat{\varepsilon}^c$, $\hat{\eta}(w) \to \hat{\eta}$. Then (22) implies*

$$\frac{\tau}{1-\tau} \leq \frac{1}{\varepsilon^c \left(\rho - \eta/\varepsilon^c\right)} \leq \frac{1}{\hat{\varepsilon}^c \left(\rho - \hat{\eta}/\hat{\varepsilon}^c\right)}.$$

Using the non-superstar elasticities delivers a *less stringent* upper bound to top marginal tax rates. Conversely, realizing that the earnings distribution is in fact the result of superstar effects pushes for a *more stringent, lower* upper bound to the top marginal tax rate. Alternatively, this provides a comparison of the predictions for optimal taxation of a superstar economy relative to a standard Mirrlees economy.

For the case of no income effects and using the Gabaix-Landier parameterization, the upper bound to the top marginal tax rate can be written as

$$\tau \leq \frac{1}{1+\rho\varepsilon} = \frac{1}{1+\rho\frac{\hat{\varepsilon}}{1-\beta\rho\hat{\varepsilon}}}.$$

For instance, based on the Gabaix-Landier numbers $\beta = 1$ and $\rho = 3$, if $\hat{\varepsilon} = 1/4$, then $\tau \leq 25\%$. Erroneously using the unadjusted $\hat{\varepsilon}$ instead of ε, one would have concluded that $\tau \leq 57\%$. The right panel in Figure 4 compares the correct upper bound to the unadjusted one as a function of the elasticity $\hat{\varepsilon} \in [0, 1/3)$.

5.3 Effort Elasticities

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

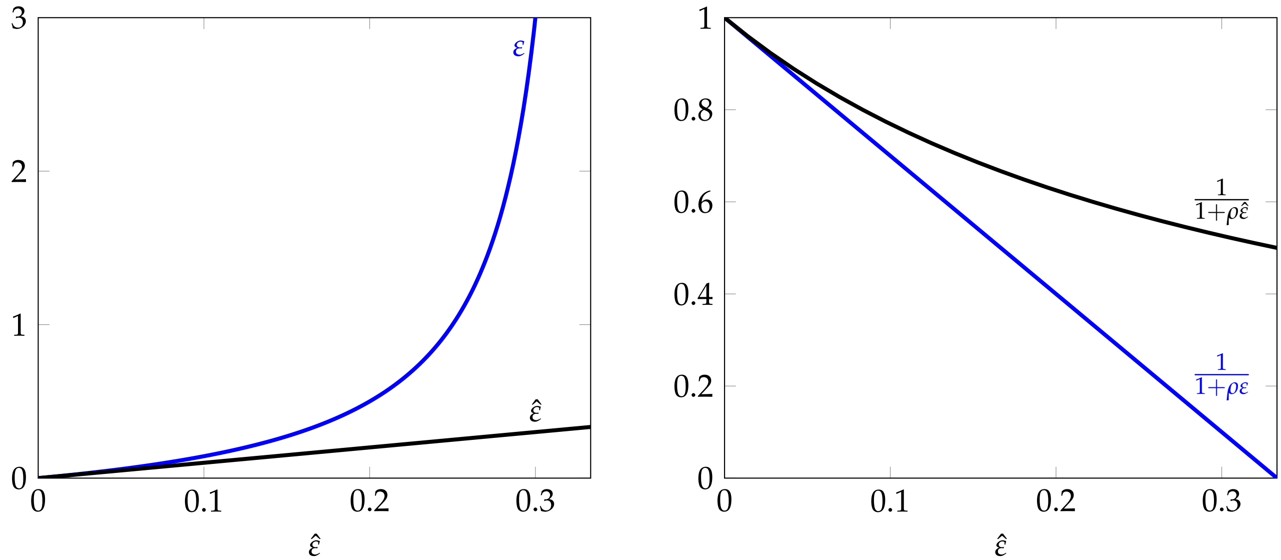

Figure 4: Unadjusted and adjusted earnings elasticities (left panel) and upper bound on top tax rate using unadjusted and adjusted earnings elasticities (right panel).

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

$$\tilde{\varepsilon}^u(y) = \frac{\partial y}{\partial(1-\tau)} \frac{1-\tau}{y}, \tag{28}$$

$$\tilde{\eta}(y) = -(1-\tau)\frac{\partial y}{\partial I} \tag{29}$$

and

$$\tilde{\varepsilon}^c(y) = \varepsilon^u(y) + \eta(y), \tag{30}$$

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 4.1. For instance, under the quasilinear and iso-elastic preference specification in Example A, the uncompensated and compensated effort elasticities coincide and are given by the constant structural parameter $\tilde{\varepsilon} = 1/(\gamma - 1)$.[18]

[18]This could be viewed as corresponding to hours elasticities (see e.g. MaCurdy, 1981; Eissa and Hoynes,

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 B.5. This property will be useful in the following, when characterizing the relationship between effort and earnings elasticities. In particular, whenever $W(y) \to \infty$ as $y \to b$ for some bound $b < \infty$ (as in the Gabaix-Landier economy for $\beta\rho > 1$), then the right-hand side of the above inequality goes to 0, which requires that $\tilde{\varepsilon}^c$ vanishes at the top. This will indeed be the case for the preferences in Example B. 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 these standard 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, let us assume here (i) that A is linear in y: $A(x, y) = a(x)y$ (so the non-superstar counterfactual indeed involves linear production) and (ii) that 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). Both assumptions are satisfied in our examples. This leads to the following result, proved in Appendix B.6:

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*

1998; Blundell et al., 1998; and Ziliak and Kniesner, 1999), although of course our notion of effective effort is broader, or 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).

defined as in (30) and (29), respectively. Under conditions (i) and (ii), we have

$$\varepsilon^c(w_0) = \frac{\tilde{\varepsilon}^c(y_0)}{\tilde{\Phi}(y_0)} > \tilde{\varepsilon}^c(y_0) \tag{31}$$

$$\frac{\eta(w_0)}{\varepsilon^c(w_0)} = \frac{\tilde{\eta}(y_0)}{\tilde{\varepsilon}^c(y_0)} \frac{1}{\frac{W'(y_0)y_0}{W(y_0)}} < \frac{\tilde{\eta}(y_0)}{\tilde{\varepsilon}^c(y_0)}, \tag{32}$$

where

$$\tilde{\Phi}(y_0) = \frac{1}{\frac{W'(y_0)y_0}{W(y_0)}} - \tilde{\varepsilon}^c(y_0) \frac{\frac{W''(y_0)y_0}{W'(y_0)}}{\frac{W'(y_0)y_0}{W(y_0)}} \in [0, 1). \tag{33}$$

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 (32) and (33). 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 and Corollary 2, both of these new adjustments push for *lower* taxes for a given earnings distribution.

Gabaix-Landier Economy. The required adjustments (31) and (33) simplify under the Gabaix-Landier specification. In particular, using the effort distribution (8) and the earnings schedule (9) we obtain, at any point in the distribution,

$$\varepsilon^c = \frac{\tilde{\varepsilon}^c}{(\beta\rho - 1)\frac{b - y_0}{y_0} - \beta\rho\tilde{\varepsilon}^c} = \frac{\tilde{\varepsilon}^c}{\frac{\beta\rho - 1}{\beta\rho n^{\frac{1}{\rho} - \beta} - 1} - \beta\rho\tilde{\varepsilon}^c}. \tag{34}$$

The first expression is written in terms of effort, the second in terms of rank (and we omitted the dependence of ε^c and $\tilde{\varepsilon}^c$ on y or n to simplify notation); both will be useful. Note that for the bottom ($n = 1$) the effort elasticity adjustment is the same as the fixed-assignment elasticity adjustment from the preceding section, given by (27); 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 $n \to 0$,

$$\varepsilon^c = \frac{\tilde{\varepsilon}^c}{1 - \beta\rho - \beta\rho\tilde{\varepsilon}^c} > \tilde{\varepsilon}^c, \tag{35}$$

