Optimal Tax with Endogenous Wage

Yena Park
Seoul National University

March 20, 2025

Abstract

This paper studies optimal tax policy in an economy with endogenous wages. In particular, we put together important sources of endogenous wage inequality—(i) human capital, (ii) superstar effects with worker-firm assignment, and (iii) general equilibrium—and analyze how they work in optimal tax design. We use the optimal tax formulas to show the role of each component and their interaction. Human capital affects the efficiency cost of tax distortion not only through the elasticity of effective labor supply—combining the elasticity of human capital investment—but also through the return to human capital which relaxes the incentive constraint. Interaction of the superstar effects, assignment of worker-firm, and imperfect substitution in production provides a force to lower optimal marginal tax rate by making the optimal tax more sensitive to the efficiency cost of tax distortion. The general equilibrium effects affect the optimal tax schedule through two channels—(i) direct response of human capital return to tax change, and (ii) reinforcement/mitigation of the effectiveness of tax distortion in relaxing the incentive constraints—whose implication depends on the degree of substitution of workers.

Keywords: Optimal Tax, Endogenous Wage Inequality, Superstars, Human Capital, General Equilibrium
JEL Classification: E21, D64, H21

1 Introduction

How is the optimal tax schedule affected by the endogenous wage? This paper studies optimal tax policy in an economy where important sources of endogenous wage inequality are operating together. In the canonical optimal taxation model (e.g., Mirrlees (1971), Saez, Slemrod, and Giertz (2012)), worker's wage distribution is simply determined by the exogenous skill distribution, and the literature largely focuses on the labor supply incentives to capture the efficiency cost of the redistributive tax distortion. How should be the optimal equity-efficiency trade off adjusted when wage distribution is endogenous? Thanks to the development of the literature, we know how the endogenous wage affects this trade-off through the supply side channel of the production. Since the relative wage enters the incentive constraints of workers, the optimal tax schedule is adjusted to compress wage schedule so that it can relax the incentive constraint—this is so-called "trickle-down" forces of Stiglitz (1982).[1]

There are, however, other important sources which makes the worker's wage distribution endogenous. In this paper, we consider three different sources of endogenous wage together: (i) human capital (ii) superstar effects with worker-firm assignment, and (iii) general equilibrium. These are all important factors in shaping the endogenous wage inequality, but we have limited understanding how these sources together work with the impact of the tax policy. Putting them together in one framework is important not only because they are required to match the realistic wage distribution but also because additional forces arise due to the interaction of these factors in sharping the optimal tax schedule.

We develop a model with three sources listed above. The key difficulty is that we need to capture the key features of wage structure in the real world with the realistic interaction of the three sources, while still maintaining the model tractable enough to analyze the theoretical optimal tax formula. Especially, we want to have a unifying framework that can generate (i) the firm size distribution and sorting between worker and firms together, (ii) complementarity between workers and managers of a firm, and (iii) human capital investment decision under the realistic return to human capital—convex wage schedule.

[1]The conventional result is that accounting for the endogenous prices makes the optimal tax schedule less progressive.

To analyze the key mechanisms of accounting for the endogenous wage in the optimal tax design, we extend a Mirrlees model. We build on the combination of the general equilibrium model with human capital of Heathcote, Storesletten, and Violante (2017) and the superstar model of Scheuer and Werning (2017), with an extension to have assignment between workers and firms and the superstar effects for both workers and entrepreneurs. Workers with heterogeneous skills invest in human capital, and supply effective labor supply combining their exogenous skill and human capital given convex earnings schedule—convex function of effective labor supply which captures the worker's superstar effects. Labor inputs from different level of human capital are not perfect substitutes, and thus the marginal productivity of worker with each human capital level, which is the price of human capital, is increasing in the labor of the human capital levels that are complements in production. Entrepreneurs with heterogeneous managerial ability own firms and produce output by optimally combining their effort and composite of worker's labor given the convex cost function of labor hiring due to the convex earnings schedule of the workers. The entrepreneur's effective effort and the composite of worker's labor are complements in production which generates span of control, and at the same time, this leads to the superstar effects of the entrepreneurs.

Then we embed this model with human capital, superstar effects and general equilibrium to the Mirrlees optimal tax framework and derive optimal tax formulas for workers and entrepreneurs. The role of each component of the endogenous wage in optimal tax schedule can be clearly investigated by adding each component one by one from the classic Mirrlees model. First, when we introduce the endogenous return to human capital, the standard optimal tax formula for workers in Mirrlees (1971) extends with two modifications. The first modification is straightforward that we need the elasticity of effective labor supply combining the working hours and human capital investment. Another change is that the effective efficiency cost (EEC) of tax distortion represented in the formula depends on the return to human capital as well as the elasticity. Higher return to human capital lowers 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 leading to a higher optimal marginal tax rate.

Next, we introduce the imperfect substitutability in production and superstar effects with

the assignment between the workers and firms while keeping the price of human capital exogenous. Then, the optimal tax formula of workers includes additional terms due to the interaction of the superstar effects, worker-firm sorting, and the imperfect substitution of workers, which provide a force of decreasing marginal tax rate by making the tax schedule more sensitive to the efficiency cost of tax distortion. In the presence of span of control of the entrepreneurs increase in worker's labor supply with lower tax is likely to be assigned to more productive entrepreneurs, increasing allocative efficiency. This entrepreneru's superstar channel increases the efficiency cost of tax distortion. In addition, convex earnings schedule of the workers and imperfect substitution workers with different human capital makes worker's hiring more costly, increasing the efficiency cost of tax distortion and the increase is magnified when the allocation efficiency channel is stronger with high degree of complementarity between the entrepreneur's effort and worker's labor. Important, this new channel of the superstar effects does not arise when only convex earnings schedule is introduced. It is the interaction with the assignment structure and imperfect substitution in production which breaks the non-neutrality of the superstar effects. On the other hand, entrepreneur's optimal tax formula is not changed at all showing the neutrality of the superstar as in Scheuer and Werning (2017).

Lastly, we introduce the standard general equilibrium so that the equilibrium prices of human capital responds to a tax reform. Then the two additional effects show up in the worker's optimal tax formula. The first channel of the general equilibrium is the standard Stiglitlz effects. The planner sets optimal tax schedule with an eye how taxes affect the relative wage across workers to relax the incentive constraints by compressing the wage schedule, which typically makes the optimal tax schedule less progressive. The second channel of the general equilibrium accounts for whether the effectiveness of the tax distortion in relaxing the incentive constraints (given price) is reinforced or mitigated by the price change driven by the tax distortion. The directional implication of this second general equilibrium channel depends on in which direction the price responds to a tax distortion. Using a simple model with linear tax, we show that this price incidence relies on the degree of substitution across labor input from different levels of human capital. If the degree of substitution is low, the shift of demand tends to dominate the shift of supply, increase in tax distortion leads to de-

crease in wage rate, which mitigates the effectiveness tax distortion in relaxing the incentive constraints, which then provides a force to lower optimal tax rate.

Our analytical investigation with the tax formula shows that the directional implication of the general equilibrium in optimal tax schedule requires quantitative investigation. We plan to explore this.

Literature review - need to fill out

- Sachs, Tsyvinski, and Werquin (2020)
- Scheuer and Werning (2017)
- Heathcote, Storesletten, and Violante (2017)
- Ales, Kurnaz, and Sleet (2015)
- Piketty, Saez, and Stantcheva (2014)
- Stantcheva (2017)

2 Environment

Our economy extends the canonical Mirrlees' model to incorporate the general equilibrium effects and super star effects on optimal taxation. 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.[2] Our economy differs from this standard model in two aspects. First, the return to human capital investment is endogenized in a general equilibrium. Second, the wage schedule of both workers and entrepreneurs are convex due to the superstar effects in the economy.

Worker's Preferences and Types Consider an economy with both workers and entreprenuers. For simplicity, we abstract from the occupation choice. There is a continuum of workers with continuous skill types. 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 of human capital investment $\tilde{e}(e, \theta)$ is function of e is the level of human capital investment and and θ is skill type of workers described below. Depending on the form of function $\tilde{e}(\cdot, \cdot)$, different types of human capital models can be captured. In the benchmark economy, we assume that

$$\tilde{e}(e, \theta) = \frac{e}{\theta},$$

where the effective effort $\tilde{e}$ required to obtain the level of human capital e is decreasing in skill θ. That is, the cost of investing in human capital e is decreasing in skill type θ in the baseline. Later, we do the sensitivity analysis with other forms of effort function $\tilde{e}(e, \theta)$.

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 pos-

[2]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.

itive and continuously differentiable density $f(\theta)$.[3] Following the Mirrlees' spirit, θ can be interpreted as comprehensive income generating ability.

Wage Structure We denote effective unit of worker's 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 $n(\theta)$. We assume that worker's productivity has the following form:

$$p(\theta, e) = \theta^{\gamma} \cdot w(e),$$

where an exogenous parameter γ controls the return to unobserved skill θ, and the return to human capital e is determined by $w(\cdot)$ which is endogenous function determined by the marginal product of labor of each human capital level.

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 of workers—return to effective labor supply is increasing at the top. For simplicity, we assume that

$$W(y) = y^{\kappa}, \quad \kappa \geq 1, \tag{1}$$

where $\kappa > 1$ allows the superstar effects of workers, and $\kappa = 1$ is assumed for the analysis with no superstar effects for workers.[4]

For later use, we also define the Hicksian coefficient of complementarity between skill θ and human capital e by $\rho(\theta, e) = \frac{p_{\theta e} \cdot p}{p_{\theta} \cdot p_e}$. Note that under our assumption on the form of productivity, the Hicksian coefficient of complementarity between skill and human capital is one. If we extend the model to the CES form of production 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 $\rho(\theta, e) = \rho$.[5]

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

[4] Note that the exact functional form of worker's earnings $W(\cdot)$ is not important, except that it is convex function to reflect the superstar effects of the workers.

[5] 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.

