Taxation and Sources of Inequality*

Yena Park

Seoul National University

November 1, 2023

Abstract

How should optimal tax be affected by sources of income inequality? To answer this question, we extend the Pareto efficient tax analysis in Mirrlees' model by incorporating endogenous wage determination that can capture various sources of income inequality— return to skill, return to education, and return to effective labor supply. We show that when an inequality factor is introduced, not only earnings distribution but also earnings elasticities change, and the impact on optimal tax rates can be in the opposite directions depending on the sources of inequality. To figure out which forces dominate, we investigate structural model, and show that effects boil down to *effective efficiency cost* of tax distortion canceling out the counterbalancing forces. Our analysis also provides two important cautions for the sufficient statistics approach. First, when sufficient statistics approach is used for a future comparative static, estimated elasticities should be adjusted according to the expected sources of inequality. Second, when testing the current tax schedule given earnings distribution, elasticities should be estimated under the current inequality factor.

Keywords: Pareto Efficiency Test, Income Inequality, Optimal Tax
JEL Classification: E21, D64, H21

*Financial support from the Seoul National University Institute for Research in Finance and Economics(SIRFE) is gratefully acknowledged.

1 Introduction

In the recent decades, the US has experienced substantial growth in income inequality, and there has been development of a huge literature that explains the rise in income inequality. Compared to the positive analysis, however, there is relatively small normative literature that studies how optimal taxes should respond to income inequality. Even those small normative studies mostly take one or two channels of generating wage inequality and investigate how optimal taxes are affected by a specific channel of inequality (e.g., Ales, Kurnaz, and Sleet (2015), Scheuer and Werning (2017), Heathcote, Storesletten, and Violante (2020)), and they have very different implications.[1] Despite the different role of each channel of inequality, we still have limited understanding on what are the key differences of these models that lead to contrasting policy suggestions. This paper tires to provide better understanding on the fundamental reason of different results so that we can have a better policy suggestion by applying a suitable analysis.

When inequality increases, the value of redistribution increases regardless of its sources, but elasticities of earnings with respect to tax rate—the efficiency cost of redistributive taxation—are also endogenous to inequality factors. Moreover, the change in elasticities does depend on different inequality factors, which can lead to the opposite implications for optimal tax rates depending on the sources. Thus we answer the following main questions in this paper. How do earnings elasticities respond to the change in different inequaltiy factors, and as a result how does the efficient tax schedule respond? The fact that elasticities are endogenous to different sources of inequality provide important cautions for using the sufficient statistic for the optimal tax analysis.What cautions should be taken for the sufficient statistics approach?

[1]For example, Ales, Kurnaz, and Sleet (2015) focus on the technical change and provide an important mechanism of shaping optimal taxation. Scheuer and Werning (2017) study optimal tax in the presence of superstar effects, where better assignment of superstars provides a force for lower marginal tax with higher responsiveness of individual earnings. Heathcote, Storesletten, and Violante (2020) is an exception which investigates the implications of two different factors of income inequality—rise in skill premium due to skill biased technical change and growth in residual wage dispersion due to increased increased labor market uncertainty—on the optimal tax progressivity, but the counterbalancing forces for optimal schedule are still very specific to the model.

To answer these questions, this paper studies the set of Pareto efficient tax schedules in an extension of Mirrlees (1971)' model.[2] Workers with exogenous skills make a decision of human capital investment and working hours to provide effective labor supply, and are paid earnings according to a wage schedule.[3] With a model that can encompass various factors—return to skill, return to human capital, and return to effective labor supply—of generating realistic earnings inequality, we investigate the Pareto efficiency test of a tax schedule as in Werning (2007).[4] We do this in a simplest possible framework. We consider a redistributive taxation that only depends on earnings, which does not depend on skill, human capital, and working hours separately. To focus on the first order effects of inequality factors on the equity-efficiency trade-off, we abstract from "Stiglitz effects"—general equilibrium effects through aggregate complementarity in production (Stiglitz (1982))—and other indirect channels such as profit spillover effects and assignment change due to the general equilibrium effects.

The main results from our analysis can be best explained by the top income tax rate example. When the skill distribution at the top exhibits Pareto distribution, under some assumptions, the Pareto efficiency test formula of the top income tax rate is simply:

$$\tau_{top} \leq \frac{1}{1 + \epsilon_\omega \times a_\omega^{top}},$$

where ϵ_ω is the earnings elasticity with respect to marginal tax rate and a_ω^{top} is the Pareto parameter of earnings distribution at the top. Higher inequality implies lower Pareto parameter a_ω^{top} regardless of its sources providing forces for higher tax rates. What is more important is that the earnings elasticity ϵ_ω also changes, and its direction depends on the sources of inequality. The effects of inequality on top income tax rate τ_{top} thus depends on how the earnings elasticity responds to each inequality factor and which effect dominates between the changes in distribution and elasticities.

[2]We take the Pareto efficient tax analysis rather than welfare maximizing optimal tax analysis. Since the change in value of redistribution is too sensitive to the social welfare, we intentionally take Pareto efficiency criterion which does not depend on social welfare weight.

[3]In the benchmark model, the effective labor supply of skill type θ is $y(\theta) = p(\theta, e) \cdot n$ with productivity of skill type θ and human capital level e $p(\theta, e)$ and working hours n. This effective labor supply $y(\theta)$ generates earnings $\omega(\theta) = W(y(\theta))$, where $W(\cdot)$ is the wage schedule.

[4]Pareto efficiency test was first developed by Werning (2007) and has been extended to a model with superstars (Scheuer and Werning (2017)) and a life-cycle model with bequest (Hosseini and Shourideh (2019)).

Our structural investigation for a comparative static shows that after canceling out the counterbalancing forces, the effect of rising inequality given a skill distribution boils down to a change in the *effective efficiency cost* (hereafter EEC) of tax distortion. The effective efficiency cost is obviously increasing in elasticity of effective labor supply. The EEC also depends on the differential of effective labor supply across skill types.[5] Higher differential of effective labor supply across skill type decreases the EEC as it tightens the incentive constraint by making higher types easy to mimic lower types which increases the effectiveness of tax for relaxing the incentive constraint. Thus an inequality factor leading to higher elasticity of effective labor supply leads to higher EEC and thus providing force for a lower tax rate τ_{top}, while an inequality factor leading to higher labor supply differential implies lower EEC which increases the upper bound of τ_{top}.

When rise in inequality is driven by higher return to unobserved (and exogenous) skill, higher differential of effective labor supply across skills reduces the EEC of tax distortion as it tightens the incentive constraints, which provides a force for higher tax rate τ_{top}. When higher return to human capital raises inequality, however, both differential of effective labor supply and the elasticity of effective labor supply increase, which have the opposite implications on the EEC. If the return to unobserved skill is relatively big (compared to the return to human capital), the elasticity channel dominates, and thus it leads to higher EEC and lower top income tax rate, which is the opposite direction from the case of rising return to skill. On the other hand, when return to effective labor supply increases, there is no impact on the EEC and Pareto efficient tax rates, as the counterbalancing forces to the distribution and earnings elasticity exactly cancel out.

Our analysis uncovers common conditions for the efficiency of a tax schedule. The Pareto efficiency test formula expressed in terms of earnings distribution is equivalent to the test formula in the standard Mirrlees model despite the endogenous wage structure with inequality factors. This implies that the neutrality result of the efficiency test of Scheuer and Werning (2017), which is obtained in a model with superstar effects, is extended to a more general

[5] The expression of EEC is simply product of the elasticity of effective labor supply and the inverse of the differential of effective labor supply across skill types when the Hicksian complementarity between education and skill is one.

framework that can capture various factors of income inequality. More precisely, the test formula is expressed in terms of common sufficient statistics—the earnings distribution and the elasticities of earnings with respect to marginal tax rate—regardless of the underlying sources of generating inequality. Even in the models with imperfect substitution between working hours and Human capital investment effort, the elasticities of effective unit of labor supply are sufficient statistics as long as the government does not regulate working hours and effort separately.

How does the EEC respond to different sources of inequality? When rise in inequality is driven by increasing return to (unobserved) skill, higher labor supply differential across skill type (with no change in elasticities) implies lower EEC and higher τ_{top}. When the increase in inequality is generated by higher return to human capital (education), however, it can have the opposite implication—higher EEC and lower τ_{top}. Both elasticity of effective labor supply and labor supply differential increases, and the elasticity effects will be larger when return to skill is relatively higher than the return to human capital, implying higher EEC. On the other hand, increase in inequality driven by return to effective labor supply has no effects on the EEC, which coincides with the result in Scheuer and Werning (2017).

The key observation from the structural investigation is that earnings elasticities are endogenous to sources of inequality. This observation also provides us important cautions for using the sufficient statistic approach for a policy suggestion at a time when inequality is changing significantly. First caution is for the application of the sufficient statistic approach for the future comparative static when rise in income inequality is expected. Suppose that one tries to use the top income tax formula discussed above for the future policy suggestion. Typical sufficient statistic approach would evaluate the formula with predicted lower level of Pareto parameter a_{ω}^{top} and earnings elasticity estimated from current (or past) data assuming the elasticities do not change in the future. We show that this approach does not cause any problem only if higher return to unobserved skill is the reason for rise in inequality. For example, in our benchmark model, with 10% decrease in Pareto parameter, typical sufficient statistic approach would predict 6.1% increase in the upper bound of top income tax rate. The upper bound of top income tax rate evaluated with correct prediction of earnings elasticities, however, would be expected to decrease by 1.3% when return to human capital is the source of inequality, and there will be no change if higher return to effective labor supply is the source.

The second caution is for using the sufficient statistic approach for testing the current tax schedule. Now the Pareto efficiency formula discussed above should be evaluated using the Pareto parameter of current earnings distribution and current earnings elasticity. Given the observed earnings distribution, as long as the earnings elasticities are estimated correctly, the sufficient statistic approach works well and sources of inequality does not matter. The problem is that, however, economists usually estimate the elasticities using the tax reform event in the past to minimize endogeneity issues in estimation. Then there is time gap between the point of estimation and the point of testing tax schedule. If the inequality gap between the two periods have been driven by rise in return to skill, we have the right estimated elasticities. However, if it was driven by the rise in return to either human capital or effective labor supply, then the estimated elasticites should be adjusted upward.

As we discussed above, there are few recent papers studying how an optimal tax schedule should respond to rise in income inequality. Heathcote, Storesletten, and Violante (2020) answers this question focusing on the response of optimal progressivity restricting the tool of the government to the parametric functional form of tax schedule—this log-linear function with two parameters is referred to as HSV tax function in the literature. They consider two factors of rising income inequality—skill premium due to the skill-biased technical change and growth in residual wage dispersion—and quantitatively show that overall optimal progressivity does not change that much (with modest decline in progressivity) as the two factors provide the opposite forces on the optimal progressivity. Wu (2021) also quantitatively investigates this issue under the same HSV tax function but with more inequality factors, and provides similar argument that rising inequality implies declining optimal progressivity.[6] Our analysis complements these Ramsey tax studies by extending analysis to the fully nonlinear tax schedule and by generalizing the model to a broader class of Mirrlees' tax model.

There is a literature on optimal Mirrlees tax analysis with a richer labor market by incorporating more realistic features of labor markets. Although the studies in this literature

[6]The decomposition results of Wu (2021) shows more emphasis on the role of aging population which has fiscal pressure.

does not directly address the main question of this paper—response of optimal tax to rising inequality, the various features of the labor market considered in this literature are the factors that can generate the realistic high inequality. The model with superstar effects considered in Scheuer and Werning (2017) is one special example of our analysis (which belongs to the first class of model) and assignment model of Ales, Kurnaz, and Sleet (2015) is another special example, but with the opposite implications on the optimal tax when earnings inequality increases. Our analysis provides a better understanding of these results in the previous literature by identifying how different sources of inequality differently affect behavioral response.

We take the Pareto efficient tax analysis rather than welfare maximizing optimal tax analysis. Werning (2007) is the first paper which developed tests for the Pareto efficiency of a tax schedule with a test formula that can be expressed in terms of sufficient statistics including current income distribution.[7] We extend this test to incorporate a broad class of model that can capture various sources of inequality. Since the observed income distribution is the result of multiple factors that generate earnings inequality, identifying the key factors that have different implications on the efficient tax schedule is important. We closely follow the analysis of Scheuer and Werning (2017) by showing the role of elasticities despite the neutrality result of the test formula. With a general framework that can encompass this superstar effect, we identify what is key feature of their model that is leading to different implications on optimal taxation relative to other models with different inequality factors. Our analysis clarifies that the fundamental difference is whether the inequality factor increases earnings differential through the effective labor supply channel or not.

[List of papers that need to be discussed:]

- Badel, Huggett, and Luo (2020), Ales and Sleet (2016), Scheuer and Slemrod (2020)
- Gruber and Saez (2002), Saez, Slemrod, and Giertz (2012), Blomquist and Selin (2010)
- Lochner, Park, Shin, and Ninth (2018)

[7] Bierbrauer, Boyer, and Hansen (2022) provides better understanding of this test formula, showing that this test is essentially equivalent to the condition that the revenue function associated with an elementary tax reform at each income level must be non-increasing.