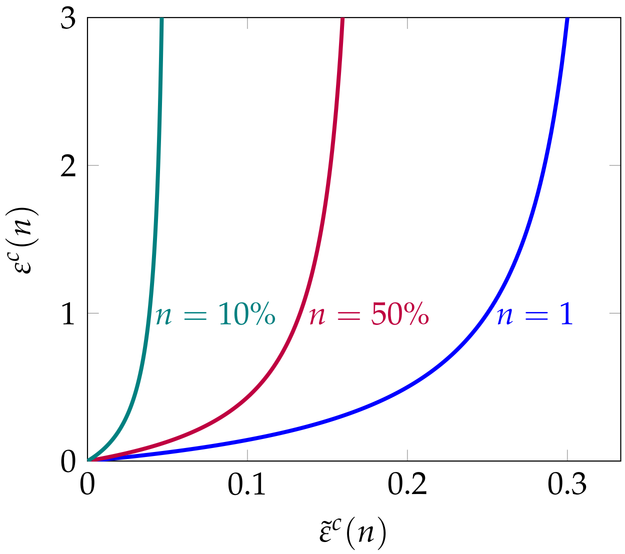

Figure 5: Adjusted earnings elasticities for effort elasticities at bottom, median and top 10%.

since, by Lemma 2, $\tilde{\varepsilon}^c \leq \frac{1-\beta\rho}{\beta\rho}$, ensuring that the denominator is positive. We see that the adjustment is increasing in the superstar parameter β and vanishes if $\beta = 0$. For a given observed earnings distribution and hence Pareto parameter ρ, this implies a lower upper bound on the set of Pareto efficient top marginal tax rates τ. For instance, with no income effects (as in Example A), we have

$$\frac{\tau}{1-\tau}\varepsilon\rho \leq 1, \tag{36}$$

which should be compared to erroneously using the effort elasticity $\tilde{\varepsilon} < \varepsilon$.

If $\beta\rho > 1$, as supported by Gabaix-Landier, then, since $\beta\rho n^{\frac{1}{\rho}-\beta} \to \infty$ as $n \to 0$, the adjustment becomes unbounded for any given $\tilde{\varepsilon}^c$. Erroneously using such an unbounded earnings elasticity in (36) would appear to imply that only a zero top tax rate can be Pareto efficient. Figure 5 illustrates this. It shows the relationship between the earnings elasticity and the effort elasticity (34) 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 $n \to 0$ 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 that property and

$$\tilde{\varepsilon}(y) = \frac{1}{\gamma+1}\frac{b-y}{y}.$$

Naively using this effort elasticity in the efficiency test (36), we would erroneously conclude that the upper bound to the set of Pareto efficient top marginal tax rates is 100%. However, substituting in (34) reveals that the earnings elasticity is in fact constant and equals[19]

$$\varepsilon = \frac{1}{\gamma(\beta\rho-1)-1} > 0. \tag{37}$$

The efficiency test for the top marginal tax rate and a given observed earnings distribution becomes

$$\frac{\tau}{1-\tau}\frac{\rho}{\gamma(\beta\rho-1)-1} \leq 1, \tag{38}$$

implying a well-defined upper bound on τ strictly less than 1.[20]

5.4 Macro Elasticities

The earnings elasticities required for Proposition 1 measure *partial equilibrium* behavioral responses, in the sense that they hold the equilibrium earnings schedule fixed when varying taxes. The third scenario is therefore one where the estimated elasticities are *macro elasticities*, which capture the general equilibrium effects of tax changes. This would be the case when elasticities are estimated using large tax reforms with aggregate effects or using cross-country comparisons.[21] These estimates will capture the fact that the entire equilibrium earnings schedule $W(y)$ shifts 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 τ 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 4.1.

[19] $\gamma(\beta\rho-1) \geq 1$ by Lemma 2.

[20] Using the Gabaix-Landier parameters $\beta = 1$ and $\rho = 3$, this is purely a function of the preference parameter γ. Depending on γ, any upper bound on the top tax rate in $(0,1)$ can be justified, and we can always adjust the underlying skill distribution to make it consistent with the numbers in Gabaix-Landier.

[21] 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).

Macro Earnings Schedules. We begin with two key observations: First, for any value of τ, the equilibrium assignment will be the same; i.e., type θ_0 will stay matched with the same firm. The macro elasticity will therefore be akin to the fixed-assignment elasticity from Section 5.2, where earnings follow the equilibrium output function $B(\theta_0|y)$. However, the difference is that the macro elasticity will also incorporate the *shift* in the earnings schedule, so even when individual θ_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 θ_0 is matched with in response to the change in τ.

As a result, the macro earnings response at w_0 is determined according to the macro earnings schedule

$$\overline{W}_0(y|\tau) = B(\theta_0, y) - \pi(\theta_0|\tau). \tag{39}$$

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 θ_0 chooses her earnings as follows:

$$\overline{w}_0(\tau) \in \arg\max_w (1-\tau)w - T(w) - \phi(\overline{W}_0^{-1}(w|\tau), \theta_0) \tag{40}$$

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

$$\overline{\varepsilon}(w_0) = \left.\frac{d\overline{w}_0}{d(1-\tau)}\right|_{\tau=0} \frac{1-T'(w_0)}{w_0}. \tag{41}$$

Adjusting Macro Elasticities. With these definitions, we have the following result (see Appendix B.7 for a proof):

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 $\overline{\varepsilon}(w_0)$ be as defined in (41), the fixed-assignment elasticity $\hat{\varepsilon}(w_0)$ as in Section 5.2, and the effort elasticity $\tilde{\varepsilon}(y_0)$ as in (28). Under conditions (i) and (ii), we have*

$$\hat{\varepsilon}(w_0) = \frac{\overline{\varepsilon}(w_0)}{\overline{\Phi}(w_0)} \tag{42}$$

where

$$\overline{\Phi}(w_0) = 1 + \frac{\chi(w_0)}{\tilde{\varepsilon}(y_0)} \frac{1}{\frac{W'(y_0)y_0}{W(y_0)}} \quad \textit{and} \quad \chi(w_0) = \left.\frac{\partial \overline{W}_0(y_0|\tau)}{\partial(1-\tau)}\right|_{\tau=0} \frac{1-T'(w_0)}{w_0}. \tag{43}$$

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 χ 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.

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.

Examples. Proposition 4 can be nicely illustrated using our examples. For the isoelastic preferences in Example A with a Pareto skill distribution, Appendix A shows that the equilibrium earnings schedule is $W(y) = ky^{\frac{1}{1-\beta\rho}} + w_0$ with $k = \tilde{C}(1-\tau)^{-\frac{\beta\rho}{(\gamma-1)(1-\beta\rho)}}$. This immediately implies that, at the top, the general equilibrium correction in (43) converges to

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

which is negative as predicted by the second part of Proposition 4. We also know that the effort elasticity is $\tilde{\varepsilon} = 1/(\gamma-1)$ and the earnings elasticity is $\varepsilon = 1/(\gamma(1-\beta\rho)-1)$ at the top by (35). Finally, by (27), the fixed-assignment elasticity at the top is

$$\hat{\varepsilon} = \frac{1}{1/\varepsilon + \beta\rho} = \frac{1}{(\gamma-1)(1-\beta\rho)} < \varepsilon.$$

Taking this together allows us to derive the macro elasticity at the top using (42) and (43):

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

The macro elasticity at the top coincides with the effort elasticity. We can therefore summarize the relationships between the four elasticity concepts in this case as follows:

$$\tilde{\varepsilon} = \bar{\varepsilon} < \hat{\varepsilon} < \varepsilon.$$

In all three scenarios, the estimated elasticity is smaller than the required superstar earnings elasticity ε, with the macro and effort elasticities $\bar{\varepsilon}$ and $\tilde{\varepsilon}$ being even smaller than the fixed-assignment elasticity $\hat{\varepsilon}$.

Similar results can be shown to apply for Example B. Indeed, at the top one finds that $\bar{\varepsilon} < \tilde{\varepsilon} = 0 < \hat{\varepsilon} < \varepsilon$.