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 productivity function $p(\theta, e)$ is endogenously determined in equilibrium, but workers take them as given. 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 $T(\cdot)$:

$$\max_{c,e,n} u(c, \tilde{e}, n) \quad s.t. \quad c \leq W(p(\theta, e) \cdot n) - T\left(W(p(\theta, e) \cdot n)\right), \tag{2}$$

where $\tilde{e} = \frac{e}{\theta}$.

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:

$$U(c, y, \theta) = \max_{e,n} u\left(c, \tilde{e}(e, \theta), n\right) \quad s.t. \quad y \leq p(\theta, e) \cdot n, \tag{3}$$

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

$$\max_{c,\omega} U(c, W^{-1}(\omega), \theta) \quad s.t. \quad c \leq \omega - T(\omega). \tag{4}$$

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

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

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 Entrepreneur's Problem Our economy extends the model with endogenous return to human capital (similar to the model in Heathcote, Storesletten, and Violante (2020)) to include the span of controls of entrepreneurs. There is continuum of entrepreneurs with continuous skill types. The skill distribution entrueprenerus is given by $G(x)$, where x is the managerial ability.[6] Entrepreneur with ability x produces output by combining his own effective effort $y(x) = x \cdot n(x)$ and a composite labor $Y(x)$ which combines worker's labor. The labor from different levels of human capital are imperfect substitutes, and the aggregation of worker's labor at firm x is represented by

$$1 = \Omega(e)^{\frac{1}{\eta}} \cdot \ell(e|x)^{\frac{\eta-1}{\eta}} \cdot Y(x)^{\frac{\epsilon(e)-\eta}{\eta}} m(e)de, \tag{5}$$

which is non-homothetic technology when $\epsilon(e)$ depends on e. This technology representation includes the standard CES technology when $\epsilon(e) = 1$ for all e, which is the case we focus on later for more detailed analysis. Then an entrepreneur with ability x produces output with technology $A(y(x), Y(x))$. We assume that $A_{yY} > 0$ which generates the span of control of the entrepreneurs with more effort.

Entrepreneurs have preferences represented by $\tilde{u}(c, n)$, where $\tilde{u}$ is increasing in consumption c and decreasing in working hour n. Given a tax schedule $T(\cdot)$ and a price schedule $w(e)$, an entrepreneur with productivity x solves the following problem:

$$\begin{aligned} \max_{\{\ell(e)\},c,n,y,Y,\omega} \quad & \tilde{u}(c, n) \\ s.t. \quad & c = \omega - T(\omega) \\ & \omega = A(y, Y) - \int W\left(w(e) \cdot \ell(e)\right) m(e)de \\ & y = x \cdot n \\ & 1 = \Omega(e)^{\frac{1}{\eta}} \cdot \ell(e)^{\frac{\eta-1}{\eta}} \cdot Y^{\frac{\epsilon(e)-\eta}{\eta}} m(e)de. \end{aligned} \tag{6}$$

This problem can be rewritten as a two-step decision: 1. chooses $(y(x), c(x))$, and 2. hires $\{\ell(e|x)\}$ given $(y(x), c(x))$. Solving backwards, in the second step, an entrepreneur with given

[6] Alternatively, we can think of $F(\theta)$ as the skill distribution of all individuals with exogenous cutoff $\bar{\theta}$ below which individuals become workers and above which individuals become entrepreneurs. We can endogenize the cutoff $\bar{\theta}$, but then an additional entry/exit channel will arise in the optimal tax formula.

effort $y(x)$ makes the hiring decision of worker's labor to maximize profit, which determines the earnings of the entrepreneurs $\omega(x) = \tilde{W}(y(x))$:

$$\begin{aligned} \tilde{W}(y(x)) &= \max_{Y} A(y(x), Y) - C(Y), && (7) \\ \text{where} \quad C(Y) &= \min_{\{\ell(e)\}} \int [w(e) \cdot \ell(e)]^{\kappa} m(e) de && (8) \\ (\lambda_c(Y)) \quad s.t. \quad 1 &= \Omega(e)^{\frac{1}{\eta}} \cdot \ell(e)^{\frac{\eta-1}{\eta}} \cdot Y^{\frac{\epsilon(e)-\eta}{\eta}} m(e) de, && (9) \end{aligned}$$

where (8) uses the functional form assumption (1). We denote the solution to the profit maximization (7) by $Y(y(x))$ and the solution to the cost minimization (8) by $\{\ell_D(e|Y(y(x)))\}$. By solving the cost-minimization (8), we can obtain the cost function $C(Y(y(x))) = \frac{\lambda_c(Y(y(x)))}{\kappa} \frac{\eta-1}{\eta}$, and the profit maximization (7) should satisfy

$$A_Y(y(x), Y(y(x))) = C'(Y(y(x))). \tag{10}$$

We remark that the superstar effects of the enterepreneurs—the convexity of the entrepreneur's earnings schedule—is endogenously generated by the supermodularity between the entrepreneur's effort and aggregate labor hiring $A_{yY} > 0$. By taking the envelope condition to (**??**), we have $\tilde{W}'(y(x)) = A_y(y(x), Y(y(x)))$, and thus

$$\tilde{W}''(y(x)) = A_{yY}(y(x), Y(y(x))) \times Y'(y(x)) + A_{yy}(y(x), Y(y(x))),$$

which shows that $A_{yY} > 0$ provides a force for convex earnings schedule of entrepreneurs $\tilde{W}''(y) > 0$.

Entrepreneurs' preferences can be represented in terms of effective effort y: $\tilde{U}(c, y, x) = \tilde{u}\left(c, \frac{y}{x}\right)$. Then, in the first-step decision, an entrepreneur with productivity x solves the following problem given the earnings function $\tilde{W}(y)$:

$$\max_{\omega} \tilde{U}(\omega - T(\omega), \tilde{W}^{-1}(\omega), x). \tag{11}$$

The solution to this problem—entrepreneur's earnings—is denoted by $\omega(x)$ which in turn determines the entrepreneur's optimal effective effort $y(x) = \tilde{W}^{-1}(\omega(x))$ and optimal aggregate labor demand $Y(y(x))$. We denote the aggregate labor of the economy by Y, which is calculated by integrating the aggregate labor demand of all entrepreneurs: $Y = \int Y(y(x))g(x)dx$.

Equilibrium An equilibrium consists of a tax schedule $T(\omega)$, schedules for workers' earnings $W(y)$ and for entrepreneurs' earnings $\tilde{W}(y)$, allocation for workers $\{c(\theta), y(\theta), e(\theta), n(\theta), \omega(\theta)\}$ and for entrepreneurs $\{c(x), y(x), n(x), Y(y(x)), \omega(x), \{\ell(e|x)\}\}$, and wage rate of labor for each e $\{w(e)\}$ such that (i) for each θ, $(c(\theta), e(\theta), n(\theta))$ solves (2) given $W(\cdot)$ and $p(\theta, e) = \theta^{\gamma} \cdot w(e)$, 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, $(c(x), n(x), y(x), Y(y(x)))$ solves (6) given $\tilde{W}(\cdot)$, with the associated demand for labor $\{\ell(\theta|x)\}$; (iii) labor markets clear

$$\int \ell(e(\theta)|x) g(x) dx = \theta^{\gamma} \cdot n(\theta);$$

and (iv) the goods market clearing condition holds:

$$\int c(\theta) dF(\theta) + \int c(x) dG(x) = \int A(y(x), Y(y(x))) g(x) dx. \tag{12}$$

3 Optimal Tax Formulas

In this section, we derive the optimal tax formulas in an economy with human capital, superstar effects and general equilibrium. We first begin with the general labor aggregation technology to obtain the general formula, and then proceed further with the CES technology assumption to obtain sharper characterization.

3.1 Planning Problem

First of all, we set up the planning problem in an extended version of Mirrlees model. We assume that for the workers, 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 for workers $(c(\theta), y(\theta))$ chosen by the planner, the allocation of $(e(\theta), n(\theta))$ solves (3), and the planner only cares for the value function is $U(c(\theta), y(\theta), \theta)$.

First Order Approach For any allocation of workers $(c(\theta), y(\theta))$, we denote the utility delivered by this allocation by

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

An allocation of workers is incentive compatible if

$$v(\theta) = \max_{\theta'} U(c(\theta'), y(\theta'), \theta; p) \quad \forall \theta. \tag{13}$$

Similarly, for any allocation of entrepreneurs $(c(x), y(x))$, we denote the utility delivered by this allocation

$$\tilde{v}(x) = \tilde{U}(c(x), y(x), x).$$

An allocation of entrepreneurs is incentive compatible if

$$\tilde{v}(x) = \max_{x'} \tilde{U}(c(x'), y(x'), x) \quad \forall x. \tag{14}$$

To derive the optimal tax formula, 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), allocation for workers $(c(\theta), y(\theta))$ can be implemented as an equilibrium for some tax schedule if and only if it satisfies resource feasibility (12) and incentive compatibility (13) and (14). 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:

$$v'(\theta) = U_\theta(c(\theta), y(\theta), \theta; p) \quad \forall \theta. \tag{15}$$

Similarly, for the entrepreneurs, we can use the first order approach with the envelope condition:

$$\tilde{v}'(x) = \tilde{U}_x(c(x), y(x), x) \quad \forall x. \tag{16}$$

Thus, we can replace the incentive-compatibility constraints by the envelope conditions (15) and (16), and drop the monotonicity conditions, and check the monotonicity ex post after solving the mechanism problem.

Planning Problem The planner also takes into account that the equilibrium wage rate of each human capital level e, $w(e)$ is determined to clear the labor markets. We denote the firm x's demand for the worker's effective labor with skill type θ by $y(\theta|x)$, then the labor market clearing condition can be expressed as follows:

$$\int y(\theta|x)g(x)dx = y(\theta) \quad \forall\theta, \tag{17}$$

which is essentially the labor market clearing condition for each human capital level $e(\theta)$. With the technology (5) which aggregates the imperfectly substitutable labor $\{\ell(e|x)\}$, the composite labor input $Y(y(x))$ of firm x's production should satisfy the following aggregation (see the Appendix for its derivation):

$$\lambda_c(Y(y(x))) = \kappa \cdot \frac{\eta}{\eta-1} \cdot \int y(\theta|x)^\kappa f(\theta)d\theta \quad \forall x, \tag{18}$$

where $\lambda_c(Y(y(x)))$ is the Lagrange multiplier of the cost minimization problem (8). We remark that this multiplier depends on the wage rate. We also note that the value function of the worker $U(c, y, \theta)$ depends on the endogenous price $\{w(e)\}$, as the productivity function is $p(\theta, e) = \theta^\gamma \cdot w(e)$.