2 Simple Framework

We describe a framework which extends the standard Mirrlees' optimal tax model. We provide the simplest possible static framework that can still incorporate various factors in the labor market so that we can generate realistic earnings inequality. In this section, we set up the planner's problem in this general framework. In the next section, we derive and analyze the Pareto efficiency test formula.

2.1 Environment

Standard Mirrlees' model assumes that the final production function is linear in the efficiency unit of labor supply—product of worker's skill and effort—and thus wage rate of each worker is simply worker's level of skill.[8] We generalize this canonical model to capture more realistic features of wage schedule and inequality. To focus on the first-order effects of inequality factors on the equity-efficiency trade-off, we intentionally abstract from "Stiglitz effects"—the general equilibrium effects through aggregate complementarity in production.

Preferences and Types We consider a model with continuous skill type and a continuum of workers. Workers have identical preferences represented by a utility function: $u(c, \tilde{e}, n)$, where c is consumption, $\tilde{e}$ is effort for the human capital investment, and n is working hours. The utility function $u(\cdot)$ is increasing in c and decreasing in $\tilde{e}$ and n, and it is concave, twice continuously differentiable, where the domain is $\mathbb{R}_+ \times [0, \bar{e}] \times [0, \bar{n}]$. $u(\cdot)$ satisfies the Inada conditions: for all $c > 0$, $lim_{e \to 0} u_e = lim_{h \to 0} u_h = 0$ and $lim_{e \to \bar{e}} u_e = lim_{h \to \bar{n}} u_n = -\infty$. The effort $\tilde{e}$ can capture general effort to increase working productivity, we will focus on the human capital investment interpretation.

We can capture two different types of human capital models with different forms of effort function $\tilde{e}(e, \theta)$, where e is the level of human capital investment and and θ is skill type of workers described below. If the effective effort $\tilde{e}(e, \theta)$ is decreasing in θ, then the cost of investing for the human capital level e is decreasing in skill type θ. We only consider two

[8] Alternative interpretation is that workers of different skill types are perfect substitutes in production and output is sum of worker's effort weighted by worker's skill level.

special cases: (a) $\tilde{e}(e,\theta) = e$ and (b) $\tilde{e}(e,\theta) = \frac{e}{\theta}$. In the first case, the cost of human capital investment is independent of skill type, although the return to human capital investment can still depend on the skill type as we describe below.

Workers are heterogeneous in their exogenously given skill types and skill types are distributed across an interval $[\underline{\theta}, \bar{\theta}]$ according to a distribution function $F(\theta)$ with strictly positive and continuously differentiable density $f(\theta)$.[9] Following the Mirrlees' spirit, θ can be interpreted as comprehensive income generating ability.

Wage Structure We denote effective unit of labor supply by $y(\theta) = p(\theta, e(\theta)) \cdot n(\theta)$, where $p(\theta, e)$ is effective productivity and $n(\theta)$ is working hours. Workers can increase the effective labor supply either by by increasing individual productivity $p(\theta, e(\theta))$ with human capital investment or by increasing labor supply $n(\theta)$. Workers who provide effective labor supply $y(\theta)$ are paid earnings $\omega(\theta) = W(y(\theta))$ according to a wage schedule $W(\cdot)$. If $W(\cdot)$ is convex at the top, it captures the superstar effects at the top income as in Scheuer and Werning (2017)—return to effective labor supply is increasing at the top.

With this simple earnings function in a static model, we can still capture various sources of inequality. Given skill distribution, we can consider rise in return to each of three factors as the sources of inequality—(i) return to skill θ ($\frac{p_\theta \cdot \theta}{p}$), (ii) return to human capital e ($\frac{p_e \cdot e}{p}$), and (iii) return to effective labor supply y ($\frac{W'(y) \cdot y}{W(y)}$). In this simple structural model, we can isolate the role of each source in determining optimal tax schedule. We do this by investigating how the earnings elasticities respond to change in each inequality factor and analyzing their impact on the Pareto efficient tax schedule. and this is because the response of earnings elasticities to each inequality factor is not the same. Although we mostly focus on these three sources in the benchmark analysis, in an extension, we also consider increase in variance of skill distribution as another source of inequality.

For later use, we also define the Hicksian coefficient of complementarity between skill θ and human capital e by $\rho(\theta, e) = \frac{p_{\theta e} \cdot p}{p_\theta \cdot p_e}$. For example, if the productivity function takes a form of Cobb-Douglas, $p(\theta, e) = \theta^\gamma \times e^\nu$, then $\rho(\theta, e) = 1$. With the CES form of production

[9] In the dynamic extension of the model, θ can be stochastic, but it is going to be still exogenous shock.

function $p(\theta,e)=\left((1-\alpha)\cdot\theta^{\gamma(1-\rho)}+\alpha\cdot e^{\nu(1-\rho)}\right)^{\frac{1}{1-\rho}}$, the Hicksian coefficient is $\rho(\theta,e)=\rho$.[10]

Worker's Problem Note that each worker with skill type θ provides labor supply $y(\theta)$ considering that they are paid according to a wage schedule $W(\cdot)$. The wage schedule $W(y)$ can be endogenously determined in equilibrium, but each worker takes it as given. In the benchmark, we assume that workers' productivity function $p(\theta,e)$ is exogenously given. $p(\theta,e)$ can be also endogenized in an extended model, but even then workers make a decision of e given $p(\theta,\cdot)$. That is, a worker with skill level θ solves the the following problem given wage schedule $W(\cdot)$, productivity function $p(\theta,\cdot)$, and income tax schedule $T(\cdot)$:

$$\max_{c,e,n} u(c,\tilde{e},n) \quad s.t. \quad c\leq W(p(\theta,e)\cdot n)-T\left(W(p(\theta,e)\cdot n)\right), \tag{1}$$

where $\tilde{e}=\tilde{e}(e,\theta)$.

To make this worker's problem comparable to the one in the standard Mirrlees' model, we rewrite the workers' problem as a two-step decision—1. choose (c,ω), and 2. choose (e,n) given (c,ω). Solving backwards, given consumption c and efficiency unit labor supply $y=W^{-1}(\omega)$, workers choose effort level e and working hours n:

$$U(c,y,\theta)=\max_{e,n} u\left(c,\tilde{e}(e,\theta),n\right) \quad s.t. \quad y\leq w(\theta,e)\cdot n, \tag{2}$$

where $U(c,y,\theta)$ is the associated value function. Workers then optimally choose c and ω to maximize the value function U:

$$\max_{c,\omega} U(c,W^{-1}(\omega),\theta) \quad s.t. \quad c\leq\omega-T(\omega). \tag{3}$$

Note that although we allow separate choice of human capital investment and working hours, as long as the tax function does not depend on e and n separately, the worker's problem boils down to the problem of choosing consumption and earnings in the standard Mirrlees model given value function U. As in the standard Mirrlees, we need single crossing condition to set up the planner's problem, but what is different is we need this condition in terms of the value function U not just in terms of primitive preferences. Thus we define the marginal rate of substitution by

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

[10]More simpler CES functional for with costant Hicksian coefficient is possible with $\gamma=\nu=1$, but in this case, return to θ and return to e is always negatively correlated.

and we assume the following Spence-Mirrleess single crossing property.

Assumption 1. *$MRS(c, y, \theta)$ is decreasing in θ.*

That is, for more skilled workers, it is less costly to provide effective labor supply y.

Technology and Firm's Problem By modeling a suitable technology and assuming (positive) sorting between firms and workers, we can endogenize either wage schedule $W(y)$ or productivity function $p(\theta, e)$. In this benchmark, we take a model that can endogenize the wage schedule $W(y)$—as in Scheuer and Werning (2017), and later discuss alternative modeling and the implication of each choice. Note that the modeling choice does not matter for the main results when we think of a change in return to each factor as an exogenous shock, which we take in the benchmark analysis.

Suppose that there is a continuum of unit measure of firms with heterogeneous productivity x, and firm's productivity is drawn from a distribution $\Gamma(x)$. We assume that a firm and a worker is matched according to the perfectly positive sorting based on the complementarity between firm's productivity x and worker's effective labor supply y. Then the assortative matching function $\tilde{\sigma}(y) = x$ is such that $F(y^{-1}(y)) = \Gamma(\tilde{\sigma}(y))$, which essentially leads to a matching function $\sigma(\theta) = \tilde{\sigma}(y(\theta))$. A firm with productivity x which is matched with a worker providing effective labor supply y produces final output $G(x, y)$, and we assume that there is complementarity between firm's productivity and worker's labor supply: $G_{xy} > 0$.

Given wage schedule $W(y)$, a firm with productivity x solves

$$\max_y G(x, y) - W(y), \tag{4}$$

whose first order condition is

$$G_y(x, y) = W'(y). \tag{5}$$

As Scheuer and Werning (2017) have shown, $G_{xy} > 0$ provides a force for a convex wage schedule $W(y)$. We can see this from

$$W''(y) = G_{xy}(\tilde{\sigma}(y), y)\tilde{\sigma}'(y) + G_{yy}(\tilde{\sigma}(y), y).$$

We want to remark that we don't have to assume that the whole wage schedule $W(y)$ is convex for our analysis.

We also remark that our model does not allow aggregate complementarity by assuming a technology with non-complementarity across $y(\theta)$'s.[11] This implies that we abstract from the general equilibrium effects ("Stiglitz effects"). We intentionally do not consider this additional channel to focus on the first order channels associated with rise in inequality.

Equilibrium An equilibrium consists of a tax schedule $T(\omega)$, an earning schedule $W(y)$, an allocation $\{c(\theta), y(\theta), e(\theta), n(\theta), \omega(\theta)\}$ and matching function $\sigma(\theta)$ such that (i) for each θ, $(c(\theta), e(\theta), n(\theta))$ solves (1) given $W(\cdot)$, with the associated effective unit of labor supply $y(\theta) = p(\theta, e(\theta)) \cdot n(\theta)$ and earnings $\omega(\theta) = W(y(\theta))$; (ii) for each x, $Y(x)$ solves $G_y(x, Y) = W'(Y)$; (iii) the goods market clearing condition (6) holds:

$$\int c(\theta) dF(\theta) \geq \int G(\sigma(\theta), y(\theta)) dF(\theta); \tag{6}$$

and (iv) the labor market clearing condition holds: $Y(\sigma(\theta)) = y(\theta)$, $F(\theta) = \Gamma(\sigma(\theta))$.

2.2 Planning Problem

In this section, we set up the planning problem and derive the conditions for Pareto optimality. This condition will be expressed in terms of elasticities, which will be used as a Pareto efficiency test formula. By investigating how this test formula responds to the introduction of each inequality factor, we analyze how Pareto efficient taxes are affected by different sources of inequality.

We assume that the planner does not regulate effort $e(\theta)$ and working hours $n(\theta)$ separately but only regulates earnings $\omega(\theta)$—or equivalently, the planner regulates the effective units of labor $y(\theta)$ as there is one to one relationship between y and ω given wage schedule $W(y)$. This is either because the government cannot observe $e(\theta)$ and $n(\theta)$ separately, or because the tax policy only depends on earnings $\omega(\theta) = W(y(\theta))$ due to other reasons (even though the government can observe $e(\theta)$ and $n(\theta)$). Then, given the optimal allocation $(c(\theta), y(\theta))$

[11] That is, different levels of effective unit of labor supply $y(\theta)$ are not complement each other in production.

chosen by the government, the allocation of $(e(\theta), n(\theta))$ solves (2), and the relevant worker's utility function is the value function $U(c(\theta), y(\theta), \theta)$.

For any allocation $(c(\theta), y(\theta))$, we denote the utility delivered this allocation by

$$v(\theta) = U(c(\theta), y(\theta), \theta).$$

An allocation is incentive compatible if

$$v(\theta) = \max_{\theta'} U(c(\theta'), y(\theta'), \theta) \quad \forall\theta. \tag{7}$$

To characterize Pareto efficient tax, we follow the conventional mechanism design approach of solving the set of optimal allocation and find the tax schedule that can implement optimal allocation as an equilibrium. As in standard Mirrlees (1971), an allocation $(c(\theta), y(\theta))$ can be implemented as an equilibrium for some tax schedule if and only if it satisfies resource feasibility (6) and incentive compatibility (7). By single crossing assumption, we can use the first order approach because the global incentive constraints are equivalent to (i) monotonicity condition: $y'(\theta) \geq 0$; and (ii) the envelope conditions:

$$v'(\theta) = U_\theta(c(\theta), y(\theta), \theta) \quad \forall\theta. \tag{8}$$

Thus, we can replace the incentive-compatibility constraint by the envelope condition (8), and drop the monotonicity condition, and check the monotonicity condition ex post after solving the mechanism problem.