5.5 Replicating the Superstar Equilibrium without Superstars

We finally address the second question we raised at the beginning of this section: How would we have tested for Pareto efficiency if we had *believed* that the observed earnings distribution came from a standard economy without superstars, holding fixed the tax T in place? Clearly, posing this question assumes a positive answer to the following: Are there economies that replicate the same outcome as in the superstar economy, but without superstar effects? We shall show this is indeed the case. This amounts to a *recalibration exercise:* We pick preferences $\hat{U}(c, w, \theta)$ and a skill distribution $F(\theta)$ for a standard Mirrlees economy with linear technology where earnings equal effort, $w = y$, such that, under the earnings tax T, we get the same earnings distribution $H(w)$ as in the superstar economy. We already know that the test *formula* in Proposition 1 equally applies to this recalibrated non-superstar economy, so the only remaining question is which elasticity corresponds to the required earnings elasticities for the recalibrated economy.

Recalibrating Preferences. We show first that one can construct $\hat{U}$ such that the correct elasticities coincide precisely with the *fixed-assignment elasticities* $\hat{\varepsilon}^u(w)$, $\hat{\eta}(w)$ and $\hat{\varepsilon}^c(w)$ from Section (5.2), for example. Fix an earnings tax T and the resulting superstar equilibrium, with effort $y(\theta)$, earnings $w(\theta) = W(y(\theta))$, output $B(\theta, y(\theta))$ and profits

$$\pi(\theta) \equiv B(\theta, y(\theta)) - w(\theta).$$

Denote by $B^{-1}(\theta, .)$ the inverse function of $B(\theta, .)$ with respect to its second argument and let

$$\Delta(\theta, w) \equiv B^{-1}(\theta, w + \pi(\theta)).$$

Using this, define the new utility function

$$\hat{U}(c, w, \theta) \equiv U(c, \Delta(\theta, w), \theta).$$

Intuitively, $\hat{U}$ incorporates the firm level production function B into preferences, holding again fixed the equilibrium matching and accounting for the fact that individuals do not earn the entire output because there are profits in the superstar equilibrium. Consider a standard Mirrlees model with linear technology (so output, effort and earnings are all the same), preferences $\hat{U}$, and endowment $E + \Pi$ with

$$\Pi = \int \pi(\theta) dF(\theta).$$

Leave the skill distribution F unchanged. Then we have the following Corollary to Proposition 2 (see Appendix B.8 for a proof):

Corollary 3. *In the non-superstar economy with preferences $\hat{U}$ and earnings tax T, each type θ chooses the same earnings and consumption as in the superstar economy with preferences U and tax T. The test for Pareto efficiency of T is given by (22) when replacing $\varepsilon^c(w)$ and $\eta(w)$ by $\hat{\varepsilon}^c(w)$ and $\hat{\eta}(w)$.*

In other words, if we believed that the observed earnings distribution was generated by a standard economy with no superstars, we would evaluate the efficiency of the existing tax schedule T based on the same condition as in the superstar economy, but using the *lower* compensated earnings elasticity $\hat{\varepsilon}^c(w)$.

Recalibrating the Skill Distribution. An alternative is to vary the underlying *skill distribution* in the non-superstar economy to match the empirical distribution of earnings in the superstar economy, instead of recalibrating preferences. We use Example A to show that this leads to similar results.

In Example A, both the skill and the equilibrium earnings distribution are (approximately) Pareto, the former with parameter α and the latter with parameter ρ given by (10). Now construct a standard economy with linear production and the *same* preferences (described by parameter γ), but with another skill distribution (in fact, a Pareto distribution with parameter $\tilde{\alpha}$), such that, under the same tax τ, we replicate the *same* earnings distribution as in the superstar economy. In the standard economy, the Pareto parameter of the earnings distribution will be

$$\tilde{\rho} = \frac{\gamma - 1}{\gamma} \tilde{\alpha},$$

so we need to set $\tilde{\alpha}$ such that $\tilde{\rho} = \rho$:

$$\tilde{\alpha} = \frac{\alpha}{\alpha\beta + 1} < \alpha.$$

This is intuitive: To generate the same earnings distribution, the underlying *skill* inequality in the non-superstar economy must be *higher* than in the superstar economy. However, since earnings and effort coincide in the standard economy, the *earnings elasticity* is simply $1/(\gamma - 1)$ and hence *less* than the earnings elasticity in the superstar economy, as we saw in Section 5.3. In other words, accounting for superstars again requires the use of a *higher* earnings elasticity than in the non-superstars counterfactual, holding fixed the observed earnings distribution.

6 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 $F(\theta)$. In other words, we ask whether, for a given skill distribution and preferences, the introduction of superstar effects leads to higher or lower taxes. We also show which statistics, such as elasticities or the parameters of the superstar technology, determine these effects.

6.1 An Efficiency Test in Terms of Primitives

As in Section 3, we begin with a test that any Pareto efficient tax schedule has to satisfy, now however expressed in terms of fundamentals.

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

$$\frac{\tau(\theta)}{1-\tau(\theta)}\xi(\theta)\left(p(\theta) - \frac{d\log\left(\frac{\tau(\theta)}{1-\tau(\theta)}\xi(\theta)\right)}{d\log\theta} - \frac{\tilde{\eta}(\theta)}{\tilde{\varepsilon}^c(\theta)}\frac{y'(\theta)\theta}{y(\theta)}\right) \leq 1 \tag{44}$$

for all θ, where

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

is the local Pareto parameter of the skill distribution and

$$\xi(\theta) \equiv \left(-\frac{\partial\log MRS(c(\theta), y(\theta), \theta)}{\partial\log\theta}\right)^{-1} \geq 0$$

is the inverse of the elasticity of the marginal rate of substitution w.r.t. θ*.*

See Appendix B.9 for the proof. Note that $\xi(\theta)$ is a preference parameter. For example, in the canonical specification $U(c,y,\theta) = u(c,y/\theta)$, the elasticity of the marginal rate of substitution w.r.t. θ is directly related to the elasticity of the marginal rate of substitution w.r.t. y, and we show in Appendix B.9 that

$$\xi(\theta) = \frac{\tilde{\varepsilon}^c(\theta)}{1+\tilde{\varepsilon}^u(\theta)},$$

leading to the following immediate corollary:

Corollary 4. *If* $U(c,y,\theta) = u(c,y/\theta)$*, any Pareto efficient earnings tax* T *is such that for all* θ

$$\frac{\tau(\theta)}{1-\tau(\theta)}\frac{\tilde{\varepsilon}^c(\theta)}{1+\tilde{\varepsilon}^u(\theta)}\left[p(\theta) - \frac{d\log\left(\frac{\tau(\theta)}{1-\tau(\theta)}\frac{\tilde{\varepsilon}^c(\theta)}{1+\tilde{\varepsilon}^u(\theta)}\right)}{d\log\theta} - \frac{\tilde{\eta}(\theta)}{\tilde{\varepsilon}^c(\theta)}\frac{y'(\theta)\theta}{y(\theta)}\right] \leq 1. \tag{45}$$

These results reveal three insights. First, the *formula* for the efficiency test (44) (and its special case (45)) 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 3, where *earnings* elasticities were the key sufficient statistic when using the earnings distribution as an input. Third, the test inequality (45) 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 natural benchmark cases where not even this matters.

Examples. Suppose preferences are quasilinear, so that $\tilde{\eta}(\theta) = 0$. Then (45) simplifies to

$$\frac{\tau(\theta)}{1-\tau(\theta)}\frac{\tilde{\varepsilon}^c(\theta)}{1+\tilde{\varepsilon}^u(\theta)}\left[p(\theta) - \frac{d\log\left(\frac{\tau(\theta)}{1-\tau(\theta)}\right)}{d\log\theta}\right] \leq 1 \quad \forall\theta, \tag{46}$$

which is only a function of the tax schedule, the skill distribution, and the local elasticity. If the tax schedule is locally linear or at the top, the efficiency test is simply

$$\frac{\tau}{1-\tau}\frac{\tilde{\varepsilon}^c(\theta)}{1+\tilde{\varepsilon}^u(\theta)}p(\theta) \leq 1.$$

The condition puts an upper bound on the marginal tax rate that is decreasing in the *effort* elasticity and the Pareto parameter. This condition is independent of the equilibrium earnings schedule $W(y)$ and the allocation $c(\theta), y(\theta)$, except for the possible dependence of the local elasticity on where it is evaluated.