Let $C(v, y, \theta)$ be the inverse function of $U(c, y, \theta)$, defined according to $v = U(C[v, y, \theta], y, \theta)$ and $\tilde{C}(\tilde{v}, y, x)$ be the inverse function of $\tilde{U}(c, y, x)$, defined according to $\tilde{v} = \tilde{U}\left(\tilde{C}[\tilde{v}, y, x], y, x\right)$. Given a Pareto weight $\Lambda_\theta(\theta)$ to workers of skill θ and a weight $\Lambda_x(x)$ to entrepreneurs of productivity x, the planner solves the following problem:

$$\begin{aligned} \max_{\{v(\theta),y(\theta)\},\{\tilde{v}(x),y(x)\}} \quad &= \quad \int \Lambda_\theta(\theta)v(\theta)f(\theta)d\theta + \int \Lambda_x(x)\tilde{v}(x)g(x)dx \\ s.t. \quad &(12), (15), (16), (17), \quad \text{and} \quad (18). \end{aligned}$$

Since the composite labor $Y(y(x))$ and the effective labor demand $y(\theta|x)$ only appears in the constraints (12), (17), and (18), we can rewrite the planner's problem as follows:

$$\begin{aligned} \max_{\substack{\{v(\theta),\, y(\theta)\},\\ \{\tilde{v}(x),\, y(x)\}}} \quad &= \quad \int \Lambda_\theta(\theta)v(\theta)f(\theta)d\theta + \int \Lambda_x(x)\tilde{v}(x)g(x)dx \qquad (19) \\ (\mu(\theta)) \quad s.t. \quad & v'(\theta) = U_\theta(C(v(\theta), y(\theta), \theta), y(\theta), \theta; p) \quad \forall\theta, \\ (\tilde{\mu}(x)) \qquad & \tilde{v}'(x) = \tilde{U}_x(C(\tilde{v}(x), y(x), x), y(x), x) \quad \forall x, \\ (\lambda) \qquad & \Psi\left(\{y(\theta)\}, \{y(x)\}; \lambda_c(\cdot)\right) - \int C(v(\theta), y(\theta), \theta)dF(\theta) - \int C(v(x), y(x), x)dG(x) \geq 0, \end{aligned}$$

where the aggregate output is

$$\begin{aligned}
\Psi\left(\{y(\theta)\},\{y(x)\};\lambda_c(\cdot)\right) &= \max_{\{y(\theta|x)\},\{Y(y(x))\}} \int A\left(y(x),Y(y(x))\right) g(x)dx \qquad (20)\\
(\gamma(x)) \quad s.t. \quad & \lambda_c(Y(y(x))) = \kappa \cdot \frac{\eta}{\eta-1} \cdot \int y(\theta|x)^{\kappa} f(\theta)d\theta \quad \forall x,\\
(\phi(\theta)) \quad & \int y(\theta|x)g(x)dx = y(\theta) \quad \forall \theta.
\end{aligned}$$

Despite the extension of the model, the planner's problem (19) itself is not different from a standard Mirrleesian problem except the distinction between the workers and entrepreneurs and the expression of the aggregate production function $\Psi(\cdot)$. This is because the followings. First, the planner does not regulate the choice of effort $e(\theta)$ and working hours $n(\theta)$ separately. Thus given the value function $U(c,y,\theta)$, the planner's problem boils down to solving optimal allocation of choosing consumption $c(\theta)$ and effective units of labor $y(\theta)$ as in the standard Mirrleesian problem. Second, the earnings schedules of workers $W(\cdot)$ and of entrepreneurs $\tilde{W}(\cdot)$ are simply captured in the marginal product of Ψ which is aggregate production function expressed in a reduced form.

However, the optimal tax formula does depend on the detailed assumptions on the wage structure and production technology which are hidden in the value functions U, $\tilde{U}$ and aggregate production function Ψ, as the elasticity of earnings and elasticity of prices do depend on those assumptions. Following lemma explicitly shows that how the marginal product of aggregate production depends on the structure of our economy.

Lemma 2. *The marginal products of the worker's effective labor supply and entruepreneurs labor supply in aggregate production given the $\lambda_c(\cdot)$ are*

$$\begin{aligned}
\frac{\partial \Psi}{\partial y(\theta)} &= W'(y(\theta)) \cdot f(\theta) \cdot \frac{\int y(\theta|x)^{\kappa} g(x)dx}{y(\theta)^{\kappa}}, \qquad (21)\\
\frac{\partial \Psi}{\partial y(x)} &= A_y(y(x),Y(y(x))) \cdot g(x) = \tilde{W}'(y(x)) \cdot g(x). \qquad (22)
\end{aligned}$$

The effect of increase in $y(\theta)$ on the aggregate production through the change in $\lambda_c(\cdot)$ is

$$\begin{aligned}
&\int \frac{\partial \Psi}{\partial \lambda_c(Y(y(x)))} \left.\frac{d\lambda_c(Y(y(x)))}{dy(\theta)}\right|_{Y(y(x))} dx\\
&= -\frac{1}{y(\theta)} \cdot \int \kappa \cdot y(\tilde{\theta})^{\kappa} \frac{\int y(\tilde{\theta}|x)^{\kappa} g(x)dx}{y(\tilde{\theta})^{\kappa}} \cdot \epsilon_{w(\tilde{\theta}),y(\theta)} f(\tilde{\theta})d\tilde{\theta} - \int \lambda_c(Y(y(x)))B(x)g(x)dx, \qquad (23)
\end{aligned}$$

where $B(x)$ captures a change in the constraint (9) through a change in $e(\theta)$ in response to increase in $y(\theta)$.

Proof See the Appendix. ■

In the standard Mirrleesian model, the marginal product of effective labor is one. Lemma 2 shows that in our economy, the marginal product of effective labor reflects both the superstar effects and general equilibrium. Although the resource constraint is simply expressed in terms of reduced form of aggregate production function Ψ, its marginal products are affected by the convexity of the earnings schedule of both workers—driven by $\kappa > 1$—and entrepreneurs—driven by $A_{yY} > 0$. Endogenous return to human capital through both the change in price schedule $w(e)$ and the change in human capital $e(\theta)$ also affects the marginal product of effective labor.

In the next lemma, we simplify the expressions in Lemma 2 under the CES technology.

Lemma 3. *Assume that the labor aggregation technology is CES:*

$$1 = \int e^{\psi} \ell(e)^{\frac{\eta-1}{\eta}} Y^{\frac{1-\eta}{\eta}} m(e) de. \tag{24}$$

Then (21) can be rewritten as

$$\frac{\partial \Psi}{\partial y(\theta)} \quad = \quad W'(y(\theta)) \cdot f(\theta) \cdot \frac{\int Y(y(x))^{\kappa} g(x) dx}{Y^{\kappa}}. \tag{25}$$

Under the CES aggregation, $\lambda_c(Y(y(x))) = \lambda_c \cdot Y(y(x))^{\kappa}$, where $\lambda_c \equiv \lambda_c(1)$. Then the effect of increase in $y(\theta)$ on the aggregate production through the change in $\lambda_c(\cdot)$ boils down to

$$\frac{\partial \Psi}{\partial \lambda_c} \frac{d\lambda_c}{dy(\theta)} = \frac{\kappa}{y(\theta)} \cdot \frac{\int Y(y(x))^{\kappa} g(x) dx}{Y^{\kappa}} \cdot \left[\begin{array}{l} -\int y(\tilde{\theta})^{\kappa} \cdot \epsilon_{w(e(\tilde{\theta})),y(\theta)} f(\tilde{\theta}) d\tilde{\theta} \\ +\psi \cdot \frac{\eta}{\eta-1} \int y(\tilde{\theta})^{\kappa} \cdot \epsilon_{e(\tilde{\theta}),y(\theta)} f(\tilde{\theta}) d\tilde{\theta} \end{array} \right]. \tag{26}$$

Proof See the Appendix. ■

3.2 Optimal Tax Formulas

The optimal tax formulas can be expressed either in terms of skill distribution or in terms of earnings distribution depending on the main purpose of the formulas. In this paper, we mostly focus on the (structural) formula expressed in terms of skill distribution, as it is more useful to understand the underlying mechanisms of how the superstar effects and general equilibrium effects affect the optimal tax schedule. Later, we also express the formula in terms of earnings distribution which can be used for sufficient statistics approach and provide brief analysis.

Easticities We first define the elasticities of effective units of labor y—for both worker's $y(\theta)$ and entrepreneur's $y(x)$—with respect to marginal tax rate change along the linear wage schedule (assuming no superstar effects) and linear budget constraint, which are standard elasticities commonly used in the canonical Mirrleesian models. Despite the extension of the model with wage structures and superstar effects, these standard elasticities are relevant ones when the optimal tax formula is expressed in terms of skill distribution.[7]

By definition, these elasticities are equivalent to the elasticities of earnings assuming $W(y) = y$ (and $\tilde{W}(y) = y$) and $T''(\omega) = 0$. Under this assumption, the function of effective units of worker's 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 worker's elasticities of y along the linear wage schedule and linear budget constraint are defined by

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

These elasticities are the combination of elasticities of working hours and human capital investment considering the return to θ and e, which is hidden in the value function U. We also note that these are elasticities given price schedule $\{w(e)\}$. The entrepreneur's elasticities

[7] When we express the tax formula in in terms of earnings distribution, the relevant elasticities are the elasticities of earnings along the nonlinear budget constraint. We define these later.

of y along the linear wage schedule and linear budget constraint are similarly defined with entrepreneur's utility function $\tilde{U}(c, y, x)$. We denote the elasticities of entrepreneurs by $\tilde{\epsilon}_e^u(y)$, $\tilde{\eta}_e(y)$, and $\tilde{\epsilon}_e^c(y)$.