Planning Problem We denote a Pareto weight to workers of skill θ by $\lambda(\theta)$, and any Pareto weight should satisfy $\int \lambda(\theta) f(\theta) d\theta = 1$. Let $C(v, y, \theta)$ be the inverse function of $U(c, y, \theta)$, defined according to $v = U(C[v, y], y, \theta)$. We also denote the production function by $Q(\theta, y(\theta)) = G(\sigma(\theta), y(\theta))$. Then the planner's problem writes:

$$\max_{v(\theta), y(\theta)} \quad = \quad \int \lambda(\theta) v(\theta) dF(\theta) \tag{9}$$

$$(\mu(\theta)) \quad s.t. \quad v'(\theta) = U_\theta(c(\theta), y(\theta), \theta) \quad \forall\theta, \tag{10}$$

$$(\eta) \qquad \int \left(Q(\theta, y(\theta)) - C(v(\theta), y(\theta), \theta)\right) dF(\theta) \geq 0, \tag{11}$$

for some given Pareto weight $\lambda(\theta)$.

The planning problem (9) itself is not different from a standard Mirrleesian problem despite the extension of of the model. First, this is because the planner does not regulate the choice of effort $e(\theta)$ and working hours $n(\theta)$ separately, and thus the planner's problem boils down to solving optimal allocation of choosing consumption $c(\theta)$ and effective units of labor $y(\theta)$ (given the value function $U(c, y, \theta)$) as in the standard Mirrleesian problem. More important, endogenous wage structures are captured in a simplified reduced form: $W'(y(\theta)) = G_y(\sigma(\theta), y(\theta)) = Q_y(\theta, y(\theta))$. Depending on the shape of production technology G and the functional form (and its parameter values) of productivity function $p(\theta, e)$ hidden in the effective units of labor $y(\theta) = p(\theta, e(\theta)) \cdot n(\theta)$, we can capture various return to each factor in an equilibrium. Also, the value function of the worker $U(c, y, \theta)$ also depends on the shape of productivity function $p(\theta, e)$. The planning problem (9), however, can be written in a general form without specific assumptions on G, $p(\theta, e)$ and $U(c, y, \theta)$, and thus the planning problem boils down to the standard problem except for the general expression of production Q in the resource constraint and the value function U in the incentive constraint.

However, when we express the Pareto efficiency test in terms of sufficient statistics, the level of the elasticity of earnings with respect to marginal tax rate—one important sufficient statistic, does depend on the detailed assumption on the wage structure. Moreover, each factor of generating rise in inequality will have different implication on the worker's wage, and thus on the elasticities. In the next section, we investigate the relationship between the Pareto efficient taxes and sources of Inequality in detail.

Optimality Conditions: General Form We derive the optimality conditions of allocation from the first order conditions of the planning problem. These common optimality conditions expressed in a general form will be used to derive the Pareto efficiency test formula in various cases.

The first order conditions of the planning problem with respect $v(\theta)$ and $y(\theta)$ are as follows, respectively:

$$\lambda(\theta)f(\theta) - \eta e_v(\theta)f(\theta) - \mu(\theta)U_{\theta c}C_v(\theta) - \mu'(\theta) = 0 \tag{12}$$

$$\eta(Q_y(\theta) - C_y(\theta))f(\theta) = \mu(\theta)\left[U_{\theta c}(\theta)C_y(\theta) + U_{\theta y}(\theta)\right]. \tag{13}$$

These conditions can be rewritten in terms of marginal rate of substitutions:

$$-\hat{\mu}'(\theta) - \hat{\mu}(\theta) \cdot MRS_c \cdot y'(\theta) = f(\theta) - \frac{\lambda(\theta)U_c(\theta)}{\eta} \tag{14}$$

$$\hat{\mu}(\theta) = \frac{T'(\omega(\theta))}{1 - T'(\omega(\theta))}\theta f(\theta) \left(-\frac{MRS_\theta \cdot \theta}{MRS(\theta)}\right)^{-1}. \tag{15}$$

See the appendix for the detailed derivation.

We say that a tax schedule is Pareto efficient if the allocation associated with a tax schedule solves the planning problem (9) for some nonnegative Pareto weight $\lambda(\theta)$. Thus an allocation implemented by a Pareto tax schedule should satisfy these optimality conditions (14) and (15) for some Pareto weight. By combining $\lambda(\theta) \geq 0$ with (14), we get the inequality condition, then we can use (15) to derive a Pareto efficiency test formula for earnings taxes $T(\omega)$.

3 Pareto Efficienct Tax and Sources of Inequality

In this section, we investigate Pareto efficiency test formulas. The formulas are expressed in terms of relevant elasticities and these elasticiteis are endogenous to the change in inequality factors. We then proceed to the comparative static analysis to show how the Pareto efficient tax rates respond to the sources of inequality.

3.1 Pareto Effiociency Test Formulas

We begin by deriving the Pareto efficiency test formulas. The formulas can be expressed either in terms of skill distribution or in terms of earnings distribution. The (structural) formula expressed in terms of skill distribution is useful for a future comparative statics when earnings distribution is expected to changed by rise in income inequality. On the other hand, the formula expressed in terms of earnings distribution can be used for a sufficient statistics approach. This sufficient statistics formula can be used either for testing current tax schedule given observed current earnings distribution or for a comparative statics for the future by predicting change in earnings distribution.

Easticities Depending on whether we express the formulas in terms of earning distribution or in terms of skill distribution, the relevant elasticites required in the formulas is different. We define both elasticities of earnings and elasticities of effective labor supply.

First, we define the elasticities of earnings with respect to change in marginal tax rate, which are required sufficient statistics for the test formula in terms of earnings distribution. Recall that given tax schedule $T(\omega)$, workers' problem boils down to the problem of choosing optimal $(c(\theta), \omega(\theta))$ which solves (3). That is, given value function $U(c, y, \theta)$, wage schedule $W(\cdot)$ and tax schedule $T(\cdot)$, the earnings function is defined by

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

Then the uncompensated earnings elasticity, income effects on earnings, and the compensated elasticity are defined respectively:

$$\epsilon_\omega^u(\omega) = \left.\frac{\partial \omega}{\partial(1-\tau)}\right|_{\tau=I=0} \cdot \frac{1-T'(\omega)}{\omega}, \quad \eta_\omega(\omega) = -\left.\frac{\partial \omega}{\partial I}\right|_{\tau=I=0} \cdot (1-T'(\omega)), \quad \epsilon_\omega^c(\omega) = \epsilon_\omega^u(\omega) + \eta_\omega(\omega)$$

This elasticity measures the response of earnings to the change in marginal tax rate considering the (potentially convex) wage schedule $W(\cdot)$ and nonlinear tax schedule $T(\cdot)$ as in Scheuer and Werning (2017). Different from Scheuer and Werning (2017), however, the earnings elasticities depend on the value function, which is essentially determined by the elasticity of human capital investment, elasticity of working hours, and return to θ and e captured in productivity function $p(\theta, e)$. This elasticity of earnings is, in principle, the one we can estimate using the taxable income data.

Next, we also define the elasticities of effective units of labor y with respect to marginal tax rate change along the linear wage schedule and linear budget constraint, which are relevant elasticities in the tax formula in terms of skill distribution. By definition, these elasticities are equivalent to the elasticities of earnings assuming $W(y) = y$ and $T''(\omega) = 0$. Under this assumption, the function of effective units of labor supply is defined by $y(1-\tau, I) \in \arg\max_y U((1-\tau)y - T(y) + I, y, \theta)$, which solves the following first order condition given constant marginal tax rate $\hat{T}' = T'(y)$:

$$MRS((1-\tau)y - T(y) + I, y, \theta) = 1 - \tau - \hat{T}'.$$

Then the elasticities of y along the linear wage schedule and linear budget constraint are defined by

$$\tilde{\epsilon}^u(y) = \frac{\partial y}{\partial(1-\tau)}\bigg|_{\tau=I=0} \cdot \frac{1-T'(y)}{y}, \quad \tilde{\eta}(y) = -\frac{\partial y}{\partial I}\bigg|_{\tau=I=0} \cdot (1-T'(y)), \quad \tilde{\epsilon}^c(y) = \tilde{\epsilon}^u(y) + \tilde{\eta}(y).$$

Different from the earnings elasticities in the standard models with exogenous wage rate, our elasticities of effective labor supply y also depend on the value function. Thus, this elasticity is again the combination of elasticities of working hours and human capital investment considering the return to θ and e. As the actual taxable income data is generated under the nonlinear wage and tax schedules, this (hypothetical) elasticity cannot be directly estimated from the data.

Efficiency Test Formulas We now express the test formulas with the elasticites we defined above. We first show the Pareto efficiency test formula in terms of earnings distribution in the next proposition.

Proposition 2. *Given any Pareto efficient tax function $T(\omega)$ inducing an earnings distribution $H(\omega)$ with its density $h(\omega)$ satisfies*

$$\frac{d}{d\omega}\underbrace{\left[1 - H(\omega) - \frac{T'(\omega)}{1-T'(\omega)} \times \omega h(\omega) \times \epsilon^c_\omega(\omega) + \int_\omega^{\bar{\omega}} \frac{T'(\tilde{\omega})}{1-T'(\tilde{\omega})}\eta_\omega(\tilde{\omega})h(\tilde{\omega})d\tilde{\omega}\right]}_{\equiv R^\omega(\omega)} \leq 0. \qquad (16)$$

As long as the government does not regulate effort $e(\theta)$ and working hours $h(\theta)$ separately, the test only requires earnings elasticities.

Proof See the appendix. ■

The test formula (16) is essentially checking whether the revenue incidence associated with a tax reform of increasing marginal tax rate at ω is nonincreasing in ω.[12] If this revenue function is increasing in ω, Bierbrauer, Boyer and Hansen(2022) show that there can be a Pareto improving two-bracket tax reform of decreasing marginal tax rate at a lower bracket of income and increasing marginal tax rate at a higher bracket of income.

[12]Note that the equation inside the bracket in (16), $R^\omega(\omega)$ is exactly the revenue incidence function associated with elementary tax reform of increasing marginal tax rate at earnings level ω.

This test formula in terms of earnings distribution is the same as in the the test formula used in the standard Mirrlees tax model (e.g. Werning (2007)), which is consistent with the neutrality result in Scheuer and Werning (2017). That is, the test formula (16) is expressed just in terms the same sufficient statistics—earnings elasticity and earnings distribution, and thus the expression of the formula itself does not depend on the value function $U(c, y, \theta)$, the shape of wage schedule $W(y)$, and production function $p(\theta, e)$. We only need earnings elasticities as the government does not regulate human capital and working hours separately.[13]

However, this does mean that the set of Pareto efficient tax rates is independent of sources of inequality, as the sufficient statistics in the formula depend on inequality factor. When income inequality is increasing, not only earnings distribution but also earnings elasticities change and the response of elasticity does depend on different sources of inequality. After canceling out the counterbalancing forces of changing distribution and elasticities, in which direction the net force will change the tax rate? To show this, we now express the Pareto efficiency test formula in terms of skill distribution in the next proposition.

Proposition 3. *Given exogenous skill distribution $F(\theta)$ with its density $f(\theta)$, any Pareto efficient tax function $T(\omega(\theta))$ satisfies*

$$\frac{d}{d\theta}\underbrace{\left[\begin{array}{r} 1 - F(\theta) - \frac{T'(\omega(\theta))}{1-T'(\omega(\theta)))} \times \theta f(\theta) \times EEC(\theta) \\ + \int_{\theta}^{\bar{\theta}} \frac{T'(\omega(\tilde{\theta}))}{1-T'(\omega(\tilde{\theta}))} EEC(\tilde{\theta}) \frac{\tilde{\eta}(\tilde{\theta})}{\tilde{\epsilon}^c(\tilde{\theta})} \frac{y'(\tilde{\theta})\tilde{\theta}}{y(\tilde{\theta})} f(\tilde{\theta}) d\tilde{\theta} \end{array}\right]}_{\equiv R(\theta)} \leq 0, \quad \textit{where} \tag{17}$$

$$EEC(\theta) \quad = \quad \epsilon^c_\omega(\omega(\theta)) \times \frac{1}{\frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}}. \tag{18}$$

Proof See the appendix. ■

[13] We also remark that the neutrality result of Pareto efficiency test holds in the absence of general equilibrium effects. If there is complementarity between different levels of $y(\theta)$ or compelementarity between different levels of human capital $e(\theta)$, then there will be additional terms in the test formula capturing "Stiglitz effects." We abstract from this type of general equilibrium effect by assuming no aggregate complementarity in production. Another type of general equilibrium effects we do not consider is profit-spillover effects considered in Ales and Sleet (2016). We abstract from this effect by assuming that the government can levy optimal tax on profit (which is 100 percent tax in this economy).

From the formula (17), ignoring the income effects, we can see that the effects of rise in inequality boils down to its effects on $EEC(\theta)$ in the formula which captures the effective efficiency cost (EEC) of tax distortion. Note that the earnings distribution is given by $\omega h(\omega) = \theta f(\theta) \times \frac{1}{\frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}}$. Equation (18) in Proposition 3 shows that $\epsilon^c_\omega(\omega(\theta)) = EEC(\theta) \times \frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}$. Thus, when inequality is increasing, by canceling out the counterbalancing forces of change in earnings distribution and earnings elasticities, the effects boil down to the change in $EEC(\theta)$ given skill distribution:

$$\omega h(\omega) \times \epsilon^c_\omega = \theta f(\theta) \times EEC(\theta).$$

The fact that the effect of rising inequality boils down to the EEC is best explained by the simple top income tax rate example.