In Example A we have $\frac{\tilde{\varepsilon}^c(\theta)}{1+\tilde{\varepsilon}^u(\theta)} = \frac{1}{\gamma}$ and $p(\theta) = \alpha$, so that

$$\frac{\tau}{1-\tau} \leq \frac{\gamma}{\alpha}. \tag{47}$$

As it turns out, this same condition obtains in Example B (see below). This shows particularly clearly, and perhaps surprisingly, that superstars are completely *neutral*: superstar effects do *not* affect the set of Pareto efficient marginal taxes at the top, for given fundamentals. The shape of the output function A and in particular parameter β, plays no role whatsoever in these conditions.

To understand this conclusion, we go back to the efficiency condition in condition (36) (which corresponds to condition (23) at the top, or the condition in Corollary 2 in the absence of income effects), expressed in terms of the Pareto parameter for earnings ρ and the earnings elasticity ε. In Example A, we obtain

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

while in Example B

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

Superstar effects change both ρ and ε. On the one hand, the Pareto parameter ρ of the *earnings* distribution is *decreasing* in β, i.e. superstar effects lead to fatter tails in the earnings distribution; on the other hand, the earnings elasticity is increasing in β, given the primitive parameters α and γ.[22] Neutrality obtains because both these changes precisely cancel out: in condition (36), the term $\varepsilon\rho = \alpha/\gamma$ is independent of β.

To sum up, 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.

[22]These observations on the elasticity are not implied by the results reported in Section 5. Unlike the results obtained there, here we hold the skill distribution and preferences fixed, α and γ, while changing a technology parameter, β.

6.2 Optimal Tax Schedules in Terms of Primitives

As in Section 3, we can exploit the tight relationship between testing for Pareto efficient taxes and *optimal* tax schedules for any given Pareto weights, this time however in terms fundamentals. This will allow 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. A natural benchmark is again the Rawlsian case where our goal is to maximize revenue from the top tax bracket.

The following result, proved in Appendix B.10, provides a formula for optimal taxes:

Proposition 6. *The optimal tax schedule for any distribution of Pareto weights Λ satisfies*

$$\frac{\tau(\theta)}{1-\tau(\theta)} = \frac{1}{\theta f(\theta)\xi(\theta)} \int_\theta^\infty \left(1 - \frac{\lambda(s)U_c(s)}{\eta f(s)}\right) \exp\left(\int_\theta^s \frac{\tilde{\eta}(t)}{\tilde{\varepsilon}^c(t)} \frac{dy(t)}{y(t)}\right) dF(s), \tag{48}$$

where η is the multiplier on (17) and λ is the density corresponding to Λ.

When $U(c,y,\theta) = u(c,y/\theta)$, we have $\xi = \tilde{\varepsilon}^c/(1+\tilde{\varepsilon}^u)$ as before, so (48) 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. In a linear tax framework, Diamond and Mirrlees (1971) 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 an implication of their result.

Examples. For instance, with the preferences from Example A, the optimal tax schedule in terms of θ is given by

$$\frac{\tau(\theta)}{1-\tau(\theta)} = \gamma \frac{\Lambda(\theta) - F(\theta)}{\theta f(\theta)},$$

independent of superstar effects. Similarly, with the preferences from Example B, it is given by

$$\frac{\tau(\theta)}{1-\tau(\theta)} = \gamma \frac{\Lambda(\theta) - F(\theta)}{(a-\theta) f(\theta)},$$

again a closed form function of preference parameters and the skill distribution only. As a result, even though the emergence of superstar technology may lead to a dramatic increase in earnings inequality for a given underlying skill distribution, it does not affect the optimal taxes for any given Pareto weights.[23]

In the Rawlsian case where positive social welfare weight is put only on the lowest skill type, we have $\Lambda(\theta) = 1$ for all θ, so formula (48) simplifies accordingly. With the preferences and skill distributions assumed in Examples A and B, for instance, the optimal Rawlsian marginal tax rates are therefore constant and given by

$$\frac{\tau}{1-\tau} = \frac{\gamma}{\alpha},$$

so they are simply the highest Pareto efficient taxes, as given by the upper bound (47).[24] Hence, the revenue-maximizing top marginal tax rate is also unaffected by superstar effects.

7 Extensions

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 alternative and more general settings with superstar effects, including the first- and second-generation span-of-control models we turn to next.

7.1 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

$$A(L, y),$$

[23] These results imply that a given individual of type θ faces the same optimal marginal tax rate, no matter whether technology features superstar effects or not. This is the standard way of reporting the predictions of optimal nonlinear taxes in the literature. Since the distribution of θ is fixed and since earnings are monotone in θ, it is equivalent to reporting marginal taxes as a function of the quantiles of the earnings distribution, a useful normalization (i.e. it answers "what is the tax on the top x%?"). In general, this does *not* mean that the tax schedule when written in terms of earnings, $T(w)$, is unchanged. The reason, of course, is that the distribution of earnings changes when superstar effects emerge.

[24] More generally, formula (48) for the Rawlsian case is the integral version of (44) when it holds as an equality everywhere.

and we again assume that A is supermodular, so $A_{Ly} > 0$. Denoting the wage per unit of labor by ω, 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 ω. Managers hire labor to maximize profits taking the wage as given, so their earnings are

$$W(y|\omega) = \max_{L} A(L,y) - \omega L \tag{49}$$

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

Superstar Effects. Here, superstar effects arise simply because more talented managers, who exert higher effort y, will hire more workers, thereby increasing their scale.[25] 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 in their effort. For example, if $A(L,y) = L^{\alpha}y$ with $\alpha \in (0,1)$, then

$$L(y|\omega) = y^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{\omega}\right)^{\frac{1}{1-\alpha}} \quad \text{and} \quad W(y|\omega) = (1-\alpha) y^{\frac{1}{1-\alpha}} \left(\frac{\alpha}{\omega}\right)^{\frac{\alpha}{1-\alpha}},$$

which are both increasing and convex in y.

Ales et al. (2015) consider a similar model, but with the important difference that output is given by $\theta^{\gamma} A(L,y)$ and assuming that A exhibits constant returns to scale. In that case, 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 γ, introduces Stiglitz (1982)-type effects, which we have abstracted from here.

Pareto Problem. For the purposes of this section, it will be convenient to index managers directly by the quantiles of the skill distribution $t = F(\theta)$, which is without loss of generality. Then managers' preferences as a function of t are simply

$$U(c, y, F^{-1}(t)).$$

[25]Similarly, in Rosen (1981), individuals can directly increase the scale of their production.

By the same arguments as in Section 3.1, the Pareto problem for the nonlinear profit tax on managers and the lump-sum tax on workers is

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

s.t.

$$U\left(c(t), y(t), F^{-1}(t)\right) \geq U\left(c(t'), y(t'), F^{-1}(t)\right) \quad \forall\, t, t' \tag{50}$$

$$\int c(t)dt + c^w \leq \int A(L(y(t)|\omega), y(t))dt \tag{51}$$

$$\int L(y(t)|\omega)dt = 1, \tag{52}$$

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, λ^w and $\lambda(t)$ are the respective Pareto weights, and the last equation (52) is the labor market clearing condition. Observe that the wage ω only appears in the last two constraints (51) and (52). This allows us to rewrite the Pareto problem as follows:

$$\max_{c(t),y(t),c^w} \lambda^w u(c^w) + \int \lambda(t) U\left(c(t), y(t), F^{-1}(t)\right) dt \tag{53}$$

s.t. (50) and

$$\int c(t)dt + c^w \leq \Psi(\{y\}), \tag{54}$$

where

$$\Psi(\{y\}) \equiv \max_{\omega} \int A(L(y(t)|\omega), y(t))dt \quad \text{s.t.} \quad \int L(y(t)|\omega)dt = 1. \tag{55}$$

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 (53) subject to (50) and (54) is the same as our original Pareto problem (16) subject to (13) and (17) from Section 3.1 when we replace $\int B(t, y(t))dt$ by $\Psi(\{y\})$. Since all our results are based on the necessary first-order conditions of the planning problem, they all go through if it remains true that the partial derivative of this aggregate output w.r.t. $y(t)$ coincides with the marginal earnings in equilibrium at that effort level. This relationship is indeed established by the following lemma, proved in Appendix B.11:

Lemma 3. *Let $\Psi(\{y\})$ be defined as in (55). Then*

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

for any ω and t.