We next define the elasticities of prices, which are used to capture the general equilibrium effects. We denote the total elasticities of the human capital price $w(e(\tilde{\theta}))$ with respect to worker's effective labor $y(\theta)$ and entrepreneur's labor $y(x)$ by

$$\epsilon_{w(e(\tilde{\theta})),y(\theta)} = \frac{dw(e(\tilde{\theta}))}{dy(\theta)} \frac{y(\theta)}{w(e(\tilde{\theta}))}, \quad \text{and} \quad \epsilon_{w(e(\tilde{\theta})),y(x)} = \frac{dw(e(\tilde{\theta}))}{dy(x)} \frac{y(x)}{w(e(\tilde{\theta}))},$$

where $\frac{dw(e(\tilde{\theta}))}{dy(\theta)} \frac{y(\theta)}{w(e(\tilde{\theta}))}$ and $\frac{dw(e(\tilde{\theta}))}{dy(x)}$ are total derivatives capturing the incidences on $w(e(\theta))$ both through the change in price schedule $w(\cdot)$ and through the change in level of human capital $e(\theta)$. Note that given $w(\cdot)$-schedule fixed, the wage rate $w(e(\theta))$ can still be changed in response to increase in $y(\theta)$ as the human capital $e(\theta)$ increases given price. Since this change in human capital given price is captured in the elasticities of earnings defined above, we also define the price elasticity net of this effect:

$$\epsilon^o_{w(e(\tilde{\theta})),y(\theta)} = \left.\frac{dw(e(\tilde{\theta}))}{dy(\theta)}\right|^o \frac{y(\theta)}{w(e(\tilde{\theta}))}, \quad \text{where} \quad \left.\frac{dw(e(\tilde{\theta}))}{dy(\theta)}\right|^o = \frac{dw(e(\tilde{\theta}))}{dy(\theta)} - \mathbb{1}_{\tilde{\theta}=\theta} \cdot w'(e(\tilde{\theta})) \cdot \left.\frac{de(\tilde{\theta})}{dy(\theta)}\right|_{w(\cdot)},$$

which is the elasticity of our interest to analyze the role of the general equilibrium below.

Here, we do not present the analytical expression of the price elasticities, but we just define them and provide notation. Later in Section 4, using a simple model with further assumptions, we investigate the analytical expression of these elasticities to better understand how the price schedule changes to the tax reform.

Optimal Tax Formulas We now express the optimal tax formulas in our economy with human capital, superstar effects and general equilibrium. Under the general labor aggregation technology which takes the form of (5).

Proposition 4. *Assume that each firm's aggregation of worker's labor takes the form of (5.*

The optimal tax schedule for workers for any Pareto weights $\{\Lambda_\theta(\theta)\}$ satisfies

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}+\frac{1}{1-T'(\omega(\theta))}\left[\begin{array}{l}\left(\frac{\int y(\theta|x)^\kappa g(x)dx}{y(\theta)^\kappa}-1\right)-\int\frac{y(\tilde\theta)^\kappa}{y(\theta)^\kappa}\cdot\frac{\int y(\theta|x)^\kappa g(x)dx}{y(\theta)^\kappa}\epsilon_{w(e(\tilde\theta)),y(\theta)}\frac{f(\tilde\theta)}{f(\theta)}d\tilde\theta\\ -\frac{1}{f(\theta)}\int\frac{\lambda_c(Y(y(x)))}{\kappa y(\theta)^{\kappa-1}}B(x)g(x)dx\end{array}\right] \tag{27}$$

$$\begin{aligned}=\quad&\hat\mu(\theta)\frac{1}{\theta f(\theta)}\cdot\frac{1+\tilde\epsilon^u(\theta)}{\tilde\epsilon^c(\theta)}\cdot\left(\frac{w_e e}{w}(\theta)+\frac{w_\theta\theta}{w}\right)\\&+\frac{1}{\theta f(\theta)}\cdot\int\hat\mu(\tilde\theta)\frac{1-T'(\omega(\tilde\theta))}{1-T'(\omega(\theta))}\cdot\frac{y(\tilde\theta)^\kappa\tilde\theta}{y(\theta)^\kappa\theta}\cdot y(\theta)\frac{d\left(\frac{w_e e}{w}(\tilde\theta)\right)}{dy(\theta)}d\tilde\theta\\&+\frac{1}{\theta f(\theta)}\cdot\int\hat\mu(\tilde\theta)\frac{1-T'(\omega(\tilde\theta))}{1-T'(\omega(\theta))}\cdot\frac{y(\tilde\theta)^\kappa/\tilde\theta}{y(\theta)^\kappa/\theta}\left(-\frac{1+\tilde\epsilon^u(\tilde\theta)}{\tilde\epsilon^c(\tilde\theta)}-\frac{U_{cy}(\tilde\theta)y(\tilde\theta)}{U_c(\tilde\theta)}\right)\epsilon^o_{w(\tilde\theta),y(\theta)}\left(\frac{w_e e}{w}(\theta)+\frac{w_\theta\theta}{w}\right)d\tilde\theta\end{aligned}$$

where $\hat\mu(\tilde\theta)=$.

3.3 Understanding the Mechanisms in the Optimal Tax

Formulas under the CES and No Income Effects For a sharp characterization of the role of the superstar effects and general equilibrium in optimal tax, we take a first step to simplify the formulas by making the following two assumptions. First, the labor aggregation technology is CES as in (24). Second, worker's preferences have no income effects in working hours ($n(\theta)$) decision and human capital ($e(\theta)$) investment.

Under these assumptions, we have the following two observations which simplify the left-hand side of the formula (27)—the effect of tax reform on the aggregate production. First, under the CES aggregation technology, proportion of worker's skill type hired at each firm satisfies

$$\frac{y(\theta|x)}{y(\theta)}=\frac{Y(y(x))}{Y},\quad\forall\theta,\forall x.$$

This implies that the distribution of worker's skill hired at each firm does not depend on the entrepreneur's productivity x. Every firm has the identical distribution of worker's labor over the worker's skill types (with the same $\frac{y(\theta|x)}{Y(y(x))}$ for all x), although the scale of the firm $Y(y(x))$ does depend on the entrepreneur's productivity x.

Second, under the CES aggregation technology, the cost minimization (8) implies that wage rate $w(e)$ of labor at human capital level e satisfies

$$w(e)=\left(\frac{\lambda_c}{\kappa}\cdot e^\psi\right)^{\frac{1}{\kappa}}\times\left(\frac{Y}{\ell(e)}\right)^{\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)}, \tag{28}$$

which shows that it depends on other type's labor only through aggregate labor Y. Then the rate of return to human capital investment $\frac{w_e e}{w}(\theta)$ at the level of human capital $e(\theta)$ does not depend on the labor supply of other levels of human capital $e(\tilde{\theta})$ for $e(\tilde{\theta}) \neq e(\theta)$. Since worker's human capital investment depends only on the relative return $\frac{w_e e}{w}(\theta)$ not on the level of wage $w(e(\theta))$ under the preferences with no income effects, the elasticity of human capital $e(\tilde{\theta})$ with respect to $y(\theta)$ is nonzero only for the change in labor of own skill type.

Then the optimal tax formula for workers (27) rewrites:

$$\begin{aligned}
&\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))} + \frac{1}{1-T'(\omega(\theta))} \left[\begin{array}{l} \left(\frac{\int Y(y(x))^\kappa g(x) dx}{Y^\kappa} - 1 \right) - \frac{\int Y(y(x))^\kappa g(x) dx}{Y^\kappa} \int \frac{y(\tilde{\theta})^\kappa}{y(\theta)^\kappa} \epsilon_{w(e(\tilde{\theta})),y(\theta)} \frac{f(\tilde{\theta})}{f(\theta)} d\tilde{\theta} \\ +\psi \cdot \frac{\eta}{\eta-1} \cdot \frac{\int Y(y(x))^\kappa g(x) dx}{Y^\kappa} \cdot \epsilon_{e(\theta),y(\theta)} \end{array} \right] \qquad (29) \\
= \quad & \hat{\mu}(\theta) \cdot \frac{1}{\theta f(\theta)} \cdot \frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)} \cdot \left(\frac{w_e e}{w}(\theta) + \frac{w_\theta \theta}{w} \right) \\
& + \hat{\mu}(\theta) \cdot \frac{1}{\theta f(\theta)} \cdot y(\theta) \frac{d\left(\frac{w_e e}{w}(\theta) \right)}{dy(\theta)} \\
& + \frac{1}{\theta f(\theta)} \cdot \int \hat{\mu}(\tilde{\theta}) \frac{1-T'(\omega(\tilde{\theta}))}{1-T'(\omega(\theta))} \cdot \frac{y(\tilde{\theta})^\kappa / \tilde{\theta}}{y(\theta)^\kappa / \theta} \left(-\frac{1+\tilde{\epsilon}^u(\tilde{\theta})}{\tilde{\epsilon}^c(\tilde{\theta})} - \frac{U_{cy}(\tilde{\theta}) y(\tilde{\theta})}{U_c(\tilde{\theta})} \right) \epsilon^o_{w(\tilde{\theta}),y(\theta)} \left(\frac{w_e e}{w}(\theta) + \frac{w_\theta \theta}{w} \right) d\tilde{\theta}.
\end{aligned}$$

Note that despite the simplified effects of tax reform on production, the formula (29) still includes the weighted integration of the cross price elasticities.