Example 4. ***(Top Income Tax)*** *Assume no income effects in labor supply ($\eta_\omega = \tilde{\eta} = 0$) and thus $\epsilon^c_\omega(\omega) = \epsilon^u_\omega(\omega) \equiv \epsilon_\omega(\omega)$. Suppose that the top skill distribution follows $Pareto(\phi)$ and these top skill types face constant top income tax rate τ_{top}. If the earnings elasticity $\epsilon_\omega(\omega)$ and earnings differential across skill $\frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}$ are locally constant at the top, then the top earnings distribution also follows $Pareto(a^{top}_\omega)$ with a Pareto parameter $a^{top}_\omega = \phi \times \frac{1}{\frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}}$ and $EEC(\theta)$ is also locally constant.*

Under these assumptions, the Pareto efficiency test formula boils down to

$$\tau_{top} \leq \tau_{ub} = \frac{1}{1 + \epsilon_\omega \times a^{top}_\omega} = \frac{1}{1 + EEC \times \phi}, \tag{19}$$

where τ_{ub} represents the upper bound of top income tax rate. The first equality in (19) shows the formula in terms of income distribution, and the second equality in (19) shows the formula in terms of skill distribution.

In this top income example, when income inequality increases, both Pareto parameter a^{top}_ω of the earnings distribution and the earnings elasticities ϵ_ω change. Rise in inequality lowers Pareto parameter a^{top}_ω and thus provides forces for higher top income tax rate τ_{top} regardless of the sources of inequality. On the other hand, the responses of earnings elasticities ϵ_ω depend on different sources of inequality.

Canceling out the couterbalancing forces through a_ω^{top} and ϵ_ω, what is the net effect of rise in inequality on the top income tax rate? Note that $a_\omega^{top} = \phi \times \frac{1}{\frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}}$ and $\epsilon_{omega} = EEC \times \frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}$. Thus, the rise in earnings differential $\frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}$ lowers Pareto parameter a_ω^{top} and raises earnings elasticity ϵ_ω, theses two effects exactly cancel out. Thus, the effects boil down to the response of the effective efficiency cost of tax distortion (EEC), which depends on the sources of inequality. In the next section, we analyze how the EEC responds to each factor of inequality in detail.

3.2 The Effects on the Effective Efficiency Cost (EEC)

Going back to the main question of the paper, we are interested in how Pareto efficient tax schedules are affected by the sources of income inequality. Our derivation of tax formula (17) in Proposition 3 and the top income tax formula (19) in the simple example show that the question boils down to figuring out how the effective efficiency cost of tax distortion responds to different sources of inequality. In this section, we analyze the effects of each inequality factor on the EEC and the the implications on the Pareto efficient tax rates.

Next proposition rewrites the EEC so that it can show the mechanisms clearly.

Proposition 5. *The effective efficiency cost of tax distortion $EEC(\theta)$ in the Pareto efficiency tax formula (17) satisfies*

$$EEC(\theta) \quad = \quad \left[\left(\frac{\tilde{\epsilon}^c(\theta)}{1+\tilde{\epsilon}^u(\theta)}\right)^{-1} \times \left(\frac{p_\theta \cdot \theta}{p} + \frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\{\tilde{e}=\frac{e}{\theta}\}}\right) + \Omega(\rho_{\theta e})\right]^{-1}. \tag{20}$$

If the Hicksian coefficient of complementarity $\rho_{\theta e}$ is 1, then $\Omega(\rho_{\theta e}) = 0$.

Proof See the appendix. ■

Equation (20) shows the expression for the EEC. For expositional simplicity, we focus on the case when the Hicksian coefficient of complementarity $\rho_{\theta e}$ is 1 and there no income effects in effective labor supply ($\tilde{\epsilon}^c = \tilde{\epsilon}^u \equiv \tilde{\epsilon}$). Then we have simpler expression for the EEC:

$$EEC(\theta) = \frac{\tilde{\epsilon}(\theta)}{1+\tilde{\epsilon}(\theta)} \times \frac{1}{\frac{p_\theta\cdot\theta}{p} + \frac{p_e\cdot e}{p} \cdot \mathbb{1}_{\{\tilde{e}=\frac{e}{\theta}\}}}. \tag{21}$$

Different from the standard Mirreesian model, the efficiency cost of tax distortion is determined by two factors. First one is the elasticity of effective labor supply $\tilde{\epsilon}$, which is is standard factor. When responsiveness of labor supply is higher, increase in marginal tax rate leads to larger decrease in labor supply, which creates higher efficiency cost. The second factor is the differential of effective labor supply $y(\theta)$ across skill type θ. With higher differential of $y(\theta)$, it is easier for the high skilled type to mimic the low skilled type, which tightens the incentive constraint. Since the tax distortion at the lower skilled type can relax the incentive constraint, the efficiency cost of tax distortion is lowered when the differential of $y(\theta)$ is higher. Higher return to skill type (i.e. higher $\frac{p_\theta \cdot \theta}{p}$) leads to higher differential which lowers EEC, but higher return to human capital (i.e. higher $\frac{p_e \cdot e}{p}$) leads to higher differential of $y(\theta)$ only when the cost of human capital decreases in skill type ($\tilde{e}(e, \theta) = \frac{e}{\theta}$).

Responses of EEC to Different Inequality Factors We now investigate how the EEC depends on each factor of inequality and its implications on the Pareto efficient tax rates. Since it is best explained by the simple top income tax rate example—Example 4—with $\rho_{\theta e} = 1$, we focus on this example.

When rise in inequality is driven by an increase in return to skill, higher $\frac{p_\theta \cdot \theta}{p}$ lowers the effective efficiency cost of tax distortion through the labor supply differential channel. Since higher skilled types gets higher return, this makes them easier to shirk by mimicking lower types, then more tax distortion helps the higher types self select their labor supply, which lower the EEC. As a result higher return to skill provides a force to increase the top income tax rate τ_{top}.

On the contrary, rise in inequality driven by an increase in return to human capital can have the opposite implications relative to the effects of higher return to skill depending on the type of human capital model and the size of the return to each factor. Consider a model where the cost of human capital investment is decreasing in skill type: $\tilde{e}(e, \theta) = \frac{e}{\theta}$. Increase in return to human capital changes the EEC via two channels. First, the labor supply differential channel lowers the EEC as in the case of higher return to skill. Higher return to human capital increases human capital investment more for the higher skilled type, and thus raises the labor supply differential across skills, leading to lower the EEC. Second, different from higher return to skill, higher return to education also increases the elasticity of effective labor supply, which raises the EEC. Thus, whether EEC increases or decreases depend on which channel dominates. If the size of return to human capital is smaller compared to the size of the return to skill, the elasticity channel dominates, thus higher return to human capital raises the EEC and lower τ_{top}. We get the same result if the human capital cost does not depend on skill because only the elasticity channel survives.

On the other hand, when the rise in return to effective labor supply is the source of higher inequality, the EEC does not change, and thus it has no effects on the Pareto efficient top income tax. This is because the counterbalancing forces on the earnings distribution and the earnings elasticity cancel out. Higher return to effective labor supply $\frac{\omega'(\theta)\cdot\theta}{\omega(\theta)}$ decreases the Pareto parameter of earnings distribution a_{ω}^{top} and increases the earnings elasticity ϵ_{ω}, and the two effects exactly offset. This result is consistent with the result of the superstar effects in Scheuer and Werning (2017).

General Complementarity between Skill and Human Capital

[to be filled in]

4 Cautions for the Sufficient Statistic Approach

One important message from the structural investigation so far is that when earnings inequality changes, not only earnings distribution but also earnings elasticities change, and the direction of the change in earnings elasticities depends on the sources of inequality. This endogeneity of the earnings elasticity can cause an issue when using a sufficient statistics approach for the optimal policy suggestion. In this section, we investigate what cautions should be taken for the sufficient statistics approach when inequality is changing significantly. We provide cautions for (i) the future comparative static and (ii) the testing Pareto efficiency of the current tax schedule.

4.1 Caution 1: Sufficient Statistics for the Future Compartive Static

The first caution is for the application of the sufficient statistic approach for the future comparative static when rise in income inequality is expected. Income inequality has been changed over time, and we expect that in the future, there will be change in technology and labor market that will increase earnings inequality in general. When this change in inequality is expected, can we apply the sufficient statistic approach to make a policy suggestion for the future? In our structural investigation, we have learned that earnings elasticities can be changed and we need to predict the sources of rise in inequality in the future to make a right prediction of the earnings elasticities. In this section, we show a typical application of sufficient statistic approach for the future ignoring the sources of inequality can lead to a wrong policy suggestion.

We use the simple top income tax rate example—Example 4—for expositional simplicity. We also assume that the Hicksian coefficient of complementarity $\rho_{\theta e}$ is 1. Suppose that we expect 10% decrease in Pareto parameter in earnings distribution due to the rise in inequality. Suppose that we want to make a policy suggestion for the future using a sufficient statistics formula (19). The usual application of the sufficient statistic approach would evaluate the formula with the predicted Pareto parameter $a_\omega^F = 0.9 \times a_\omega^{US}$ where the a_ω^{US} is the current Pareto parameter of the observed earnings distribution. For the earnings elasticity ϵ_ω, however, without knowing how the elasticity will change, a typical approach would be to use the estimated earnings elasticities from the current or past taxable income data. How wrong can it be with this typical approach?

One immediate answer is that this approach is not problematic if the elasticities of earnings do not change despite the increase in earnings inequality. Our structural investigation this can be the case when the expected rise in inequality is driven by the increase in return to skill. Next corollary shows this.

Corollary 6. *Suppose that there is no income effect in effective labor supply with $\eta_\omega = \tilde{\eta} = 0$ and the Hicksian coefficient of complementarity $\rho_{\theta e}$ is 1. Then the earnings elasticity $\epsilon_\omega(\omega(\theta))$*

satisfies following:

$$\epsilon_\omega(\omega(\theta)) = \tilde{\epsilon}(\theta) \times \frac{1}{\frac{y'(\theta)\cdot\theta}{y(\theta)}} \times \frac{\omega'(\theta)\cdot\theta}{\omega(\theta)} = \tilde{\epsilon}(\theta) \times \frac{W'(y(\theta))\cdot y(\theta)}{W(y(\theta))}. \tag{22}$$

From this corollary, we can see that return to skill does not show up in the earnings elasticity expression (22), and thus there is no direct effect of return to skill on the earnings elasticity. Of course, if the elasticity of labor supply $\tilde{\epsilon}(\theta)$ and the return to effective labor supply $\frac{W'(y)y}{W(y)}$ depends on where they are evaluated, there can be indirect change in ϵ_ω due to the change in level of effective labor supply $y(\theta)$. Below, we show the example where return to skill has no effects on the earnings elasticity.

On the other hand, if the expected rise in inequality is because of the rise in either return to human capital or return to effective labor supply, then they have direct effects on the earnings elasticities. In the expression of earnings elasticity in (22), the rise in return to human capital increases the elasticity of effective labor supply $\tilde{\epsilon}(\theta)$ and the rise in return to effective labor supply directly raises $\frac{W'(y)y}{W(y)}$. Below we have a numerical investigation to show how wrong can the typical sufficient statistic approach be when ignoring these changes.

4.1.1 Numerical Example: Calibration

We now investigate a numerical example to show the relationship between the Pareto efficient top income tax rate and sources of inequality. To capture the return to each inequality factor with one parameter each, we consider simple functional forms for the wage schedule and productivity function. The wage schedule has a power functional form $W(y) = y^\kappa$, where the return to effective labor supply $\frac{W'(y)\cdot y}{W(y)}$ is simply determined by a parameter κ.[14] We assume that the productivity function takes a form of $p(\theta, e) = \theta^\gamma \cdot e^\nu$. With this Cobb-Douglas functional form, the Hicksian coefficient of complementarity is one. Note also that the return to skill $\frac{p_\theta\cdot\theta}{p}$ is equal to a parameter γ and the return to human capital $\frac{p_e\cdot e}{p}$ is equal to a parameter ν. Then a rise inequality can be generated by an increase in κ, γ, or ν, which represents each inequality factor.

[14] In section 2, we have endogenized $W(y)$ with a firm's problem given technology function $G(x, y)$. Here, we directly assume the functional form of $W(y)$, but we can invert out the technology G that can induce $W(y) = y^\kappa$.

We assume the following preferences with no income effects in effective labor supply:

$$u(c, \tilde{e}, n) = c - \zeta \cdot \tilde{e}^{\frac{1}{\zeta}} - \sigma \cdot n^{\cdot\frac{1}{\sigma}}, \tag{23}$$

where ζ controls the elasticity of human capital investment and σ controls the elasticity of working hours. We consider the following two types of human capital models:

$$\tilde{e}(e, \theta) = \begin{cases} e, & \text{HK-A} \\ \frac{e}{\theta}, & \text{HK-B.} \end{cases}$$

In the benchmark, we assume that the cost of human capital investment is decreasing in skill by taking the HK-B model.