Because of Lemma 3, all of our results in Propositions 1 to 6 immediately extend to the earnings tax for managers in this model.

7.2 Span-of-Control with Heterogeneous Workers

In this section, we 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).[26] Here, 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 an 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).[27] 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

[26] See also Garicano and Rossi-Hansberg (2015) for a recent survey of this literature.

[27] Extending our results to larger numbers of 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

$$1 = hn(1 - y_p) \quad \Rightarrow \quad n = \frac{1}{h(1 - y_p)}. \tag{56}$$

As a result of this time constraint, a manager can increase his team size only by being matched more talented workers, who provide more effort and can therefore solve more problems on their own.[28]

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 (56). 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

$$\Pi = n(A(y_m, y_p) - w_p(y_p)) - W_m(y_m) = \frac{y_m - w_p(y_p)}{h(1 - y_p)} - W_m(y_m) \tag{57}$$

and firms choose y_m and y_p to maximize Π 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

[28] By single-crossing preferences, the effort schedule y will again be increasing in talent t in any 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 any equilibrium.

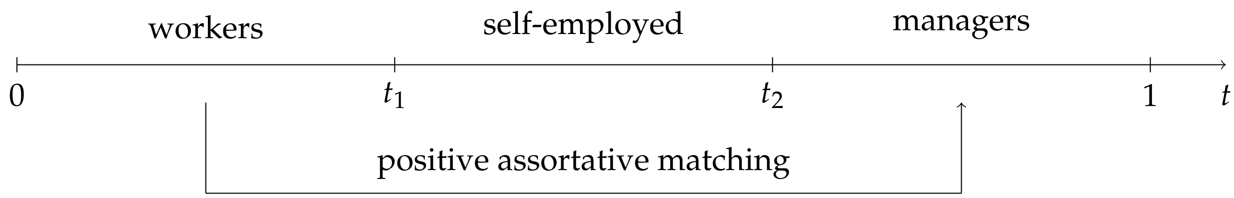

Figure 6: Equilibrium sorting and matching.

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 workers, and (iv) there is strictly positive assortative matching between workers in $[0, t_1]$ and managers in $[t_2, 1]$. Figure 6 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 on problems without help from managers. We assume there is risk pooling among these self-employed, so their earnings are simply equal to their expected output and hence effective effort: $W(y(t)) = y(t)$ for $t \in (t_1, t_2)$. Note that this set of self-employed may not exist depending on parameters: we may have $t_1 = t_2$ in equilibrium.

Matching Managers and Workers. By property (iv), 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

$$h \int_{P(t)}^{t_1} (1 - y(s))ds = \int_t^1 ds \quad \forall t \leq 1. \tag{58}$$

This constraint simply 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

$$P'(t) = \frac{1}{h(1 - y(P(t)))} \quad \forall t \leq 1, \tag{59}$$

with initial condition $P(1) = t_1$, where (59) follows from differentiating (58) w.r.t. t. Observe that this then also uniquely pins down the manager cutoff t_2, namely such that $P(t_2) = 0$.

Interestingly, and in contrast to the models we have considered so far, (59) implies that 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 of type t 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: For this reason, $P'(t)$ is increasing in $y(P(t))$.

Superstar Effects. This also reveals how superstar effects are different in this framework compared to the previous subsection: Better managers here 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.

Indeed, we can verify that the equilibrium earnings schedule is convex. The firm's first-order condition for y_m in (57) implies

$$W'_m(y(t)) = \frac{1}{h(1-y(P(t)))}, \tag{60}$$

which is increasing along the y-schedule among managers. Similarly, 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

$$w'_p(y(P(t))) = W_m(y(t))h, \tag{61}$$

which is also increasing along the y-schedule. Of course, the earnings schedule is linear in the intermediate segment between $y(t_1)$ and $y(t_2)$, if it exists, leading to an overall earnings schedule $W(y)$ that is weakly convex.[29]

[29]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 (60) and (61) 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.

Pareto Problem. By the same arguments as in Section 3.1, we are now ready to write down the Pareto problem for the earnings tax T:

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

s.t.

$$\int_0^1 c(t)dt \leq \int_{t_2}^1 \frac{y(t)}{h(1-y(P(t)))}dt + \int_{t_1}^{t_2} y(t)dt, \tag{62}$$

(50), (59) and $P(1) = t_1$, $P(t_2) = 0$. The first integral on the right-hand side of (62) 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 (59) and (62), which allows us to decompose the Pareto problem as follows:

$$\max_{c,y} \int_0^1 \lambda(t)U(c(t),y(t),F^{-1}(t))dt \tag{63}$$

s.t. (50) and

$$\int_0^1 c(t)dt \leq \Psi(\{y\}), \tag{64}$$

where

$$\Psi(\{y\}) = \max_{P,t_1,t_2} \int_{t_2}^1 \frac{y(t)}{h(1-y(P(t)))}dt + \int_{t_1}^{t_2} y(t)dt \tag{65}$$

s.t. (59) and $P(1) = t_1$, $P(t_2) = 0$. Hence, similar to the previous subsection, $\Psi(\{y\})$ is aggregate output in the economy under the equilibrium assignment for any given (monotone) effort schedule y.[30] The Pareto problem (63) s.t. (50) and (64) is therefore again identical to our original Pareto problem (16) subject to (13) and (17) from Section 3.1 when we replace $\int B(t,y(t))dt$ by $\Psi(\{y\})$. Since all our results are based on the necessary first-order conditions of the planning problem, they all go through if it remains true that the partial derivative of this aggregate output w.r.t. $y(t)$ coincides with the marginal earnings in equilibrium at that effort level. The following lemma again establishes precisely this relationship:

Lemma 4. *Let $\Psi(\{y\})$ be defined as in (65). Then*

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

for all t.

[30] By single-crossing and (50), any incentive compatible effort schedule y must be monotone, so we will evaluate Ψ only for monotone schedules.

The proof is in Appendix B.12. Despite the additional complexities of this model, in particular the richer effects of taxes on equilibrium matching, this implies again that our results extend to the earnings tax for both workers and managers in this model.

7.3 A General Production Function Approach

Both of our previous extensions concluded by showing that, in equilibrium, there exists an aggregate production function $\Psi(\{y\})$ that maximizes total output, given effort choices $\{y\}$, as well as the property that marginal earnings $W'(y(t))$ equal the derivative of Ψ 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 now 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 $y : [0,1] \to Y$ 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

$$W'(y(t)) = \frac{\partial \Psi(\{y\})}{\partial y(t)}, \tag{66}$$

the desired marginal condition.[31] Our results then follow directly for this more general formulation.

[31]More formally, by $\frac{\partial}{\partial y}\Psi(\{y\})$ we mean the following. The Fréchet derivative of Ψ at some $\{y\}$ in direction h is a linear function

$$\partial\Psi(\{y\}, h) = \int \delta(t)dh(t).$$

It is worth pointing out that this formulation 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 (66) 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.

8 Conclusion

This paper extends the Mirrlees optimal taxation model by enriching the labor market to incorporate an assignment problem between workers and firms. As first shown by Rosen (1981), assignment models of this sort are capable of producing superstar phenomena, where innate skill differences are greatly magnified in terms of earnings. Despite the potential to dramatically increase inequality, our results show that these superstar effects do not provide a basis for higher taxation. Indeed, depending on the interpretation, they are either neutral or provide a force for lower taxes.

It is worth clarifying that this force for lower taxes does not, by itself, imply that taxes at the top of the earnings distribution should be low or lower than current tax rates. Other forces or considerations, unrelated to superstars, may be at play at the top shaping the optimal level of taxes. Here we mention a few.