Rewriting the Formula with Normalization Under the CES technology and preferences with no income effects, we can proceed further with normalization to interpret the exact roles of general equilibrium in the optimal tax schedule. We normalize the wage rates by the common factor across $w(e)$'s, which allows us to greatly simplify the terms related to the cross price elasticities so that we can better understand the mechanisms of the superstar effects and general equilibrium. From the wage rate expression (28), the common endogenous factor across all human capital levels are λ_c and Y. We thus normalize wage rates by these common factors and denote it by

$$\tilde{w}(e) \equiv \frac{1}{\lambda_c^{\frac{1}{\kappa}} \cdot Y^{\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)}} \cdot w(e), \quad \forall e. \qquad (30)$$

We also normalize $p(\theta, e)$, $y(\theta)$, and $Y(y(x))$ by the same factor: $\tilde{p}(\theta, e) \equiv \frac{p(\theta,e)}{\lambda_c^{\frac{1}{\kappa}} \cdot Y^{\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)}} = \theta^\gamma \cdot \tilde{w}(e)$, $\tilde{y}(\theta) \equiv \frac{y(\theta)}{\lambda_c^{\frac{1}{\kappa}} \cdot Y^{\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)}} = \tilde{w}(e(\theta)) \cdot \ell(e(\theta))$, and $\tilde{Y}(y(x)) \equiv \frac{Y(y(x))}{\lambda_c^{\frac{1}{\kappa}} \cdot Y^{\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)}}$ respectively.

We then rewrite the planner's problem with these normalized variables. The problem is the same as in (19) except that the envelope condition (15) of the workers is rewritten with the redefined value function $U(c, \tilde{y}, \theta; \tilde{p}) = \max_e u\left(c, \frac{e}{\theta}, \frac{\tilde{y}}{\tilde{p}(\theta,e)}\right)$ and the production function in the resource constraint is replaced by the following function:

$$\begin{aligned}
\Psi_n\left(\{y(x)\}, \{\tilde{y}(\theta)\}; \lambda_c\right) \quad = \quad & \max_{\{\tilde{y}(\theta|x)\}, \{\tilde{Y}(y(x))\}, Y} \int A\left(y(x), \tilde{Y}(y(x)) \cdot \lambda_c^{\frac{1}{\kappa}} \cdot Y^{\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)}\right) g(x) dx \qquad (31) \\
& s.t. \quad \lambda_c \cdot \tilde{Y}(y(x))^{\kappa} = \kappa \cdot \frac{\eta}{\eta - 1} \cdot \int \tilde{y}(\theta|x)^{\kappa} f(\theta) d\theta \quad \forall x, \\
& \int \tilde{y}(\theta|x) g(x) dx = \tilde{y}(\theta) \quad \forall \theta, \\
& \lambda_c^{\frac{1}{\kappa}} \cdot \int \tilde{Y}(y(x)) g(x) dx = Y^{1-\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)}. \qquad (32)
\end{aligned}$$

Note that we also maximize over the aggregate composite labor Y to optimally determine the level of worker's labor as well as the distribution of worker's labor, but this composite labor Y should be consistent with the aggregation $\tilde{Y}(y(x))$ over the entrepreneurs which requires additional constraint (32) in the aggregate production maximization. We also remark that with the aggregate production (31) in terms of normalized variables, the effect of increase in $\tilde{y}(\theta)$ on the aggregate production Ψ_n through the change in λ_c—through the change in labor cost—is zero as $\frac{\partial \Psi_n}{\partial \lambda_c} = 0$. Then the incidence of a tax reform on production can be represented without cross price elasticities, which simplifies the optimal tax formula greatly.

From the first order condition of this planning problem, we can derive an alternative optimal tax formula expressed in terms of the normalized variables which characterizes the same optimal tax schedule as the formula (29) does. The decomposition of the alternative formula provides much clear understanding in the role of superstar effects and general equilibrium effects. We present this in the next proposition.

Proposition 5. *Assume that each firm's aggregation of worker's labor takes the form of (24), and the worker's preferences have no income effects in $n(\theta)$ and $e(\theta)$. For any Pareto weights*

$\{\Lambda_\theta(\theta)\}$, *the optimal tax schedule for workers satisfies*

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))} + \frac{1}{1-T'(\omega(\theta))}\cdot\Omega_1 = \frac{1-F(\theta)}{\theta f(\theta)}\cdot\frac{\hat{\mu}(\theta)}{1-F(\theta)}\cdot EEC_{GE}(\theta)^{-1}, \quad \forall\theta \tag{33}$$

where

$$\Omega_1 = \frac{\kappa\eta}{\eta-1}\cdot\frac{\int Y(y(x))^\kappa g(x)dx}{Y^\kappa} - 1$$

$$EEC_{GE}(\theta)^{-1} = \underbrace{\frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)}\cdot\left(\frac{w_e e}{w}(\theta)+\frac{w_\theta\theta}{w}\right)}_{=EEC(\theta)^{-1}} + \Omega_{GE}, \tag{34}$$

$$\Omega_{GE}(\theta) = y(\theta)\frac{d\left(\frac{w_e e}{w}(\theta)\right)}{dy(\theta)} + \left(-\frac{1+\tilde{\epsilon}^u(\tilde{\theta})}{\tilde{\epsilon}^c(\tilde{\theta})} - \frac{U_{cy}(\tilde{\theta})y(\tilde{\theta})}{U_c(\tilde{\theta})}\right)\cdot\epsilon^o_{w(\tilde{\theta}),y(\theta)}\cdot\left(\frac{w_e e}{w}(\theta)+\frac{w_\theta\theta}{w}\right). \tag{35}$$

The optimal tax formula (33) is different from the standard Mirrlees formula because our model deviates from the canonical Mirrlees model in several aspects: (i) wage structure with endogenous human capital, (ii) general equilibrium with endogenous wage schedule, (iii) imperfect substitution of worker's labor in production (iv) superstar effects for workers with $\kappa \geq 1$, (v) superstar effects for entrepreneurs with $A_{yY} > 0$, and (vi) assignment of workers to firms. To understand the role of each component in the optimal tax schedule, we analyze the formula from the stripped-down version and investigate mechanisms by adding each component.

Role of each Component: (1) Wage with Human Capital We start with the special case which shuts down the components (ii)–(vi) mentioned above but allows general wage structure with human capital. Consider an economy which consists of workers only (without entrepreneurs). With no general equilibrium effects (exogenous prices $w(e)$) and no superstar effect for workers ($\kappa = 1$), $\Omega_1 = \Omega_2(\theta) = 0$ and $EEC_{GE}(\theta)^{-1} = EEC(\theta)^{-1}$. Next proposition provides the tax formula for workers in this special case.

Proposition 6. *Suppose that the economy consists of only workers and the the wage schedule* $\{w(e)\}$ *is exogenously given. Assume that* $\kappa = 1$. *Then the optimal tax schedule satisfies*

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))} = \frac{1-F(\theta)}{\theta f(\theta)}\cdot EEC(\theta)^{-1}\cdot\frac{\hat{\mu}(\theta)}{1-F(\theta)}, \quad \textit{where} \tag{36}$$

$$EEC(\theta)^{-1} = \frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)}\cdot\left(\frac{w_e e}{w}(\theta)+\frac{w_\theta\theta}{w}\right).$$

The formula (36) is equivalent to the standard ABC-formula in the Mirrlees model except that the term of the effective efficiency depends on the returns to skill and human capital as well as the elasticity of labor supply. In the standard Mirrlees economy without endogenous human capital, $p(\theta, e(\theta)) = \theta$, and thus $\frac{w_e e}{w}(\theta) + \frac{w_\theta \theta}{w} = 1$ which implies $EEC^{-1} = \frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)}$. In the presence of general wage structure with human capital, however, the efficiency cost of tax distortion EEC is determined not only by the elasticity of labor supply but also returns to skill and human capital. With a higher return to either skill or human capital, it is easier for the high skilled type to mimic the low skilled type, which tightens the incentive constraint. Since the tax distortion can relax the incentive constraint, the efficiency cost of tax distortion is lowered when these returns are higher.

Role of each Component: (2) Imperfect Substitution and Superstar Next, we extend the analysis by incorporating the complementarity in production and superstar effects—(iii),(iv),(v), and (vi) mentioned above—while still turning off the general equilibrium. That is, we still assume that the exogenous wage schedule $\{w(e)\}$ is given, but we now introduce imperfect substitution of workers' labor in production ($\eta < \infty$) and superstar effects for both workers ($\kappa > 1$) and entrepreneurs ($A_{yY} > 1$) with the structure of assigning workers to firms. Next proposition shows how these added features of production and superstar effects change the optimal tax schedule.

Proposition 7. *Consider an economy with both entrepreneurs and workers described in Section 2, but we assume that the wage schedule $\{w(e)\}$ is exogenously given. Then the optimal tax schedule for workers satisfies*

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))} \quad + \quad \frac{1}{1-T'(\omega(\theta))} \cdot \Omega_1 = \frac{1-F(\theta)}{\theta f(\theta)} \cdot EEC(\theta)^{-1} \cdot \frac{\hat{\mu}(\theta)}{1-F(\theta)}, \quad \forall \theta. \tag{37}$$

And the optimal tax schedule for entrepreneurs satisfies

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))} \quad = \quad \frac{1-G(\theta)}{xg(\theta)} \cdot \frac{1+\tilde{\epsilon}^u(x)}{\tilde{\epsilon}^c(x)} \cdot \frac{\hat{\tilde{\mu}}(x)}{1-G(x)}, \quad \forall x. \tag{38}$$

We first remark that the optimal tax formula for entrepreneurs (38) is the same with the standard ABC-formula in the Mirrlees model despite the superstar effects of the entrepreneurs. This neutrality of the superstar effects is consistent with the result in Scheuer

and Werning (2017). Introduction of the superstar effects raises both the elasticity of earnings and earnings inequality, but these two effects exactly cancel out and thus the tax formula is not changed at all when it is expressed in terms of the skill distribution. Although our economy has the wage structure with endogenous human capital, it does not affect the incentive constraint of the entrepreneurs as we assume that only workers invest in human capital. Thus, the effective efficiency cost of entrepreneurs is simply determined by the entrepreneurs' elasticities of labor supply.