Although we will focus on the Pareto efficiency of the top income tax rate, to calibrate all the necessary parameters, we need to set the distribution of the entire skill types. We assume that log skills are drawn from an exponentially modified Gaussian (EMG) distribution in the bottom 90% and the top 10% skill distribution follows Pareto distribution. More precisely, we assume that log of the unobservable skill has two components: $\log \theta = \epsilon_N + \epsilon_E$, where $\epsilon_N \sim N(\mu_\epsilon, \sigma_\epsilon^2)$ and $\epsilon_E \sim exp(\lambda_\epsilon)$ so that $\log \theta \sim EMG(\mu_\epsilon, \sigma_\epsilon^2, \lambda_\epsilon)$. Then the level skill distribution is Pareto log normal. We then modify the top 10% skill distribution so that $\theta \sim Pareto(\phi)$ for $\theta \geq \hat{\theta}$, where $\hat{\theta}$ is the top ten percentile skill level.

For a calibration purposes, we also need to set the current US tax schedule. We approximate the current tax schedule using the parametric tax function adopted in Heathcote, Storesletten, and Violante (2017) (hereafter HSV), where taxes net of transfers are given by the following function of earnings: $T(\omega) = \omega - \lambda^{US} \cdot \omega^{1-\tau^{US}}$. We assume that at the top earnings distribution, top income tax rate is constant.

Under these functional form assumptions, we have a closed for solution for human capital, effective labor supply, and earnings (in the HK-B case):

$$\begin{aligned} e(\theta) &= \left[\lambda^{US} \cdot (1 - \tau^{US}) \cdot \kappa\right]^{\frac{\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \theta^{\frac{1-(1-\tau^{US})\kappa(\sigma-\gamma\zeta)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \nu^{\frac{\zeta(1-(1-\tau^{US})\kappa\sigma)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}}, && (24) \\ y(\theta) &= \left[\lambda^{US} \cdot (1 - \tau^{US}) \cdot \kappa\right]^{\frac{\sigma+\nu\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \theta^{\frac{\gamma+\nu}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \nu^{\frac{\nu\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}}, && (25) \\ \omega(\theta) &= \left[\lambda^{US} \cdot (1 - \tau^{US}) \cdot \kappa\right]^{\frac{\kappa(\sigma+\nu\zeta)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \theta^{\frac{\kappa(\gamma+\nu)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \nu^{\frac{\kappa\nu\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}}. && (26) \end{aligned}$$

The closed form expression for the earnings shows that if $\log\theta$ is drawn from an EMG distribution for $\theta < \hat{\theta}$, $\log\omega(\theta)$ also follows the EMG distribution for $\omega < \omega(\hat{\theta})$. It also shows that if distribution of θ is Pareto for $\theta \geq \hat{\theta}$, then distribution of $\omega(\theta)$ is also Pareto for $\omega \geq \omega(\hat{\theta})$.

In the benchmark calibration, we choose $\sigma = 0.2$ so that the Frisch elasticity ($\frac{\sigma}{1-\sigma}$) is 0.25. The progressivity parameter τ^{US} of the HSV tax function is set 0.181 adopting the estimation of Heathcote, Storesletten, and Violante (2017).

For the calibration of the skill distribution, we take the strategy similar to the one in Heathcote and Tsujiyama (2021). We calibrate the parameters of the skill distributions to match the empirical earnings distribution. In our model, the distribution of log earnings is $EMG(\mu_\omega, \sigma_\omega^2, \lambda_\omega)$ distribution with

$$\sigma_\omega^2 = \left(\frac{\kappa(\gamma+\nu)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}\right)^2 \times \sigma_\epsilon^2, \qquad \lambda_\omega = \left(\frac{\kappa(\sigma+\nu\zeta)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}\right)^{-1} \times \lambda_\epsilon.$$

Thus, given the parameters for the preferences (σ, ζ) and the parameters for the wage structure (γ, ν, κ) we can infer the parameters $(\sigma_\epsilon^2, \lambda_\epsilon)$ which matches the estimated parameters $(\sigma_\omega^2, \lambda_\omega)$. We use the estimates of Heathcote and Tsujiyama (2021): $\sigma_\omega^2 = 0.412$ and $\lambda_\omega = 2.2$.[15] Similarly, the distribution of top earnings under the constant top income tax rate is $Pareto(a_\omega^{top})$ distribution with

$$a_\omega^{top} = \left(\frac{\kappa(\gamma+\nu)}{1-\kappa(\sigma+\nu\zeta)}\right)^{-1} \times \phi.$$

We infer Pareto parameter of the skill distribution ϕ to match $a_\omega^{top} = 1.7$ in the benchmark. This benchmark a_ω^{top} is somewhat larger than the numbers measured by the Adjusted Gross Income (AGI) data. Since the AGI includes capital gains which are not relevant for our earning tax, we use 1.7, which is calculated by Badel, Huggett, and Luo (2020) by excluding dividends and long-term capital gain. We do sensitivity analysis for lower a_ω^{top} in the appendix.

We calibrate the return to skill parameter γ using the Mincerian wage regression. In the benchmark calibration, we assume that the proxy of the skill variable θ in our model is the

[15] Heathcote and Tsujiyama (2021) estimates these parameters using the Survey of Consumer Finances data to include enough high income households.

cognitive skill, and use the regression coefficient of AFQT test score from the NSLY data. In the benchmark use the result from the estimation without controlling for schooling to avoid the endogeneity issue related to the schooling variable, and we set $\gamma = 0.76$. More detail is explained in the appendix. In the empirical studies, how to estimate the return to unobserved skill is very controversial. I discuss this in more detail below, and provide some sensitivity analysis.

We then jointly calibratethe remaining parameters (ζ, ν, κ) to match (i) the change in average education in response to increase in return to education, (ii) college wage premium, and (iii) Pareto parameter of earnings distribution. Given (ν, κ), the elasticity of human capital investment ζ is calibrated to match average schooling response. Note that from equation (24) $E[e_t] = \Psi_t \cdot \nu_t^{A_t(\zeta)}$, where $A_t(\zeta) = \frac{\zeta(1-(1-\tau_t^{US})\kappa_t\sigma)}{1-(1-\tau_t^{US})\kappa_t(\sigma+\nu_t\zeta)}$. If Ψ_t and A_t are constant over time, $A(\zeta)$ should satisfy

$$A(\zeta) = \frac{\log\left(E[e_{2016}]/E[e_{1980}]\right)}{\log\left(\nu_{2016}/\nu_{1980}\right)}$$

We compute $A(\zeta)$ using the calculation in Heathcote, Storesletten, and Violante (2020) who measure e as years of education above mandatory schooling.[16] We then infer ζ given (ν, κ, τ^{US}).

Given κ, the return to human capital prameter ν is determined to mainly match the college premium 1.9. We use the calculation of Heathcote, Perri, Violante, and Zhang (2023) who define the average hourly wage of workers with at least 16 years of schooling (also in Heathcote, Perri, and Violante (2010)). Since we do not have years of schooling in our model, we redefine the college graduates as the workers with top 37.7% of human capital following the empirical fraction of college graduates (defined as more that 16 years of schooling) in Heathcote, Perri, Violante, and Zhang (2023). Then the parameter for the return to effective labor supply κ is determined to mainly match the top income inequality measured by the Pareto parameter $a_\omega^{top} = 1.7$. The resulting parameters are $\zeta = 0.46$, $\nu = 0.20$, and $\kappa = 1.52$.

[16] ν_t can be obtained from a Mincerian regression coefficient r_t for each year. ($\nu_t = \frac{d \log w_t}{d e_t} e_t$).

Table 1: Benchmark: Pareto Efficient Top Income Tax Rate

$a_\omega^{US} = 1.7$	$a_\omega^F = 1.53$ (−10%)			
$\tau_{ub}^{US} = \frac{1}{1+a_\omega^{US}\times\epsilon_\omega^{US}}$	$\tau_{ub}^{\text{suff}} = \frac{1}{1+a_\omega^{F}\times\epsilon_\omega^{US}}$	$\tau_{ub}^{F} = \frac{1}{1+a_\omega^{F}\times\epsilon_\omega^{F}}$		ϵ_ω^F
0.425	**0.451** (+6.1%)	$\gamma \uparrow$	0.451 (+6.1%)	0.80 ($=\epsilon_\omega^{US}$)
		$\nu \uparrow$	**0.419** (−1.3%)	0.90 (+13.7%)
		$\kappa \uparrow$	**0.425** (+0.0%)	0.88 (+11.1%)

4.1.2 Numerical Result: Benchmark

In this calibrated economy, how wrong can the typical sufficient statistic approach be if we ignore the sources of inequality? To answer this we consider a situation where we expect rise in income inequality in the future, as a result, we expect that the Pareto parameter of top earnings distribution a_ω^{US} will decrease by 10%. This 10% decrease in Pareto parameter can be generated by three different sources: (i) increase in return to skill γ, (ii) increase in return to human capital ν, and (iii) in crease in return to effective labor supply κ. We denote the upper bound of the Pareto efficient top tax rate for the future by

$$\tau_{ub}^F = \frac{1}{1+\epsilon_\omega^F \times a_\omega^F}, \quad \text{where } a_\omega^F = 0.9 \times a_\omega^{US}.$$

The future elasticity ϵ_ω^F depends on the sources inequality, and thus the upper bound of top income tax rate is different for each source.

On the other hand, we denote the upper bound of the Pareto efficient top income tax rate following the usual sufficient statistic approach by τ_{ub}^{suff}, which is computed by

$$\tau_{ub}^{\text{suff}} = \frac{1}{1+\epsilon_\omega^{US} \times a_\omega^F}.$$

In this typical application, although the distribution parameter is evaluated with the predicted value for the future, the the earnings elasticity estimated from the current data is use ignoring the exact sources of inequality for the rise in inequality.

The typical sufficient statistic approach suggest that the upper bound τ_{ub}^{suff} should increase relative to the current upper bound of Pareto efficient top income tax rate, as we only take into account decrease in the Pareto parameter a_{ω}^{top} when income inequality rises. The first two columns of Table 1 shows that the upper bound τ_{ub}^{suff} based on the sufficient statistic increases by 6.1% relative to the current upper bound tax rate τ_{ub}^{US} ignoring where the source is coming from. The third column of Table 1 presents the correct prediction of upper bound tax rate τ_{ub}^{F} based on each source of inequality. As we already discussed above, the sufficient statistic approach is correct only when the return to skill—rise in γ—is the source of rise in inequality.

When the higher return to human capital—higher ν—is the only source of rise in inequality, however, Table 1 shows that the correct prediction of upper bound tax rate τ_{ub}^{F} should decrease by 1.3%, implying that the suggestion based on a typical sufficient approach is not only incorrect but also leads to the wrong direction. This decrease in top income tax rate is because the elasticity channel dominates. As the fourth column of Table 1 shows, when with higher return to human capital, the earnings elasticity increases ϵ_{ω}^{F} by 13.7% and this dominates the earnings distribution channel (10% decrease of a_{ω}^{F}). As the typical sufficient statistic ignores this dominant forces, the policy suggestion goes to the wrong direction.

Another way of understanding the case with higher ν is to look at the formula in terms of skill distribution ($\tau^{top} \leq \frac{1}{1+EEC\times\phi}$) after canceling out the counterbalancing forces through a_{ω}^{F} and through ϵ_{ω}^{F}. Our numerical result shows that the effective efficiency cost of tax distortion (EEC) increases with higher ν. Recall that as we discussed in section 3.2, the expression of EEC in (21) shows two channels of changing the EEC—elasticity channel and labor supply differential channel. When ν increases, the elasticity channel dominates if the role of return to human capital ν is relatively smaller than the role of return to skill γ in the labor supply differential. Our calibration of γ in the benchmark is relatively high, and thus EEC increases in ν. We discuss this in more detail in the sensitivity analysis in Section 4.1.3.

On the other hand, when the higher return to effective labor supply—higher κ—is the only source of rise in inequality, Table 1 shows that the correct prediction of upper bound tax rate τ_{ub}^{F} is equal to current upper bound tax rate τ_{ub}^{US}, which is different from the prediction of 6.1% increase in tax rate based on the wrong application of the sufficient static approach.