Stiglitz effects. As mentioned in the introduction, an active current literature explores how general equilibrium effects affect the optimal policy when tax instruments are limited. This literature revisits Stiglitz (1982) but using richer Roy models where relative wages across different sectors enter incentive constraints and tax policy seeks to manipulate wages to relax these constraints, typically implying lower taxes at the top. Since these "Stiglitz effects" are separate from superstar effects, we have abstracted from them in this paper.

Inefficiencies and uncertainty. It has been argued that some top earners earn excessive incomes not just because of superstar effects, but because their pay exceeds their marginal product. This could be due to rent-seeking (Rothschild and Scheuer, 2014a), inefficient bargaining (Piketty et al., 2014) or other inefficiencies (Rothschild and Scheuer, 2014b; Lock-

for some $\delta : [0,1] \to \mathbb{R}$. We define

$$\frac{\partial}{\partial y(t)} \Psi(\{y\}) = \delta(t).$$

wood et al., 2014) at the top of the income distribution. On the other hand, pay can be below marginal product, for instance for innovators that only appropriate a fraction of the value of their innovations, as in endogenous growth models. Finally, some top incomes may be due to luck. Inefficiencies and uncertainty of this kind provide Pigouvian and insurance motives for taxes. However, these issues arise even in standard Mirrlees models without superstar effects. To isolate the role of superstars, we have, therefore, ignored them here.

Low effort elasticities. One may argue that superstars at the top of the earnings distribution have very low labor supply elasticities at the margin, 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 completely unrelated to the presence of superstar effects and is also present in a standard Mirrleesian model. Whatever the effort elasticity of top earners, our second result shows that this elasticity needs to be adjusted upwards when accounting for superstar effects. For instance, our Example B 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

_ , Andres Bellofatto, and Jessie Wang, "Taxing Atlas: Using Firm Data to Derive Optimal Income Tax Rates," *Working Paper*, 2015. 1, 7.1

Ales, Laurence, Musab Kurnaz, and Christopher Sleet, "Tasks, Talents, and Taxes," *American Economic Review*, 2014, *forthcoming.* 1

Auten, Gerald and Robert Carroll, "The Effect of Income Taxes on Household Income," *Review of Economics and Statistics*, 1999, *81*, 681–693. 21

Blundell, Richard, Alan Duncan, and Costas Meghir, "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 1998, *66*, 827–862. 18

Chetty, Raj, "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, *80* (969-1018). 15

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. 21

Diamond, Peter and James Mirrlees, "Optimal Taxation and Public Production: I–Production Efficiency," *American Economic Review*, March 1971, *61* (1), 8–27. 2.1, 6.2

Edmans, Alex and Xavier Gabaix, "Executive Compensation: A Modern Primer," *Working Paper*, 2015. 1

Eissa, Nada and Hilary Hoynes, "The Earned Income Tax Credit and the Labor Supply of Married Couples," *NBER Working Paper 6856*, 1998. 18

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. 21

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. 21

Fuchs, William, Luis Garicano, and Luis Rayo, "Optimal Contracting and the Organization of Knowledge," *Review of Economic Studies*, 2015, *forthcoming.* 7.2, 7.2, 7.2

Gabaix, Xavier and Augustin Landier, "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 2008, *123*, 49–100. 1, 1, 2.1, 2.3, 17

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. 7.2, 7.2

— **and** — , "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 2006, *121*, 1383–1435. 7.2, 27

— **and** — , "Knowledge-Based Hierarchies: Using Organizations to Understand the Economy," *Annual Review of Economics*, 2015, *7*, 1–30. 1, 26

Gruber, Jon and Emmanuel Saez, "The Elasticity Of Taxable Income: Evidence And Implications," *Journal of Public Economics*, 2002, *84*, 1–32. 21

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. 5.3

Jaffe, Sonia and Scott Kominers, "Taxation in Matching Markets," *Working Paper*, 2014. 7.3

Keane, Michael, "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 2011, *49*, 961–1075. 15

— **and Richard Rogerson**, "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom," *Journal of Economic Literature*, 2012, *50*, 464–76. 15

Kopczuk, Wojciech, "Tax Bases, Tax Rates and the Elasticity of Reported Income," *Journal of Public Economics*, 2005, *89*, 2093–2119. 21

Lockwood, Ben, Charles Nathanson, and Glen Weyl, "Taxation and the Allocation of Talent," *Working Paper*, 2014. 8

Lucas, Robert, "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 1978, *9*, 508–523. 1, 7.1

MaCurdy, Thomas, "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, 1981, *89*, 1059–1085. 18

Mirrlees, James, "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208. 1, 2.1

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, 8

Prescott, Edward, "Why Do Americans Work so Much More Than Europeans?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2004, *28*, 2–13. 21

Rosen, Sherwin, "The Economics of Superstars," *American Economic Review*, 1981, *71*, 845–858. 1, 2, 3, 25, 8

—, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 1982, *13*, 311–323. 1, 7.1

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," *Working Paper*, 2014. 1, 8

— **and** —, "A Theory of Income Taxation under Multidimensional Skill Heterogeneity," *Working Paper*, 2014. 8

Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 2001, *68* (1), 205–29. 1, 1, 4.2, 5.3

—, "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. 21

—, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010, *2*, 180–212. 18

—, **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. 15

Samuelson, Paul, *The Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press, 1947. 5.2

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

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. 21

Stiglitz, Joseph, "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 1982, *17*, 213–240. 1, 7.1, 8

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

Werning, Iván, "Pareto Efficient Income Taxation," *Mimeo, MIT*, 2007. 1, 4, 4.2, 4.2, 6.1

Ziliak, James and Thomas Kniesner, "Estimating Life Cycle Labor Supply Tax Effects," *Journal of Political Economy*, 1999, *107*, 326–359. 18

A Appendix: Derivations for Examples A and B

A.1 Example A

Exact Earnings Schedule and Approximate Pareto Skill Distribution. We begin with Example A in Section 2.3 and construct the skill distribution $F(\theta)$ as claimed in the text. Taking $W(y)$ as given by (9), individuals solve

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

The first-order condition is

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

and the second-order condition for a maximum is satisfied if $\gamma(1-\beta\rho) > 1$. Substituting $y(n)$ from (8) for y yields

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

which simplifies after substituting $\tilde{\kappa}$ to

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

or equivalently

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

This implicitly defines $n = 1 - F(\theta)$ given θ 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, θ has a Pareto tail with parameter $\alpha > 0$. Conversely, solving for ρ yields (10) 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, δ, w_0. Taking τ 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

$$y(\theta) = \theta^{\frac{\gamma}{\gamma-\delta}} \left(\delta k(1-\tau)\right)^{\frac{1}{\gamma-\delta}} \tag{67}$$

as well as its inverse

$$\Gamma(y) = y^{\frac{\gamma-\delta}{\gamma}} \left(\delta k(1-\tau)\right)^{-\frac{1}{\gamma}}.$$

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

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

Matching coefficients requires

$$\delta k = \left(\delta k(1-\tau)\right)^{-\frac{\alpha\beta}{\gamma}} \underline{\theta}^{-\alpha\beta}$$

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 δ and k in (67) 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 (67), 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 (10). Hence, (8) holds approximately for small enough n. Again using (10), we have $\delta = 1/(1-\beta\rho)$, so

$$W(y) = ky^{\frac{1}{1-\beta\rho}} + w_0.$$

For large enough y, this approximates (9). This finally implies that (7) also holds approximately.

A.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 τ 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

$$\frac{(b-y)^{\delta-\gamma}}{(1-\tau)\delta k} = (a-\theta)^{-\gamma}. \tag{68}$$

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

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

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 δ:

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

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 (68), 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 (11). Hence, we match (8) for some appropriate κ. The same is true for (9) and hence (7).

B Appendix: Omitted Proofs

B.1 Proof of Lemma 1

Take any allocation satisfying conditions (1), (12) and (13). 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, (12) and (13).

B.2 Proof of Proposition 1

Pareto problem. The Pareto problem (16) s.t. (12), (14) and (17) can be rewritten as

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

s.t.