On the other hand, with the introduction of the imperfect substitution and superstar effects, the optimal tax formula for workers (37) has additional positive term $\frac{1}{1-T'(\omega(\theta))} \cdot \Omega_1$ relative to the standard ABC-formula, which lowers optimal marginal tax rate. By rewriting (37) in the following way, we can more clearly see that $\Omega_1 > 0$ lowers the optimal tax rate by making the marginal tax rate T' more sensitive to efficiency cost:

$$T'(\omega(\theta)) = \frac{RHS(\theta) - \Omega_1}{1 + RHS(\theta)}, \quad \text{where} \quad RHS(\theta) = \frac{1 - F(\theta)}{\theta f(\theta)} \cdot EEC(\theta)^{-1} \cdot \frac{\hat{\mu}(\theta)}{1 - F(\theta)}.$$

Why do these added features in production lower the optimal marginal tax rate? To understand the mechanisms, we decompose the new term Ω_1 into three additive terms—each of them captures different channel—as follows:

$$\Omega_1 = \left(\frac{\int Y(y(x))^{\kappa} g(x) dx}{Y^{\kappa}} - 1 \right) + \Lambda_s \cdot \frac{(\kappa - 1)\eta}{\eta - 1} + \Lambda_s \cdot \frac{1}{\eta - 1},$$

where $\Lambda_s = \frac{\int Y(y(x))^{\kappa} g(x) dx}{Y^{\kappa}} \geq 1$. An immediate observation from the decomposition is that all three terms are positive, thus all three channels provide forces to lower the optimal tax rate. Looking into each channel, the first term $\left(\frac{\int Y(y(x))^{\kappa} g(x) dx}{Y^{\kappa}} - 1 \right)$ arises as the tax reform can affect the allocative efficiency. In the presence of entrepreneur's superstar effects with $A_{yY} > 0$, increase in worker's labor supply is likely to be assigned to more productive firms with higher x, improving the allocative efficiency. This channel of allocation raises the efficiency cost of tax distortion and it is magnified when labor hiring cost is convex with $\kappa > 1$. The second term $\Lambda_s \cdot \frac{(\kappa-1)\eta}{\eta-1}$ is positive when $\kappa > 1$. Convex earnings schedule of the workers with $\kappa > 1$ makes hiring worker's labor more costly, which also magnifies the efficiency cost of tax distortion especially more when labor inputs are less substitutable. The third term $\Lambda_s \cdot \frac{1}{\eta-1}$ arises as the workers with heterogeneous skill types (and thus with different levels

of human capital) hired in the same firm are imperfectly substitutable. Reduction in labor supply caused by a tax distortion is more costly when labor inputs are less substitutable (as decrease in labor supply of skill type θ also reduces marginal products of other skill types.) Note that the second and third channels are amplified by $\Lambda_s > 1$.

Thus, the superstar effects for workers in our economy have non-neutrality in optimal tax schedule, which is different from the result in Scheuer and Werning (2017). What are the fundamental differences in two economies which lead to different implications? In our economy, many workers with different skills are assigned to the same firm and they are imperfectly substitutable, while Scheuer and Werning (2017) assumes the one-to-one assignment between firm and entrepreneurs. The convex earnings schedule of workers alone does not break the neutrality, but when it is combined with many-to-one assignment structure, the worker's earnings schedule affects the entrepreneur's cost of labor hiring, which is the key reason why the cost of tax distortion increases with the superstar effects in our model.

Role of each Component: (3) General Equilibrium Lastly, we incorporate the endogenous prices and analyze the general equilibrium effects. In an economy with the endogenous prices, the effect of a tax reform on the incentive constraints through the price change should be also considered. Comparison of the formula without general equilibrium effects (37) to the formula with general equilibrium effects (33) shows that when the general equilibrium is introduced, the optimal tax formula has an additional term $\Omega_{GE} = EEC_{GE}(\theta)^{-1} - EEC(\theta)^{-1}$ expressed in (35) in the inverse of the effective efficiency cost. This new term exactly captures how the change in prices driven by a tax reform tightens or relaxes the worker's incentive constraints. In which direction the addition of Ω_{GE} changes the optimal marginal tax rate depends on the signs of price elasticities. As we investigate below, the signs of these elasticities can be either positive or negative, thus the implication of the general equilibrium can be either direction. We investigate this further in Section 4 with a simple example.

For now, we just describe the mechanisms of the general equilibrium effects without determining its directional implication on the optimal tax schedule. Regardless of the introduction of the endogenous prices, the optimal marginal tax rate is still decreasing in effective efficiency cost of tax distortion (EEC_{GE}), but now the cost includes the new term Ω_{GE} due

to the general equilibrium effects. Whether this change increases or decreases the effective efficiency cost depends on the sign of the price elasticities. Note that the new term can be additively decomposed into two terms: $\Omega_{GE} = \Omega_{GE,1} + \Omega_{GE,2}$, where

$$\Omega_{GE,1} = y(\theta)\frac{d\left(\frac{w_e e}{w}(\theta)\right)}{dy(\theta)}, \quad \text{and} \quad \Omega_{GE,2} = \left(-\frac{1+\tilde{\epsilon}^u(\tilde{\theta})}{\tilde{\epsilon}^c(\tilde{\theta})} - \frac{U_{cy}(\tilde{\theta})y(\tilde{\theta})}{U_c(\tilde{\theta})}\right)\cdot\epsilon^o_{w(\tilde{\theta}),y(\theta)}\cdot\left(\frac{w_e e}{w}(\theta) + \frac{w_\theta \theta}{w}\right),$$

which capture the effects of change in rate of return $\frac{w_e \cdot e}{w}$ and the change in level of wage $w(e(\theta))$ respectively, and each term reflects different channel of general equilibrium.

The general equilibrium affects optimal tax schedule through the following two channels. First, since higher return to human capital implies tighter incentive constraints—as it makes the high skilled easier to mimic the low skilled, the optimal tax schedule needs to be adjusted to the direction that reduces the return to human capital to relax the incentive constraints. This channel is captured by the term $\Omega_{GE,1}$ which includes the tax incidence on rate of return $\frac{w_e \cdot e}{w}$. For example, if a tax reform of increasing labor supply $\tilde{y}(\theta)$—which reduces the marginal tax rate—raises the return $\frac{w_e e}{w}$, this implies that an increase in marginal tax rate relaxes the incentive constraints, and the first channel of general equilibrium provides a force to increase optimal tax rate

Second, the general equilibrium can also affect how effective the tax distortion is in relaxing the incentive constraints, and the second term $\Omega_{GE,2}$ captures this channel. Given prices, the planner uses the tax wedges to relax the incentive constraints as taxing the earnings of the low skilled can prevent workers with higher skills to mimic the lower types. The effectiveness of tax distortion in achieving this goal depends on the direction of the price change when general equilibrium is introduced. Consider the case with $\epsilon^o_{w(\tilde{\theta}),y(\theta)} > 0$. This implies that when the planner increases θ type's tax rate—decreases retention rate—to relax the incentive constraint, wage rate of the same type decreases in response to this tax reform. This price change makes other skill types easier to mimic θ-type, making the tax distortion less effective in relaxing the incentive constraints. Thus positive $\epsilon^o_{w(\tilde{\theta}),y(\theta)}$ leads to a higher efficiency cost of tax distortion relative to the cost without general equilibrium and lowers the optimal marginal tax rate—vice versa with negative $\epsilon^o_{w(\tilde{\theta}),y(\theta)}$.

Summing up, $\frac{d\left(\frac{w_e e}{w}(\theta)\right)}{dy(\theta)} > 0$ and $\epsilon^o_{w(\tilde{\theta}),y(\theta)} < 0$ implies that increasing marginal tax rate can relax the incentive constraints. Thus under these signs the tax incidence, general equilibrium

provides a force to raise optimal marginal tax rate with a lowered efficiency cost of tax distortion.

3.4 Comparison of the Formulas with and without Normalization

We have focused on the optimal tax formulas expressed with normalization ((33) and (**??**)) to understand the underlying mechanisms of the each component of the model. By comparing the formulas, we can see that the terms including the cross-price elasticities in formulas without normalization do not appear in the formulas with normalization. These formulas with and without normalization, however, characterize the same optimal tax schedule. Then why don't we see the cross price elasticities in the formula with normalization? We briefly discuss this in this section.

By comparing the formulas term by term, we can show that the new terms arising due to the imperfect substitution and superstar effect can be expressed alternatively with price elasticities. Next corollary provides this alternative expression.

Corollary 8. *Assume that each firm's aggregation of worker's labor takes the form of (24), and the worker's preferences have no income effects in $n(\theta)$ and $e(\theta)$. The second and third channels in the term Ω_1 of the worker's formula (33) can be alternatively expressed as follows:*

$$
\begin{aligned}
& -\left(\frac{(\kappa-1)\eta}{\eta-1}+\frac{1}{\eta-1}\right)\cdot\frac{\int Y(y(x))^{\kappa}g(x)dx}{Y^{\kappa}} \\
= \;& \frac{\int Y(y(x))^{\kappa}g(x)dx}{Y^{\kappa}}\cdot\left(\int\frac{y(\tilde{\theta})^{\kappa}}{y(\theta)^{\kappa}}\left(\epsilon_{w(e(\tilde{\theta})),y(\theta)}-\frac{\psi\eta}{\eta-1}\epsilon_{e(\tilde{\theta}),y(\theta)}\right)\frac{f(\tilde{\theta})}{f(\theta)}d\tilde{\theta}\right) \\
& -\frac{1}{\theta f(\theta)}\int\hat{\mu}(\theta)\left(1-T'(\omega(\tilde{\theta}))\right)\frac{y(\tilde{\theta})^{\kappa}/\tilde{\theta}}{y(\theta)^{\kappa}/\theta}\left(\frac{1+\tilde{\epsilon}(\tilde{\theta})}{\tilde{\epsilon}(\tilde{\theta})}+\frac{U_{cy}(\tilde{\theta})y(\tilde{\theta})}{U_c(\tilde{\theta})}\right)\epsilon^{o}_{w(e(\tilde{\theta})),y(\theta)}\left(\frac{w_e e}{w}(\tilde{\theta})+\frac{w_{\tilde{\theta}}\tilde{\theta}}{w}\right)d\tilde{\theta}.
\end{aligned}
$$

Corollary 8 shows that the second and third channels in the new term Ω_1 of the formula with normalization can be represented in terms of cross-price elasticities, although these terms arise not because of the general equilibrium effects but because of the imperfect substitution and superstar effects. This is because in a general equilibrium with endogenous price, the prices are determined by the marginal product of labor inputs, and thus any change in production can be represented as the change in prices. However, as we have discussed above,

the fundamental reason of having a new term Ω_1 is the complementarity in production and superstar effects, and the same term shows up even when we assume exogenous prices.