Figure 1: Pareto efficient upper bound

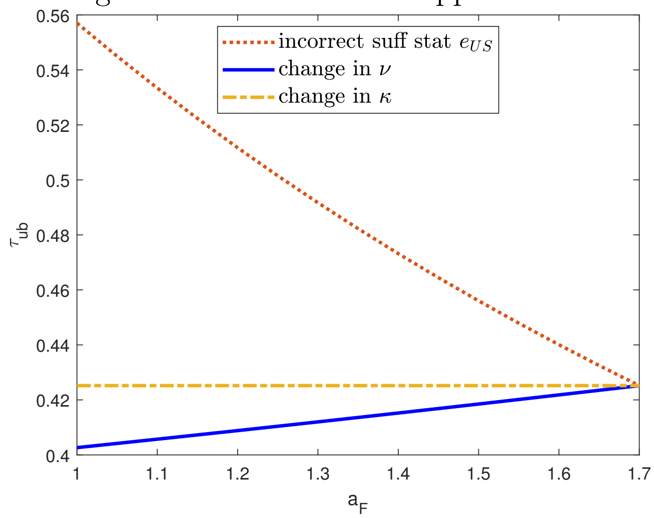
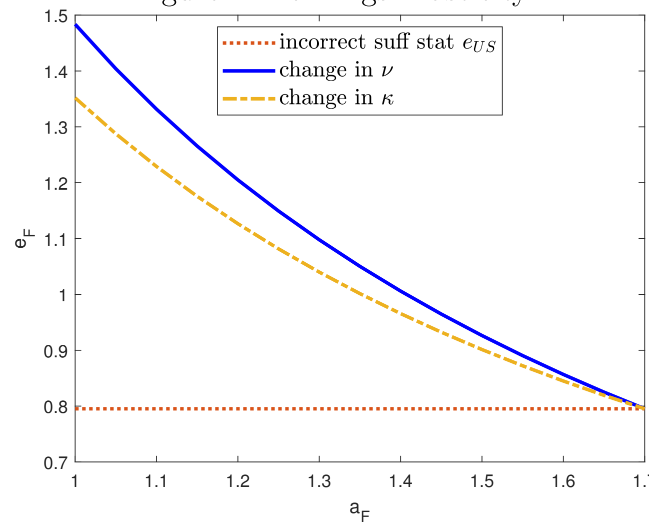


Figure 2: Earnings Elasticity



This no change in Pareto efficient tax rate is because the 10% decrease in Pareto parameter ϵ_ω^F is exactly offset by the 11.1% $(=(\frac{1}{0.9}-1)\times 100\%)$ increase in earnings elasticity, confirming the result from the structural investigation in section 3. As a result, there is no change in the EEC, and upper bound tax rate does not change.

Figure 1 shows how wrong the typical application of sufficient statistic approach could be for each level of future inequality represented by the Pareto parameter a_ω^F. As the change in inequality is larger, there will be even stronger errors in the predictions if the sources of inequality is not return to skill. Figure 1 shows that the difference between τ_{ub}^{suff} and τ_{ub}^F can be more than 10% point. This misleading policy suggestion is because significant increase in earnings elasticity driven by change in ν or κ is ignored in the usual sufficient statistic approach, which is clearly observed in Figure fig:elasticity1.

4.1.3 Numerical Result: Sensitivity

In this section, we carry out sensitivity analysis for the numerical investigation we provided in Section 4.1.2. Here we investigate alternative calibration of return to skill γ and alternative human capital model (HK-A). In the appendix, we provide further sensitivity analysis with respect to calibration of elasticity of human capital investment ζ, Pareto parameter of earnings distribution a_ω^{US}, and value of the elasticity of working hours σ.

Alternative Calibration for γ We first consider alternative calibration of return to skill γ. From the analysis so far, we have learned that the change in return to skill γ does not cause any problem of using sufficient statistics approach, but the level of γ is important for the effects of change in return to skill ν on the Pareto efficient tax rate. This is because when EEC is changed by the rise in ν, the level of γ is crucial for determining which channel dominates between the elasticity channel and and labor supply differential channel. In the next proposition, we show this more formally.

Proposition 7. *Consider an economy in Example 4. Suppose that preferences of worker has the form of (23) and productivity of worker has the form: $p(\theta, e) = \theta^{\gamma} \cdot e^{\nu}$.*

1. *If $\gamma > \frac{\sigma}{\zeta}$, then the EEC is increasing in ν.*
2. *If $\gamma < \frac{\sigma}{\zeta}$, then the EEC is decreasing in ν.*

In the benchmark numerical investigation, our calibration of γ satisfies the first case in Proposition 7, and thus higher return to human capital ν has led to higher EEC and lower Pareto efficient tax rate. As we have already discussed in Section 4.1.1, measuring the return to unobserved skill is very difficult task in the empirical literature. In the benchmark, we used the Mincerian wage regression without controlling schooling variables to get the return to skill, but in this regression, the coefficient of the cognitive skill tends to be over-estimated because it includes both direct effect of skill and indirect effect of skill through education. We now consider an alternative calibration of γ by using the Mincerian wage regression controlling schooling variables. In this alternative calibration, $\gamma = 0.243$, and it satisfoes the second case in Proposition 7.

Figure 3 shows that with this alternative calibration which leads to lower level of return to skill γ, the implication of higher ν is changed from the result in benchmark (Figure 1). When rise in income inequality is driven by higher return to human capital, the upper bound of Pareto efficient top income tax rate τ_{ub}^F increases as the decrease in Pareto parameter a_{ω}^F dominates the increase in earnings elasticity ϵ_{ω}^F. Thus, in this alternative calibration, the policy suggestion based on the typical sufficient statistics has the same directional implication with the correct suggestion implied by τ_{ub}^F. Figure 3, however, shows that the difference

Figure 3: Pareto efficient upper bound: low γ

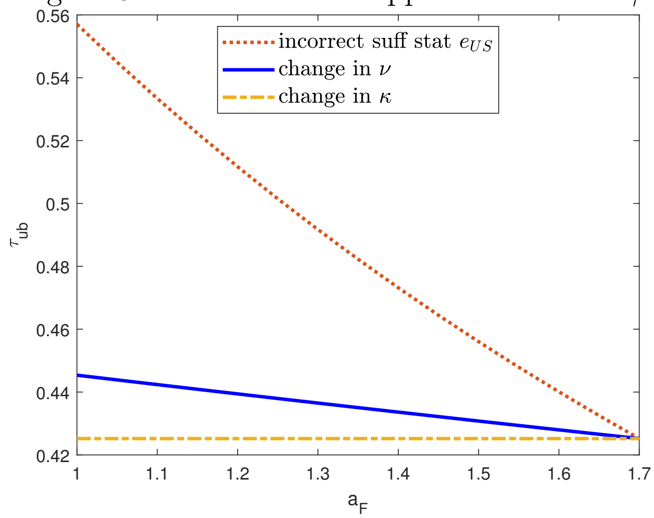
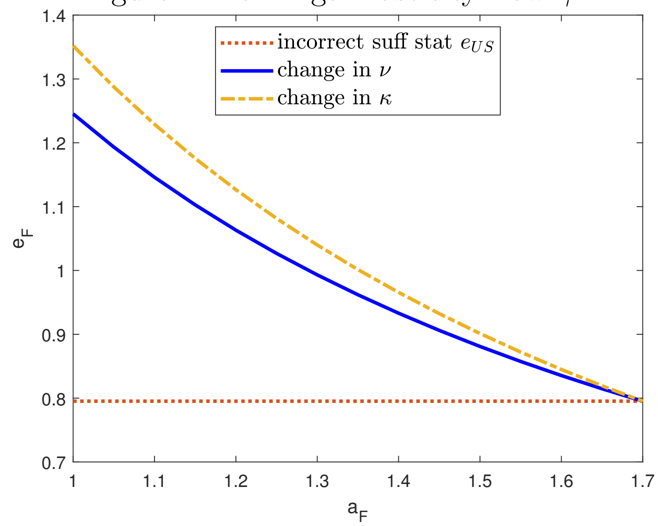


Figure 4: Earnings Elasticity: low γ



between τ_{ub}^{suff} and τ_{ub}^{F} can be still very large because the increase in earnings elasticity (in Figure 4) is completely ignored in the sufficient statistics.

[More discussion on the calibration of γ: to be filled]

Alternative Human Capital Model We now consider an alternative human capital model. In the benchmark, we took a version of human capital model (HK-B) where the cost of human capital investment is decreasing in skill type by setting $\tilde{e}(e,\theta) = \frac{e}{\theta}$. We now take an alternative human capital model (HK-A) where the cost of human capital investment is independent of skill type with $\tilde{e}(e,\theta) = e$.

With the special form of preferences and productivity function we took, the elasticity of earnings does not depend on the type of human capital model:

$$\epsilon_\omega(\omega(\theta)) = EEC \times \frac{\omega'(\theta)\cdot\theta}{\omega(\theta)} = \frac{\sigma + \nu\zeta}{\gamma + \nu \times \mathbb{1}_{HK-B}} \times \frac{\kappa \times (\gamma + \nu \times \mathbb{1}_{HK-B})}{1 - \kappa(\sigma + \nu\zeta)} = \frac{\kappa(\sigma+\nu\zeta)}{1-\kappa(\sigma+\nu\zeta)}.$$

As a result, both types of human capital models have the same current level of the earnings elasticities ϵ_ω^{US}. This implies that both economies have the same upper bound of current top income tax τ_{ub}^{US}, upper bound based on the suffcienent statistic τ_{ub}^{suff}. Since we know that future upper bound associated with γ is $\tau_{ub}^{F} = \tau_{ub}^{\text{suff}}$ and the one associated with κ is $\tau_{ub}^{F} = \tau_{ub}^{US}$, they are the same in both economies. Thus we only need to look at the case when the source of inequality if ν.

Table 2: Human Capital Model Comparison: Pareto Efficient Top Tax

$a_\omega^{US} = 1.7$	$a_\omega^{F} = 1.53 \quad (\nu \uparrow)$			
$\tau_{ub}^{US} = \frac{1}{1+a_\omega^{US} \times \epsilon_\omega^{US}}$	$\tau_{ub}^{\text{suff}} = \frac{1}{1+a_\omega^{F} \times \epsilon_\omega^{US}}$	$\tau_{ub}^{F} = \frac{1}{1+a_\omega^{F} \times \epsilon_\omega^{F}}$		ϵ_ω^{F}
0.425	**0.451** (+6.1%)	**HK-A**	**0.397** (−6.7%)	0.99 (+25.1%)
		HK-B	**0.419** (−1.3%)	0.90 (+13.7%)

Table 2 shows that the in an alternative human capital model (HK-A) where the cost of human capital investment does not depend on skill type, higher return to human capital ν leads to even more significant decrease of top income tax rate compared to the decrease in the benchmark case. That is, the error of the typical sufficient statistics approach can be even more severe in an alternative human capital model economy. As we discussed above, among the two channels of higher ν affecting the effective efficiency cost (EEC) of tax distortion, the labor supply differential channel does not exist in the HK-A model, and thus higher elasticity of labor supply always increases the EEC.

5 Caution 2: Sufficient Statistics for Testing the Current Tax

We now provide cautions for the application of the sufficient statistics approach for testing the Pareto efficiency of the current tax schedule given earnings distribution. Recall the formula (16) in Section 3.1. The formula is expressed in terms of current marginal tax rate, observed earnings distribution, and current elasticities of earning. Thus, as long as we measure the current earnings elasticities correctly, there is no problem of using the sufficient statistic approach for the Pareto efficiency test. Given the right measure of earnings elasticities, we don't have to know the sources of inequality. In that sense, the sources of inequality are neutral for this Pareto efficiency test formula.

The lesson from the structural investigation, however, gives us an important caution message that the estimation of earnings elasticities should be carried out under the factors generating current income inequality. Since the earnings ealsticities are endogenous to each inequality factor, if the inequality factors during the period of the elasticity estimation are

Table 3: Change in Sources of Inequality and Adjustment

ϵ_{ω}^{e}	$\epsilon_{\omega}^{\text{true}}$		τ_{ub}^{suff}	τ_{ub}^{true}	
0.80	$\gamma \uparrow$ (+10%)	0.80	0.425	$\gamma \uparrow$	0.425
	$\nu \uparrow$ (+10%)	0.84		$\nu \uparrow$	0.411
	$\kappa \uparrow$ (+10%)	0.95		$\kappa \uparrow$	0.382

not consistent with the factors generating current earnings distribution, the test results can be misleading.

Thus, the essential caution message is very similar to the caution in Section 4. We need to be careful about whether the estimated earnings elasticity is the right sufficient statistic that can be used to evaluate the Pareto efficiency test formula. If there has been change in inequality between the estimation period and current, then adjustment of estimated elasticities are required. What is slightly different from the case of the future comparative static is that for testing the current tax, (i) whether the adjustment is needed or not depends on when exactly the earnings elasticity was estimated, and (ii) the adjustment can be made at least based on the data by comparing the data of the past and current to infer change in sources of inequality. For the future, however, we need prediction of the future elasticities in any case because the future is not observed, and the adjustment of elasticities can only be made based on the prediction of change in sources of inequality.

To show how much adjustment of elasticity we need when there is inequality gap between the estimation period and current, we do the following hypothetical exercise. Suppose that we have estimated the elasticity in the past, and compared to the period of estimation there has been change in sources of inequality. For 10% increase of each return, how much adjustment do we need make? we use the same numerical example and the benchmark calibration as in Section 4.