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

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

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

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

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

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

for any non-negative Pareto weights g. Define $\hat{\mu} \equiv U_c \mu/\eta$, so

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

Substituting this yields

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

Note that

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

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

$$-\hat{\mu}' - \hat{\mu} MRS_c y' \leq f. \tag{69}$$

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

$$\eta(B_y - MRS) f = \mu(U_{\theta c} MRS + U_{\theta y})$$

or

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

By (18), this becomes

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

Also,

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

so the first-order condition for y finally becomes

$$
\frac{\tau}{1-\tau}f=-\hat{\mu}\frac{\partial\log MRS}{\partial\theta}. \tag{70}
$$

Elasticities. To rewrite these first-order conditions for Pareto efficiency in terms of elasticities, we compute the earnings elasticities defined in Section 4.1. Recall that the earnings function $w(1-\tau,I)$ is 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 (dropping arguments)

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

and hence the (uncompensated) earnings elasticity is

$$
\varepsilon^u=\left.\frac{\partial w}{\partial(1-\tau)}\right|_{\tau=I=0}\frac{1-T'}{w}=\frac{W'/w-MRS_c}{MRS_c+\frac{MRS_y}{MRS}-\frac{W''}{W'}+\frac{T''}{1-T'}W'}. \tag{71}
$$

Moreover,

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

so

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

and

$$
\varepsilon^c=\varepsilon^u+\eta=\frac{W'/w}{MRS_c+\frac{MRS_y}{MRS}-\frac{W''}{W'}+\frac{T''}{1-T'}W'}. \tag{72}
$$

As a result, for later use,

$$
\frac{\varepsilon^c-\varepsilon^u}{\varepsilon^c}=\frac{\eta}{\varepsilon^c}=MRS_c\frac{w}{W'} \tag{73}
$$

and

$$
\frac{1}{w\varepsilon^c}=\frac{1}{W'}\left(MRS_c+\frac{MRS_y}{MRS}-\frac{W''}{W'}\right)+\frac{T''}{1-T'}. \tag{74}
$$

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

$$\left(MRS_c(1 - T') + MRS_y \frac{1}{W'}\right) w' + MRS_\theta = \left(-T''W' + (1 - T')\frac{W''}{W'}\right) w',$$

so

$$w'(\theta) = \frac{-MRS_\theta}{MRS_c(1 - T') + \frac{MRS_y}{W'} + T''W' - (1 - T')\frac{W''}{W'}}$$

or, dividing both numerator and denominator by $MRS = (1 - T')W'$,

$$w'(\theta) = \frac{-\partial \log MRS/\partial\theta}{\frac{1}{W'}\left(MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'}\right) + \frac{T''}{1-T'}} = -\frac{\partial \log MRS}{\partial\theta} w\varepsilon^c(w) \tag{75}$$

by (74). Next, define $\hat{\mu}(w) \equiv \mu(w^{-1}(w))$ where $w^{-1}(w)$ is the inverse of $w(\theta)$, so

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

Using this and $w'(\theta) = W'(y(\theta))y'(\theta)$ in (69) yields

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

or

$$-\hat{\hat{\mu}}'(w) - \hat{\hat{\mu}}(w)\frac{\partial MRS(w)}{\partial c}\frac{1}{W'(W^{-1}(w))} \leq h(w). \tag{76}$$

Similarly, we can rewrite (70) using (75) as

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

and so

$$\hat{\hat{\mu}}(w) = \frac{T'(w)}{1 - T'(w)}\varepsilon^c(w)wh(w). \tag{77}$$

Test Inequality. Taking logs of (77) yields

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

and differentiating w.r.t. $\log w$:

$$\frac{\hat{\mu}'(w)w}{\hat{\mu}(w)} = \frac{d\log\hat{\mu}(w)}{d\log w} = \frac{d\log\left(\frac{T'(w)}{1-T'(w)}\varepsilon^c(w)\right)}{d\log w} + \frac{d\log h(w)}{d\log w} + 1. \tag{78}$$

Returning to (76), 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 (77) on the RHS and (78) on the LHS, and use (73) for the income effect, to get

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

which simplifies to

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

as claimed in Proposition 1.

B.3 Proof of Corollary 1

The Rawlsian optimal tax schedule obtains as a special case where $\lambda(\theta) = 0$ for all types except for the lowest, and hence where (76) holds with equality. Integrating this ODE yields

$$\hat{\mu}(w) = \int_w^\infty \exp\left[\int_w^s \frac{\partial MRS(t)}{\partial c}\frac{1}{W'(W^{-1}(t))}dt\right]h(s)ds = \int_w^\infty \exp\left[\int_w^s \frac{\eta(t)}{\varepsilon^c(t)}\frac{dt}{t}\right]dH(s)$$

where the second equality follows from (73). Substituting this in (77) delivers the result.

B.4 Proof of Proposition 2

We obtain the fixed-assignment elasticities as a special case of (71) and (72) replacing W by $\hat{W}_0$ and evaluating at w_0, y_0. Using $\hat{W}_0'(y_0) = W'(y_0)$, this yields

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

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

and

$$\hat{\varepsilon}^c = \hat{\varepsilon}^u + \hat{\eta} = \frac{W'/w_0}{MRS_c + \frac{MRS_y}{MRS} + \frac{T''}{1-T'}W' - \frac{\hat{W}_0''}{W'}}. \tag{79}$$

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

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

as claimed in the proposition.

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

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

so the necessary second-order condition for a maximum is

$$W''(1-T') - W'^2 T'' - MRS_c W'(1-T') - MRS_y \leq 0$$

or equivalently

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

This implies

$$\frac{W'/w_0}{\hat{\varepsilon}^c} \geq \frac{W'' - \hat{W}_0''}{W'} = \frac{B_{y\theta}\Gamma'}{W'}$$

and hence

$$\hat{\varepsilon}^c \frac{B_{y\theta}\Gamma' w_0}{W'^2} \leq 1 \quad \Leftrightarrow \quad \hat{\Phi} \geq 0.$$

B.5 Proof of Lemma 2

The individual's problem is

$$\max_y U((1-\tau)W(y), y, \theta).$$

The necessary first-order condition is

$$(1-\tau)U_c(c, y, \theta)W'(y) + U_y(c, y, \theta) = 0$$

or equivalently

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

The necessary second-order condition is (after using the first-order condition to cancel terms and dropping arguments)

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

or

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

where the last step uses (80).

B.6 Proof of Proposition 3

The effort elasticities can be obtained as a special case from (71), (72) and (73) as a special case for $W(y) = y$ and $T'(w) = \tau$. Then we have

$$\tilde{\varepsilon}^c = \frac{1/y}{MRS_c + \frac{MRS_y}{MRS}} \tag{80}$$

and

$$\frac{\tilde{\eta}}{\tilde{\varepsilon}^c} = MRS_c y. \tag{81}$$

Comparing the former with (72) and following the same steps as in the proof of Proposition 2 yields (31) and (33). Comparing the latter with (73) yields (32). 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, $\tilde{\Phi} > 0$ follows from Lemma 2.

B.7 Proof of Proposition 4

Under the assumption that $B(\theta, y) = b(\theta)y$, we have

$$\overline{W}_0^{-1}(w|\tau) = (w + \pi(\theta_0|\tau))/b(\theta_0)$$

so the first-order condition corresponding to (40) 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 $\phi_y = MRS$)

$$\left.\frac{d\overline{w}_0}{d(1-\tau)}\right|_{\tau=0} = \frac{W' - \frac{MRS_y}{W'} \left.\frac{\partial \pi(\theta_0|\tau)}{\partial(1-\tau)}\right|_{\tau=0}}{\frac{MRS_y}{W'} + T''W'}$$

and hence

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

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 (79)), (ii)

$$\frac{MRS_y y}{MRS} = \frac{1}{\tilde{\varepsilon}}$$

by (80), and (iii)

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

by (39).

For the second part, write the problem of an individual of type θ as

$$y(\theta|\tau) \in \arg\max_y (1-\tau)b(\theta)y - T(b(\theta)y - \pi(\theta|\tau)) - e(y,\theta). \tag{82}$$

By Topkis's theorem, we know that (i) y is increasing in $1-\tau$ for each θ when holding π fixed, and that (ii) y is increasing in π for each θ holding τ fixed when $T'' \geq 0$.