4 Simple Model

In Section 3, we have provided optimal tax formulas and have investigated the role of superstar effects and general equilibrium in optimal tax schedule. For the role of human capital and superstar effects, we could have relatively sharp characterization. Regarding the role of general equilibrium, however, the implication of the general equilibrium on the tax schedule is much more involved, as the formulas are expressed in terms of highly endogenous price elasticities whose signs are not deterministic. In this section, we focus on the simple example with linear tax where the endogenous human capital prices can be solved analytically to characterize the role of general equilibrium in optimal tax.

4.1 Environment of the Simple Model

Suppose that the income tax schedule is linear for both workers and entrepreneurs. The government sets a constant tax rate τ_w for workers and a constant tax rate τ_e for entrepreneurs. Under this linear tax system, by assuming special preferences, we can derive analytical expression for the price elasticities. Then we can analyze in which direction the optimal tax rate is changed in the presence of general equilibrium.

Workers We assume the following simple form of worker's preferences with no income effects in labor supply and human capital investment:

$$u\left(c, \frac{e}{\theta}, n\right) = c - \zeta \cdot \left(\frac{e}{\theta}\right)^{\frac{1}{\zeta}} - \sigma \cdot n^{\frac{1}{\sigma}}, \tag{39}$$

where ζ controls the elasticity of human capital investment and σ determines the elasticity of working hours. We solve for the price of human capital by the "guess and verify". We guess the functional form of the price of human capital as follows:

$$w(e) = \Psi(\tau) \cdot e^{\nu}, \tag{40}$$

Given the functional form (40), we solve the worker's problem of maximizing utility (39) by choosing human capital investment e and working hours n subject to the budget constraint $c = (1-\tau_w)\cdot(\theta^\gamma \cdot w(e))^\kappa + Tr$, where Tr is lumpsum transfer from the government. Using the first order conditions with respect to e and n, we can solve for the optimal decision of e and n given Ψ and ν:

$$e(\theta) = \left((1-\tau_w)\cdot\kappa\cdot\Psi^\kappa\cdot\nu^{1-\kappa\sigma}\right)^{\frac{\zeta}{1-\kappa(\sigma+\nu\zeta)}}\cdot\theta^{\frac{1-\kappa(\sigma-\gamma\zeta)}{1-\kappa(\sigma+\nu\zeta)}}, \tag{41}$$

$$n(\theta) = ((1-\tau_w)\cdot\kappa)^{\frac{\sigma}{1-\kappa\sigma}}\cdot(\Psi\cdot e(\theta))^{\frac{\kappa\sigma}{1-\kappa\sigma}}\cdot\theta^{\frac{\kappa\sigma\gamma}{1-\kappa\sigma}}. \tag{42}$$

We denote the inverse of $e(\theta)$ by $\theta(e)\equiv e^{-1}(e)$. Then the labor supply $\ell_S(e)=\theta(e)^\gamma\cdot n(\theta(e))$ can be written as a function of human capital level e:

$$\ell_S(e) = ((1-\tau_w)\cdot\kappa\cdot\Psi^\kappa)^{\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}\cdot\nu^{-\frac{\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}\cdot e^{\frac{\kappa\sigma\nu+\gamma-\kappa\gamma\nu\zeta}{1-\kappa(\sigma-\gamma\zeta)}}. \tag{43}$$

Entrepreneurs We also assume that the entrepreneurs have the following form of preferences with no income effects in labor supply:

$$\tilde{u}(c,n) = c - \sigma\cdot n^{\frac{1}{\sigma}}, \tag{44}$$

which implies the same elasticity of working hours with the worker's elasticity.

For simplicity, we make a further assumption on the entrepreneur's technology. The production function of the entrepreneur has the following form

$$A(y,Y) = a\cdot y\cdot Y, \tag{45}$$

where y is the entrepreneur's effective labor supply and Y is the composite labor which combines worker's labor with the CES aggregation technology (24). Note that (45) satisfies $A_{yY}=a>0$, generating the superstar effects of the entrepreneurs.

In this simple model, the labor supply elasticities which appear in the optimal tax formulas can be expressed in a closed form. Without income effects, there is no distinction between compensated and uncompensated elasticities, and we denote the elasticities of effective labor supply (given prices) of the workers and entrepreneurs by

$$\tilde{\epsilon}^W(\theta) \equiv \frac{d\log y(\theta)}{d\log(1-\tau_w)}, \quad \text{and} \quad \tilde{\epsilon}^E(x) \equiv \frac{d\log y(x)}{d\log(1-\tau_e)}.$$

Under the special preferences (39) and (44), they have the following closed forms (see the Appendix):

$$\tilde{\epsilon}^W(\theta) = \frac{\sigma + \nu\zeta}{1 - \kappa(\sigma + \nu\zeta)}, \quad \tilde{\epsilon}^E(x) = \frac{\sigma(\kappa - 1)}{(1 - \sigma)\kappa - 1}.$$

Note that they are constant over θ and over x respectively, and thus we simply denote these labor supply elasticities by $\tilde{\epsilon}^W$ and $\tilde{\epsilon}^E$.

Equilibrium Prices As we have described in Section 2, by aggregating the solution of the cost-minimization problem (8), we obtain the labor demand curve $\ell_D(e)$ for each e given λ_c and Y_D:

$$w(e) = \lambda_c^{\frac{1}{\kappa}} \cdot Y_D^{\frac{1}{\kappa}\left(\frac{1}{\eta} + \kappa - 1\right)} \cdot \left(\frac{e^{\psi}}{\kappa}\right)^{\frac{1}{\kappa}} \cdot \ell_D(e)^{-\frac{1}{\kappa}\left(\frac{1}{\eta} + \kappa - 1\right)}, \tag{46}$$

where $\lambda_c^{\frac{1}{\kappa}} \cdot Y_D^{\frac{1}{\kappa}\left(\frac{1}{\eta} + \kappa - 1\right)}$ is the aggregate demand-shifter which still depends on the endogenous λ_c and Y_D.[8] Thus we need to solve for λ_c and Y_D to characterize the price completely.

Under the preferences (44), we can derive the closed-form solution for the aggregate demand Y_D by aggregating the solution $Y(y(x))$ of the entrepreneur's problem (see the Appendix):

$$Y_D = a^{\frac{1}{(1-\sigma)\kappa - 1}} \cdot \lambda_c^{-\frac{1-\sigma}{(1-\sigma)\kappa - 1}} \cdot (1 - \tau_e)^{\frac{\sigma}{(1-\sigma)\kappa - 1}} \cdot \int x^{\frac{1-\sigma}{(1-\sigma)\kappa - 1}} g(x) dx.$$

In equilibrium, λ_c which is function of wage rates is determined to satisfy the market clearing $Y_D = Y_S$, and by substituting λ_c and Y_D into (46), we obtain the labor demand curve $\ell_D(e)$ for each human capital level e.

In equilibrium, labor markets should clear for each human capital level e, which implies $\ell_D(e) = \ell_S(e)$ for every e. Substituting $\ell_S(e)$ to the equation of the labor demand curve leads

[8] Demand curve shifts to the right if either Y_d or λ_c increases. Increase in aggregate labor demand (Y_D) induces increase in demand for each effective labor input $\ell_D(e)$ for all levels of human capital due to the complementarity. On the other hand, increase in marginal cost of aggregate labor (λ_c) given price $w(e)$ makes the labor input $\ell_D(e)$ relatively cheaper, leads to more demand of $\ell_D(e)$.

to the following expression for the equilibrium price of human capital (see the Appendix):

$$
\begin{aligned}
w(e) &= B \times e^{\frac{\psi}{\kappa}-\frac{1}{\kappa}\cdot\left(\frac{1}{\eta}+\kappa-1\right)\cdot\frac{\kappa\sigma\nu+\gamma-\kappa\sigma\nu\zeta}{1-\kappa(\sigma-\gamma\zeta)}}, \quad \text{where} \qquad (47)\\
B &= (1-\tau_e)^{\frac{\sigma}{\kappa}\cdot\frac{\eta-1}{\eta}} \times (1-\tau_w)^{\frac{\sigma(\eta-1)+1}{\eta}\cdot\frac{\sigma+\nu\zeta}{1-\kappa(\sigma+\nu\zeta)}-\frac{1}{\kappa}\cdot\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}\\
&\times \quad \Psi^{\frac{1}{\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)}-\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}} \times \quad \nu^{\frac{\nu\zeta}{\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)}+\frac{1}{\kappa}\cdot\left(\frac{1}{\eta}+\kappa-1\right)\frac{\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}\\
&\times \quad a^{\frac{\eta-1}{\kappa\eta}} \times \quad \kappa^{\frac{1}{\kappa\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)}-\frac{1}{\kappa}-\frac{1}{\kappa}\cdot\left(\frac{1}{\eta}+\kappa-1\right)\cdot\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}\\
&\times \left(\int x^{\frac{1-\sigma}{(1-\sigma)\kappa-1}}g(x)dx\right)^{\frac{(1-\sigma)\kappa-1}{\kappa}\cdot\frac{\eta-1}{\eta}} \times \left(\int \theta^{\frac{\kappa(\gamma+\nu)}{1-\kappa(\sigma+\nu\zeta)}}f(\theta)d\theta\right)^{\frac{\sigma(\eta-1)+1}{\kappa\eta}}.
\end{aligned}
$$