Table 3 shows the numerical adjustment required in the benchmark calibration. It confirms that when the return to skill has been increased from the estimation period, we do not have to make any adjustment for the earnings elasticities. We can use the estimated elasticity with no problem despite change in inequality. Since the observed earnings distribution is already given, the typical sufficient statistic approach works well without any adjustment of sufficient statistics.

When either return to human capital ν or return to effective labor supply κ increases, the earnings elasticity should be adjusted upward. The degree of adjustment depends on each source. Table 3 shows that for the same 10% increase of the return, the adjustment is much larger in case of rise in return to effective labor supply κ. Return to effective labor supply essentially increase return to total labor combining the human capital and working hour together, and thus both human capital investment and working hours do respond to rise in κ, and thus it has stronger effects on the earnings elasticities.

6 Conclusion

In this paper, we have investigated the role of the sources of inequality on the Pareto efficient tax schedule. The key message from our paper is that the earnings elasticity is endogenous to the sources of inequality and its direction of change does depend on each source. Thus higher inequality can lead to either higher or lower tax rates depending on its sources. This endogenous elasticity also provide cautions for the application of sufficient static approach during the period inequality is significantly changing. We need to make proper adjust of of earnings elasticities if the estimated elasticity is not estimated under the factors that generate the earnings distribution.

References

Ales, L., M. Kurnaz, and C. Sleet (2015): “Technical Change, Wage Inequality, and Taxes,” *American Economic Review*, 105(10), 3061–3101.

Ales, L., and C. Sleet (2016): “Taxing Top CEO Incomes,” *American Economic Review*, 106(11), 3331–3366.

Badel, A., M. Huggett, and W. Luo (2020): “Taxing Top Earners: a Human Capital Perspective,” *The Economic Journal*, 130(629), 1200–1225.

Bierbrauer, F. J., P. C. Boyer, and E. Hansen (2022): “Pareto-improving tax reforms and the Earned Income Tax Credit,” *Working Paper*.

Blomquist, S., and H. Selin (2010): “Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates,” *Journal of Public Economics*, 94(11), 878–889.

Gruber, J., and E. Saez (2002): “The elasticity of taxable income: evidence and implications,” *Journal of Public Economics*, 84(1), 1–32.

Heathcote, J., F. Perri, and G. L. Violante (2010): “Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006,” *Review of Economic Dynamics*, 13(1), 15–51.

Heathcote, J., F. Perri, G. L. Violante, and L. Zhang (2023): “More unequal we stand? Inequality dynamics in the United States, 1967–2021,” *Review of Economic Dynamics*, 50, 235–266.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): “Optimal Tax Progressivity: An Analytical Framework*,” *The Quarterly Journal of Economics*, 132(4), 1693–1754.

——— (2020): “How Should Tax Progressivity Respond to Rising Income Inequality?,” *Journal of the European Economic Association*, 18(6), 2715–2754.

Heathcote, J., and H. Tsujiyama (2021): “Optimal Income Taxation: Mirrlees Meets Ramsey,” *Journal of Political Economy*, 129(11), 3141–3184.

Hosseini, R., and A. Shourideh (2019): “Retirement Financing: An Optimal Reform Approach,” *Econometrica*, 87(4), 1205–1265.

Lochner, L., Y. Park, Y. Shin, and B. I. S. Ninth (2018): “Wage Dynamics and Returns to Unobserved Skill,” *Working Paper*.

Mirrlees, J. A. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *The Review of Economic Studies*, 38(2), 175–208.

Saez, E., J. Slemrod, and S. H. Giertz (2012): “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Literature*, 50(1), 3–50.

Scheuer, F., and J. Slemrod (2020): “Taxation and the Superrich,” *Annual Review of Economics*, 12(1), 189–211.

Scheuer, F., and I. Werning (2017): “The Taxation of Superstars,” *The Quarterly Journal of Economics*, 132(1), 211–270.

Stiglitz, J. E. (1982): “Self-selection and Pareto efficient taxation,” *Journal of Public Economics*, 17(2), 213–240.

Werning, I. (2007): “Pareto efficient income taxation,” Discussion paper.

Wu, C. (2021): “More unequal income but less progressive taxation,” *Journal of Monetary Economics*, 117, 949–968.

Appendix (For Online Publication)

A Proofs of Equations in Section 2

We first derive the first order conditions of the planning problem in a genera framework provided in section 2. We present the planning problem (9) of a general framework and Lagrangian of this planning problem is given by

$$\begin{aligned} L &= \int \lambda(\theta)v(\theta)dF(\theta) + \eta \int \left(G(\theta, y(\theta)) - C(v(\theta), y(\theta), \theta)\right) dF(\theta) \\ &\quad + \int \mu(\theta)v'(\theta)d\theta - \int \mu(\theta)U_\theta \left(C(v(\theta), y(\theta), \theta)\right) d\theta \\ &= \int \lambda(\theta)v(\theta)dF(\theta) + \eta \int \left(G(\theta, y(\theta)) - C(v(\theta), y(\theta), \theta)\right) dF(\theta) \\ &\quad - \int \mu'(\theta)v(\theta)d\theta - \int \mu(\theta)U_\theta \left(C(v(\theta), y(\theta), \theta)\right) d\theta, \end{aligned}$$

where the second equality is by integral by parts and using $\mu(\bar{\theta}) = \mu(\underline{\theta}) = 0$. The first order conditions of this Lagrangian with respect to $v(\theta)$ and $y(\theta)$ yields (12) and (13) in the main text.

We now express the optimality conditions in terms of marginal rate of substitution. First of all, we obtain the optimality condition from a workers' problem. Given the value function $U(c, y, \theta)$ associated with sub-problem (2), workers with skill level θ chooses the optimal level of earnings ω by solving

$$\max_{\omega} U(\omega - T(\omega), W^{-1}(\omega), \theta)$$

The first order condition of this worker's problem yields:

$$MRS(c(\theta), y(\theta), \theta) = (1 - T'(\omega))W'(y(\theta)), \tag{27}$$

where $MRS(c, y, \theta) = -\frac{U_y(c,y,\theta)}{U_c(c,y,\theta)}$.

From now on, when it does not cause confusion, we will omit the explicit arguments (c, y, θ) and simply write the functions as function of θ. Define $\hat{\mu}(\theta) = \frac{\mu(\theta)U_c(\theta)}{\eta}$. Then the first order condition of the planning problem (13) can be rewritten as

$$\frac{G_y(\theta) - C_y(\theta)}{MRS(\theta)} = \hat{\mu}(\theta) \cdot \frac{U_{\theta c}(\theta)C_y(\theta) + U_{\theta y}(\theta)}{U_c(\theta)MRS(\theta)}.$$

Note that the derivative of expenditure function C with respect to y is:

$$C_y(\theta) = -\frac{U_y(\theta)}{U_c(\theta)} = MRS(\theta) = W'(y(\theta))\left(1 - T'(\omega(\theta))\right).$$

Then by combining this with firm's optimality condition (5), we obtain

$$\frac{T'(\omega)}{1 - T'(\omega)} = \frac{G_y(\theta) - C_y(\theta)(\theta)}{MRS(\theta)}.$$

Also,

$$\frac{U_{\theta c}(\theta)C_y(\theta) + U_{\theta y}(\theta)}{U_c(\theta)MRS(\theta)} = \frac{-U_{\theta c}(\theta)\frac{U_y(\theta)}{U_c(\theta)^2} + \frac{U_{\theta y}(\theta)}{U_c(\theta)}}{MRS(\theta)} = -\frac{\frac{\partial MRS}{\partial \theta}}{MRS(\theta)},$$

so the first-order condition with respect to y becomes

$$\hat{\mu}(\theta) = \frac{T'(\omega(\theta))}{1 - T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta \cdot \theta}{MRS(\theta)}\right)^{-1}. \tag{28}$$

The first-order condition of the planning problem with respect to $v(\theta)$ (12) is rewritten as (using $C_v(\theta) = \frac{1}{U_c(\theta)}$)

$$-\frac{\mu'(\theta)U_c(\theta)}{\eta} - \frac{\mu(\theta)U_{\theta c}(\theta)}{\eta} = f(\theta) - \frac{\lambda(\theta)U_c(\theta)}{\eta}.$$

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

$$\begin{aligned} -\frac{\mu'(\theta)U_c(\theta)}{\eta} - \frac{\mu(\theta)U_{c\theta}(\theta)}{\eta} &= -\hat{\mu}'(\theta) + \hat{\mu}(\theta)\frac{\left[U_{cc}(\theta)MRS(\theta) + U_{cy}(\theta)\right]y'(\theta)}{U_c(\theta)} \\ &= -\hat{\mu}'(\theta) + \hat{\mu}(\theta)\cdot(-MRC_c)\cdot y'(\theta), \end{aligned}$$

where the second inequality uses $-\frac{\partial MRS}{\partial c} = \frac{\partial \frac{U_y}{U_c}}{\partial c} = \frac{U_{cc}(-\frac{U_y}{U_c}) + U_{cy}}{U_c} = \frac{U_{cc}MRS + U_{cy}}{U_c}$. By replacing the left-hand side of the first-order condition with respect to v, we obtain

$$-\hat{\mu}'(\theta) - \hat{\mu}(\theta)\cdot MRS_c \cdot y'(\theta) = f(\theta) - \frac{\lambda(\theta)U_c(\theta)}{\eta}. \tag{29}$$

B Proof of Propositions in Section 3 and Section ??

Test Formula in terms of Skill Distribution and MRS First, we obtain the Pareto efficiency test formula in terms of skill distribution and marginal rate of substitutions. This formula will be used to derive the formulas in Section 3 and Section **??**.

Combining (14) with $\lambda(\theta) \geq 0$ yields

$$-\frac{\hat{\mu}'(\theta)\theta}{\hat{\mu}(\theta)} - MRS_c \cdot y'(\theta) \cdot \theta \leq \frac{\theta f(\theta)}{\hat{\mu}(\theta)}$$

We then replace $\frac{\theta f(\theta)}{\hat{\mu}(\theta)}$ using (15) to obtain

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\left(-\frac{MRS_\theta \cdot \theta}{MRS(\theta)}\right)^{-1}\left[-\frac{d\log\hat{\mu}(\theta)}{d\log\theta} - MRC_c(\theta)\cdot y'(\theta)\cdot\theta\right] \leq 1.$$

Once again, using (15), we replace $\frac{d\log\hat{\mu}(\theta)}{d\log\theta}$ to yield

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\left(-\frac{MRS_\theta \cdot \theta}{MRS(\theta)}\right)^{-1}\left[\begin{array}{l} -\frac{d}{d\log\theta}\log\left(\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\right) \\ -MRC_c(\theta)\cdot y'(\theta)\cdot\theta \end{array}\right] \leq 1, \quad (30)$$

which is the test formula in terms of skill distribution and marginal rate of substitution. We can also rewrite (30) to express the equivalent formula in an alternative way:

$$\begin{array}{rlcl} & (30) & & \\ \Leftrightarrow & -\theta\cdot\frac{1}{\theta f(\theta)}\cdot\frac{d}{d\theta}\left[\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\right] & & \\ & \quad -\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\cdot MRS_c(\theta)\cdot y'(\theta)\cdot\theta & \leq & 1 \\ \Leftrightarrow & -f(\theta) - \frac{d}{d\theta}\left[\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\right] & & \\ & \quad -\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\cdot MRS_c(\theta)\cdot y'(\theta)\cdot\theta f(\theta) & \leq & 0 \\ \Leftrightarrow & \frac{d}{d\theta}\left[\begin{array}{l} 1-F(\theta) - \frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1} \\ +\int_\theta^{\bar{\theta}}\frac{T'(\omega(\tilde{\theta}))}{1-T'(\omega(\tilde{\theta}))}\left(-\frac{MRS_\theta\cdot\tilde{\theta}}{MRS(\tilde{\theta})}\right)^{-1}\cdot MRS_c(\tilde{\theta})\cdot y'(\tilde{\theta})\cdot\tilde{\theta}f(\tilde{\theta})d\tilde{\theta} \end{array}\right] & \leq & 0. \end{array}$$

This alternative condition can be stated as

$$\frac{d}{d\theta}R(\theta) \leq 0, \quad \forall\theta, \quad (31)$$

where

$$\begin{aligned} R(\theta) &= 1-F(\theta) - \frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\theta}{MRS(\theta)}\right)^{-1} \\ &\quad +\int_\theta^{\bar{\theta}}\frac{T'(\omega(\tilde{\theta}))}{1-T'(\omega(\tilde{\theta}))}\left(-\frac{MRS_\theta\tilde{\theta}}{MRS(\tilde{\theta})}\right)^{-1}MRS_c(\tilde{\theta})y'(\tilde{\theta})\tilde{\theta}dF\tilde{\theta}. \end{aligned}$$

B.1 Test Formulas in terms of Earnings Distribution

In this section, we derive the test formulas in terms of earnings distribution (16) in Proposition 2 and alternative representation (17). Proof of Proposition **??**—expression of the earnings elasticities in two classes of models—is provided after deriving the formulas in terms of skill distribution.