Next, from the firms' problem (3) and by the envelope theorem,

$$\pi'(\theta|\tau) = b'(\theta)y(\theta|\tau) \quad \forall\theta$$

and, since profits are zero for the lowest firm,

$$\pi(\theta|\tau) = \int_{\underline{\theta}}^{\theta} b'(s)y(s|\tau)ds. \tag{83}$$

Hence, π is increasing in $1-\tau$ for each θ if $y(\theta|\tau)$ is increasing in $1-\tau$ for each θ. Therefore, the fixed point of (82) and (83), i.e., the schedules $y(\theta|\tau)$ and $\pi(\theta|\tau)$ that satisfy (82) and (83) simultaneously for any given τ, involves a profit schedule such that

$$\left.\frac{\partial\pi(\theta|\tau)}{\partial(1-\tau)}\right|_{\tau=0} > 0$$

if $T'' \geq 0$. This implies $\chi < 0$ in the proposition.

B.8 Proof of Corollary 3

In the recalibrated non-superstar economy, individuals solve

$$\max_{w} \hat{U}(w - T(w), w, \theta) = U(w - T(w), \Delta(\theta, w), \theta)$$

with first-order condition

$$(1 - T'(w))U_c(w - T(w), \Delta(\theta, w), \theta) + \frac{U_y(w - T(w), \Delta(\theta, w), \theta)}{B_y(\theta, \Delta(\theta, w))} = 0.$$

To see that the original $w(\theta)$ satisfies this for each θ, substitute

$$(1 - T'(w(\theta)))W'(y(\theta))U_c(w(\theta) - T(w(\theta)), y(\theta), \theta) + U_y(w(\theta) - T(w(\theta)), y(\theta), \theta) = 0,$$

and observe that this is the first-order condition in the original superstar economy since $B_y(\theta, \Delta(\theta, w(\theta))) = B_y(\theta, y(\theta)) = W'(y(\theta))$. Note that each θ will also have the same $y(\theta) = \Delta(\theta, w(\theta))$ as before.

We now compute the earnings elasticities corresponding to $\hat{U}$ in the non-superstar economy. The uncompensated earnings function is

$$\hat{w}(1-\tau, I) \in \arg\max_{w} \hat{U}((1-\tau)w - T(w) + I, w, \theta) = U((1-\tau)w - T(w) + I, \Delta(\theta, w), \theta)$$

with first-order condition

$$(1 - \tau - T'(w))B_y(\theta, \Delta(\theta, w)) = MRS((1-\tau)w - T(w) + I, \Delta(\theta, w), \theta).$$

Hence, the uncompensated earnings elasticity is

$$\frac{B_y/w - MRS_c}{MRS_c + \frac{MRS_y}{MRS} + \frac{T''}{1-T'}B_y - \frac{B_{yy}}{B_y}} = \hat{\varepsilon}^c$$

since $B_y = \hat{W}_0'$ and $B_{yy} = \hat{W}_0''$ when evaluated at θ such that $w(\theta) = w_0$. The equality of the income effect and compensated earnings elasticity can be shown analogously.

We also note that $\hat{U}$ satisfies single-crossing locally at the original allocation. The marginal rate of substitution associated with $\hat{U}$ is

$$M\hat{R}S(c, w, \theta) = \frac{MRS(c, \Delta(\theta, w), \theta)}{B_y(\theta, \Delta(\theta, w))}$$

and we require this to be decreasing in θ for single-crossing. Differentiating, we require

$$\frac{MRS_\theta + MRS_y \Delta_\theta}{B_y} - \frac{MRS}{B_y^2}\left(B_{y\theta} + B_{yy}\Delta_\theta\right) < 0.$$

We know $MRS_\theta < 0$ by single-crossing of the original utility function U, and $B_{y\theta} > 0$. The remaining terms are

$$\left(\frac{MRS_y}{MRS} - \frac{B_{yy}}{B_y}\right)\Delta_\theta.$$

By definition of $\Delta(\theta, w)$,

$$B(\theta, \Delta(\theta, w)) = w + \pi(\theta) \quad \forall\, w, \theta.$$

Differentiating w.r.t. θ,

$$B_\theta + B_y \Delta_\theta = \pi'$$

and hence

$$\Delta_\theta(\theta, w) = \frac{\pi'(\theta) - B_\theta(\theta, \Delta(\theta, w))}{B_y(\theta, \Delta(\theta, w))}.$$

By the envelope theorem, $\pi'(\theta) = B_\theta(\theta, y(\theta)) = B_\theta(\theta, \Delta(\theta, w(\theta)))$. This implies that $\Delta_\theta = 0$ when evaluated at $w(\theta)$. Hence, the remaining terms vanish at the original allocation.

B.9 Proof of Proposition 5

Recall the first-order conditions for V and y from Appendix B.2, given by (69) and (70), and rewrite them as

$$-\hat{\mu} MRS_c y' = -\hat{\mu}\frac{\tilde{\eta}}{\tilde{\varepsilon}^c}\frac{y'}{y} = f - \lambda U_c/\eta + \hat{\mu}' \leq f + \hat{\mu}' \tag{84}$$

and

$$\hat{\mu} = \frac{\tau}{1-\tau}\theta f\left(-\frac{MRS_\theta \theta}{MRS}\right)^{-1} = \frac{\tau}{1-\tau}\theta f \xi, \tag{85}$$

where we used $\tilde{\eta}/\tilde{\varepsilon}^c = MRS_c y$ and the definition of ξ. Take logs of (85) 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 (84), dropping the Pareto weights, by $\theta/\hat{\mu}$ yields

$$-\frac{\hat{\mu}'\theta}{\hat{\mu}} - \frac{\tilde{\eta}}{\tilde{\varepsilon}^c}\frac{y'\theta}{y} \leq \frac{\theta f}{\hat{\mu}} = \frac{1}{\frac{\tau}{1-\tau}\xi}$$

or finally

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

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 (80) and (81), this implies

$$\xi=\frac{\tilde{\varepsilon}^c}{1+\tilde{\varepsilon}^u}$$

as claimed in the text.

B.10 Proof of Proposition 6

For general Pareto weights Λ, (84) 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 (85) yields the result.

B.11 Proof of Lemma 3

The first-order condition corresponding to (55) 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 η is the multiplier on the labor market clearing constraint (52). The first-order condition corresponding to (49) is

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

Hence $\eta=w$. Next,

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

B.12 Proof of Lemma 4

Recall the definition of Ψ

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

s.t.

$$P'(t) = \frac{1}{h(1-y(P(t)))} \quad \forall t \in [t_2, 1]$$

and $P(1) = t_1$, $P(t_2) = 0$. The corresponding Lagrangian is, after integrating by parts

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

For any $t \in (t_2, 1)$, we therefore immediately have

$$\frac{\partial \Psi(\{y\})}{\partial y(t)} = \frac{1}{h(1-y(P(t)))}.$$

Together with (60), this implies

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

for the managers as desired.

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

$$M'(t) = h(1-y(t)) \quad \forall t \in [0, t_1] \tag{86}$$

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:

$$\begin{aligned} \int_{t_2}^{1} \frac{y(t)}{h(1-y(P(t)))} dt &= \int_{P(t_2)}^{P(1)} \frac{y(M(t))}{h(1-y(t))} h(1-y(t)) dt \\ &= \int_{0}^{t_1} y(M(t)) dt. \end{aligned}$$

Hence, we can rewrite Ψ as

$$\Psi(\{y\}) = \max_{M,t_1,t_2} \int_{0}^{t_1} y(M(t)) dt + \int_{t_1}^{t_2} y(t) dt$$

s.t. (86) and and $M(t_1) = 1$, $M(0) = t_2$. The corresponding Lagrangian is

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

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,

$$\frac{\partial\Psi(\{y\})}{\partial y(t)} = hW_m(y(M(t))).$$

Together with (61), this shows that

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

for workers as desired.

Finally, as for the self-employed, trivially

$$\frac{\partial\Psi(\{y\})}{\partial y(t)} = 1 = W'(y(t)).$$