Derivation of Earnings Elasticities In the main text, we defined the elasticities of earnings along nonlinear budget constraint and considering convex wage schedule $W(\cdot)$. These elasticities are derived here. The earnings function $\omega(1-\tau, I)$ which solves the worker's problem should satisfy the following first order condition:

$$G = MRS\left((1-\tau)\omega - T(\omega) + I, W^{-1}(\omega), \theta\right) - (1-\tau-T'(\omega))W'(W^{-1}(\omega)) = 0.$$

Using the implicit function theorem, we obtain

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

and

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

Thus the elasticities are

$$\epsilon_\omega^u \equiv \left.\frac{\partial \omega}{\partial(1-\tau)}\right|_{\tau=I=0} \frac{1-T'}{\omega} = \frac{\frac{W'}{\omega} - MRS_c}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'} \tag{32}$$

$$\eta_\omega \equiv -(1-T')\left.\frac{\partial \omega}{\partial I}\right|_{\tau=I=0} = \frac{MRS_c}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'} \tag{33}$$

$$\epsilon_\omega^c \equiv \epsilon_\omega^u + \eta_\omega = \frac{\frac{W'}{\omega}}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'}. \tag{34}$$

For later use, the ratio between η_ω and ϵ_ω^c is

$$\frac{\eta_\omega}{\epsilon_\omega^c} = MRS_c \frac{\omega}{W'(y)} = MRS_c \cdot y \cdot \frac{1}{\frac{W'(y)y}{W(y)}}. \tag{35}$$

Proof of Proposition 2 We now derive the Pareto efficiency test formulas in terms of earnings distribution. To relate the test formulas (30) and (31) to the earnings distribution, note that earnings of skill type θ is determined by $\omega(\theta) = W(y(\theta))$ in equilibrium, and the earnings distribution and the skill distribution have the relationship $H(\omega(\theta)) = F(\theta)$. Thus the associated density functions satisfy

$$f(\theta) = h(\omega(\theta))\omega'(\theta). \tag{36}$$

Next, earnings differential across skill $\omega'(\theta)$ is derived from the workers' optimality condition. Workers' choice of earnings $\omega(\theta)$ should satisfy the following first order condition:

$$G = MRS\left(\omega - T(\omega), W^{-1}(\omega), \theta\right) - (1 - T'(\omega))\, W'(W^{-1}(\omega)) = 0.$$

Then the earnings differential yields

$$\begin{aligned}
\omega'(\theta) &= -\frac{\frac{\partial G}{\partial \theta}}{\frac{\partial G}{\partial \omega}} = -\frac{MRS_\theta}{MRS_c(1-T') + MRS_y \frac{1}{W'} - \frac{W''}{W'}(1-T') + T''W'} \\
&= \frac{-\frac{MRS_\theta \theta}{MRS} W'}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'} \cdot \frac{1}{\theta} \\
&= \frac{\frac{W'}{\omega}}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'} \cdot \frac{\omega}{\theta} \cdot \left(-\frac{MRS_\theta \theta}{MRS}\right) \\
&= \epsilon^c_\omega \cdot \frac{\omega}{\theta} \cdot \left(-\frac{MRS_\theta \theta}{MRS}\right).
\end{aligned}$$

That is, the relationship between the derivative of marginal rate of substitution and elasticity of earnings is

$$\left(-\frac{MRS_\theta \theta}{MRS}\right)^{-1} \frac{\omega'(\theta)\theta}{\omega(\theta)} = \epsilon^c_\omega(\omega) \tag{37}$$

Also, note that from (35)

$$MRS_c(\theta) y'(\theta)\theta = \frac{\eta_\omega(\omega)}{\epsilon^c_\omega(\omega)} \cdot \frac{y'(\theta)\theta}{y(\theta)} \cdot \frac{W'(y(\theta))y(\theta)}{W(y(\theta))} = \frac{\eta_\omega(\omega)}{\epsilon^c_\omega(\omega)} \cdot \frac{\omega'(\theta)\theta}{\omega(\theta)}. \tag{38}$$

Using (36), (37), and (38), we can rewrite (30) as

$$\frac{T'(\omega)}{1-T'(\omega)} \epsilon^c_\omega(\omega) \left[-\frac{d}{d\log \omega} \log\left(\frac{T'(\omega)}{1-T'(\omega)} \omega h(\omega) \epsilon^c_\omega(\omega) \right) - \frac{\eta_\omega(\omega)}{\epsilon^c_\omega(\omega)} \right].$$

We represent the the revenue incidence of a tax reform of increasing tax rate at earnings level ω by $R^\omega(\omega)$. Note that the revenue incidence function represented in terms of skill distribution $R(\theta)$ is related to $R^{\omega}(\omega)$ by $R^{\omega}(\omega(\theta)) = R(\theta)$, and this implies that $\frac{dR(\theta)}{d\theta} = \frac{dR^{\omega}(\omega)}{d\omega}\omega'(\theta)$. Since $\omega'(\theta) = W'(y(\theta))y'(\theta) \geq 0$, we can also rewrite the alternative representation of the formula (31). Using

$$\left(-\frac{MRS_{\theta}(\theta)\theta}{MRS(\theta)}\right)^{-1} MRS_c(\theta)y'(\theta)\theta = \epsilon^c_{\omega}(\omega)\frac{\eta_{\omega}(\omega)}{\epsilon^c_{\omega}(\omega)} = \eta_{\omega}(\omega),$$

we obtain the test formula (2) in Proposition 2: $\frac{dR^{\omega}(\omega)}{d\omega} \leq 0$, where

$$R^{\omega}(\omega) = 1 - H(\omega) - \frac{T'(\omega)}{1-T'(\omega)} \times \omega h(\omega) \times \epsilon^c_{\omega}(\omega) + \int_{\omega}^{\bar{\omega}} \frac{T'(\tilde{\omega})}{1-T'(\tilde{\omega})}\eta_{\omega}(\tilde{\omega})h(\tilde{\omega})d\tilde{\omega}.$$

B.2 Test Formulas in terms of Skill Distribution and Elasticities

In this section, we derive the test formulas in terms of skill distribution and elasticities. We have already expressed the formulas in terms of skill distribution and marginal rate of substitutions above, thus we only need to relate the marginal rate of substitutions to the elasticities.

Derivation of Labor Supply Elasticities In the main text, we defined the elasticities of effective units of labor supply y along the linear budget constraint ($T'' = 0$) and assuming linear wage schedule ($W'' = 0$). We now derive these elasticities. Recall that under the assumption sof $T'' = 0$ and $W'' = 0$, the effective labor supply function $y(1-\tau, I)$ should satisfy the following first-order condition given a constant marginal tax rate $\hat{T}' = T'(y)$:

$$G = MRS\left((1-\tau)y - T(y) + I, y, \theta\right) - (1 - \tau - \hat{T}') = 0.$$

Using the implicit function theorem, we obtain

$$G = MRS\left((1-\tau)\omega - T(\omega) + I, W^{-1}(\omega), \theta\right) - (1 - \tau - T'(\omega))W'(W^{-1}(\omega)) = 0.$$

Using the implicit function theorem, we obtain

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

and

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

Thus the elasticities are

$$\begin{aligned}
\tilde{\epsilon}^u &\equiv \left.\frac{\partial y}{\partial(1-\tau)}\right|_{\tau=I=0} \frac{1-T'}{y} = \frac{\frac{1}{y} - MRS_c}{MRS_c + \frac{MRS_y}{MRS}} & (39)\\
\tilde{\eta} &\equiv -(1-T') \left.\frac{\partial \omega}{\partial I}\right|_{\tau=I=0} = \frac{MRS_c}{MRS_c + \frac{MRS_y}{MRS}} & (40)\\
\tilde{\epsilon}^c &\equiv \tilde{\epsilon}^u + \tilde{\eta} = \frac{\frac{1}{y}}{MRS_c + \frac{MRS_y}{MRS}}. & (41)
\end{aligned}$$

Fro later use, we also note that

$$\begin{aligned}
\frac{\tilde{\epsilon}^c}{1+\tilde{\epsilon}^u} &= \frac{\frac{1}{y}}{\frac{1}{y} + \frac{MRS_y}{MRS}} = \left(1 + \frac{MRS_y \cdot y}{MRS}\right)^{-1} & (42)\\
\frac{\tilde{\eta}}{\tilde{\epsilon}^c} &= MRS_c \cdot y & (43)
\end{aligned}$$

Note that although the wage rate structure $w(\theta, e)$ and specific utility form over (e, h) matters for the schedule of marginal rate of substitution (MRS), but given MRS schedule, the expression of the elasticities of y is equivalent to those in standard Mirrlees model.

Proof of Proposition ?? and Proposition ?? We now derive the test formulas in terms of skill distribution in two classes of models that we considered in the main text. Given the formulas (30) and (31) in terms of skill distribution and MRS, we relate the derivatives of MRS to the function of elasticities. This relationship between the MRS and elasticities depend on the types of models we consider.

Given that the elasticities are expressed in terms of MRS_y but that the formulas (30) and (31) include MRS_θ terms, we identify the relationship between MRS_θ and MRS_y. This relationship does depend on the wage structure $w(\theta, e)$ and primitive utility $\tilde{u}(c, e, h)$. Recall that workers choose (e, h) given (c, y, θ):

$$\begin{aligned}
U(c, y, \theta) &= \max_{e,h} \tilde{u}(c, e, h) \qquad s.t. \qquad w(\theta, e) l(e, h) = y\\
&= \max_{e} \tilde{u}\left(c, e, l^{-1}\left(e, \frac{y}{w(\theta, e)}\right)\right)
\end{aligned}$$

Then $MRS(c, y, \theta) = -\frac{U_y}{U_c} = -\frac{\tilde{u}_h}{\tilde{u}_c w l_h}$, where the second equality is from the envelope theorem.

Differentiating the MRS with respect to θ yields

$$\begin{aligned}
\frac{\partial MRS}{\partial \theta} &= -\frac{\tilde{u}_{hh} w_\theta l \tilde{u}_c - \tilde{u}_h \left[-\tilde{u}_{ch} w_\theta l + \tilde{u}_c w_\theta l_h - \tilde{u}_c l_{hh} w_\theta l\right]}{\tilde{u}_c^2 w^2 l_h^2}, \text{ and} \\
-MRS_\theta \theta &= -\frac{\tilde{u}_h}{\tilde{u}_c w l_h} \cdot \frac{w_\theta \theta}{w} + \left[-\frac{\tilde{u}_{hh}}{\tilde{u}_c w^2 l_h^2} + \frac{\tilde{u}_h \tilde{u}_{ch}}{\tilde{u}_c^2 w^2 l_h^2} + \frac{\tilde{u}_h l_{hh}}{\tilde{u}_c w l_h^3}\right] \cdot w_\theta \theta l \\
&= MRS \cdot \frac{w_\theta \theta}{w} + \left[-\frac{\tilde{u}_{hh}}{\tilde{u}_c w^2 l_h^2} + \frac{\tilde{u}_h \tilde{u}_{ch}}{\tilde{u}_c^2 w^2 l_h^2} + \frac{\tilde{u}_h l_{hh}}{\tilde{u}_c w l_h^3}\right] \cdot y \cdot \frac{w_\theta \theta}{w} \\
&= MRS \cdot \frac{w_\theta \theta}{w} + \frac{\partial MRS}{\partial y} \cdot y \cdot \frac{w_\theta \theta}{w}.
\end{aligned}$$

This implies

$$-\frac{MRS_\theta \theta}{MRS} = \left[1 + \frac{MRS_y \cdot y}{MRS}\right] \frac{w_\theta \theta}{w}.$$

Using (42), we obtain the following relationship between MRS_θ and elasticities:

$$-\frac{MRS_\theta \theta}{MRS} = \left(\frac{\tilde{\epsilon}^c(\theta)}{1 + \tilde{\epsilon}^u(\theta)}\right)^{-1} \frac{w_\theta \theta}{w} \tag{44}$$

Note that this relationship holds regardless of specific form of $W(\cdot)$. Using (43) and (44), (30) and (31) are rewritten in terms of skill distribution and elasticities. Proposition **??** assumes that $w(\theta, e) = \theta$, and thus $\frac{w_\theta \theta}{w} = 1$, then we obtain the formula (**??**). On the other hand, Proposition **??** assumes that $w(\theta, e) \neq \theta$, then the term of $\frac{w_\theta \theta}{w}$ does not disappear, and this yields the formula (**??**).

Proof of Proposition ??

In the first class of model with $W''(\cdot) > 0$ and $w(\theta, e) = \theta$, comparing the compensated earnings elasticity considering the convex wage schedule ϵ_ω^c derived in (34) to the compensated elasticity of effective units of labor supply along the linear budget and wage schedule $\tilde{\epsilon}^c$ derived in (41) yields the first equality in (**??**). Also, combining (35) and (44) with $\frac{w_\theta \theta}{w} = 1$ yields the second equality in (**??**).

In the second class of model with $W''(\cdot) = 0$ and $w(\theta, e) \neq \theta$, once again, we combine (35) and (44), but in this case the term of $\frac{w_\theta \theta}{w}$ does not disappear, and thus we obtain the relationship in (**??**).