Note that our guess of the functional form is verified by (47). We can also obtain the analytical expression of the price by equating $w(e) = \Psi \cdot e^{\nu}$ and (47). From $\Psi = B$, the closed form of the coefficient Ψ yields

$$
\begin{aligned}
\Psi &= (1-\tau_e)^{\frac{1}{D}\cdot\frac{\sigma}{\kappa}\cdot\frac{\eta-1}{\eta}} \times (1-\tau_w)^{\frac{1}{D}\cdot\left(\frac{\sigma(\eta-1)+1}{\eta}\cdot\frac{\sigma+\nu\zeta}{1-\kappa(\sigma+\nu\zeta)}-\frac{1}{\kappa}\cdot\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}\right)}\\
&\times \nu^{\frac{1}{D}\left(\frac{\nu\zeta}{\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)}+\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)\frac{\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}\right)} \times a^{\frac{1}{D}\cdot\frac{\eta-1}{\kappa\eta}} \times \kappa^{\frac{1}{D}\left(\frac{1}{\kappa\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)}-\frac{1}{\kappa}-\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}\right)}\\
&\times \left(\int x^{\frac{1-\sigma}{(1-\sigma)\kappa-1}}g(x)dx\right)^{\frac{1}{D}\cdot\frac{(1-\sigma)\kappa-1}{\kappa}\cdot\frac{\eta-1}{\eta}} \times \left(\int \theta^{\frac{\kappa(\gamma+\nu)}{1-\kappa(\sigma+\nu\zeta)}}f(\theta)d\theta\right)^{\frac{1}{D}\cdot\frac{\sigma(\eta-1)+1}{\kappa\eta}}, \qquad (48)
\end{aligned}
$$

where $D = 1 - \frac{1}{\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)} + \left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}$. In addition, the exponent ν is the solution that satisfies the equation

$$
\nu = \frac{\psi}{\kappa} - \frac{1}{\kappa}\cdot\left(\frac{1}{\eta}+\kappa-1\right)\cdot\frac{\kappa\sigma\nu+\gamma-\kappa\sigma\nu\zeta}{1-\kappa(\sigma-\gamma\zeta)}. \qquad (49)
$$

4.2 Price Elasticities

We have derived the analytical expression of the price of human capital with the simplifying assumptions. Given this expression, we can further analyze the tax incidence on the endogenous wage, which is the key for the general equilibrium effects. We first characterize the price elasticities and then investigate its implications on the optimal tax schedule.

First of all, we remark that the equation (49) for ν does not include the tax rates τ_w and τ_e. That is, under the linear tax system and the CES aggregation technology, the return to human capital investment $\frac{w_e \cdot e}{w}$, which is equal to the exponent ν of the human capital price, does not depend on the tax rates and it is completely determined by the exogenously given

parameters. Thus the terms associated with the incidence of $\frac{w_e \cdot e}{w}$ in the optimal tax formula disappears under this simple example, and we only need to consider the responses of Ψ and e to the tax reform to analyze the general equilibrium effects.

In the optimal tax formulas (33) and (**??**), the terms capturing the general equilibrium effects include the elasticities with respect to effective labor supply. Note that these price elasticities can be rewritten as the elasticities with respect to the retention rates multiplied by the inverse of the labor supply elasticities:

$$\epsilon_{w(e(\theta)),y(\theta)} = \frac{d\log w(e(\theta))}{d\log(1-\tau_w)} \cdot \frac{1}{\frac{d\log y(\theta)}{d\log(1-\tau_w)}}, \quad \epsilon_{w(e(\theta)),y(x)} = \frac{d\log w(e(\theta))}{d\log(1-\tau_e)} \frac{1}{\frac{d\log y(\theta)}{d\log(1-\tau_w)}}.$$

Similarly, the elasticities of human capital can be rewritten as:

$$\epsilon_{e(\theta),y(\theta)} = \frac{d\log e(\theta)}{d\log(1-\tau_w)} \cdot \frac{1}{\frac{d\log y(\theta)}{d\log(1-\tau_w)}}, \quad \epsilon_{e(\theta),y(x)} = \frac{d\log e(\theta)}{d\log(1-\tau_e)} \cdot \frac{1}{\frac{d\log y(\theta)}{d\log(1-\tau_w)}}.$$

Thus to analyze the role of the price elasticities in the optimal tax, we need to characterize the elasticities with respect to the retention rate.

With the price schedule in the form of $w(e) = \Psi \cdot e^{\nu}$, note that the price elasticities—the incidence of a tax reform on the prices—can be decomposed into:

$$\frac{d\log w(e(\theta))}{d\log(1-\tau_j)} = \frac{d\log \Psi}{d\log(1-\tau_j)} + \nu \cdot \frac{d\log e(\theta)}{d\log(1-\tau_j)}, \quad j \in \{w, e\}. \tag{50}$$

The first term in (50) is the elasticity of human capital price $w(e)$ with respect to retention rate $(1-\tau_j)$ holding the level of human capital at e, which is constant for all e's under the wage schedule in the form of (40). We denote this constant elasticity by $\epsilon_{w,1-\tau_j}|_e$ for $j \in \{w, e\}$. The second term in (50) captures the change in price through the change in human capital. We can further decompose the elasticity of human capital with respect to worker's retention rate $(1-\tau_w)$ into the response of $e(\theta)$ given Ψ and through change in Ψ:

$$\begin{aligned} \frac{d\log e(\theta)}{d\log(1-\tau_w)} &= \epsilon_{e,1-\tau_w}|_{w(\cdot)} + \frac{\kappa\zeta}{1-\kappa(\sigma+\nu\zeta)} \times \epsilon_{w,1-\tau_w}|_e \\ &= \frac{\zeta}{1-\kappa(\sigma+\nu\zeta)} + \frac{\kappa\zeta}{1-\kappa(\sigma+\nu\zeta)} \times \epsilon_{w,1-\tau_w}|_e. \end{aligned} \tag{51}$$

On the other hand, the elasticity of human capital with respect to entrepreneur's retention rate $(1-\tau_e)$ only consists of the the response through change in Ψ:

$$\epsilon_{e,1-\tau_e} = \frac{d\log e(\theta)}{d\log(1-\tau_e)} = \frac{\kappa\zeta}{1-\kappa(\sigma+\nu\zeta)} \times \epsilon_{w,1-\tau_e}|_e. \tag{52}$$

Summing up, we can characterize the terms reflecting the price incidence in the optimal tax formulas. Combining our derivation above, the price elasticities with respect to worker's labor supply and entrepreneur's labor supply are represented as follows:

$$\epsilon^W_{w,y} = \epsilon_{w,1-\tau_w} \times \frac{1}{\tilde{\epsilon}^W} = \left(\left(1+\nu\frac{\kappa\zeta}{1-\kappa(\sigma+\nu\zeta)}\right)\cdot \epsilon_{w,1-\tau_w}|_e + \nu\cdot\epsilon_{e,1-\tau_w}|_{w(\cdot)}\right)\times\frac{1}{\tilde{\epsilon}^W} \quad (53)$$

$$\epsilon^E_{w,y} = \epsilon_{w,1-\tau_e} \times \frac{1}{\tilde{\epsilon}^E} = \left(1+\nu\frac{\kappa\zeta}{1-\kappa(\sigma+\nu\zeta)}\right)\cdot \epsilon_{w,1-\tau_e}|_e \times\frac{1}{\tilde{\epsilon}^E}. \quad (54)$$

Note that since these elasticities are constant over θ's of the workers and over x's of the entrepreneurs, we omit θ and x from the notation. Different from $\epsilon^E_{w,y}$, the elasticity $\epsilon^W_{w,y}$ includes the change in price through the change in human capital holding the price fixed ($\epsilon_{e,1-\tau_w}|_{w(\cdot)}$). We denote the price elasticity net of this change by $\epsilon^{W,o}_{w,y}$:

$$\epsilon^{W,o}_{w,y} = \epsilon^W_{w,y} - \nu\epsilon_{e,1-\tau_w}|_{w(\cdot)} \times \frac{1}{\tilde{\epsilon}} = \left(1+\nu\frac{\kappa\zeta}{1-\kappa(\sigma+\nu\zeta)}\right)\cdot\epsilon_{w,1-\tau_w}|_e\times\frac{1}{\tilde{\epsilon}_w}, \quad (55)$$

which is the elasticity appearing in the worker's tax formula to capture the general equilibrium effects.

Note that the sign of the general equilibrium effect terms are essentially determined by the sign of $\epsilon_{w,1-\tau_j}|_e$, $j\in\{e,w\}$, and thus we provide analytical expression of $\epsilon_{w,1-\tau_j}|_e$ from the closed-form solution of Ψ in (48):

$$\epsilon_{w,1-\tau_w}|_e = \frac{d\log\Psi}{d\log(1-\tau_w)} = \frac{\frac{\sigma(\eta-1)+1}{\eta}\cdot\frac{\sigma+\nu\zeta}{1-\kappa(\sigma+\nu\zeta)} - \frac{1}{\kappa}\cdot\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}{1-\frac{1}{\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)}+\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}, \quad \text{and} \quad (56)$$

$$\epsilon_{w,1-\tau_e}|_e = \frac{d\log\Psi}{d\log(1-\tau_e)} = \frac{\frac{\sigma}{\kappa}\cdot\frac{\eta-1}{\eta}}{1-\frac{1}{\eta}\cdot\frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)}+\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)}}. \quad (57)$$

The signs of these elasticitis can be either positive or negative depending on primitive parameters, which we further analyze below.

4.3 Implications of the General Equilibrium Effects in Optimal Tax

We now analyze the role of the general equilibrium in the optimal tax formulas. Under the linear tax restriction, the planner can only choose constant tax rate (each for workers and entrepreneurs), thus the worker's optimal tax formulas can now be obtained from a tax

perturbation of increasing $\tilde{y}(\theta)$ for all skill type θ and the entrepreneur's optimal tax formulas is obtained from a tax reform of increasing $y(x)$ for all productivity type x. Next proposition provides the optimal linear tax formulas which are the counterpart of the nonlinear formulas (33) and (**??**).

Proposition 9. *Assume linear tax system for both workers and entrepreneurs. Then the optimal tax rate for workers τ_w satisfies*

$$\begin{aligned}
\frac{\tau_w}{1-\tau_w} \quad + \quad \frac{1}{1-\tau_w}\cdot\Omega_1 &= \int \frac{\hat{\mu}(\theta)}{\theta f(\theta)}\cdot w_{super}(\theta) f(\theta) d\theta \times EEC_{GE}^{-1}, \quad \textit{where} \qquad (58)\\
\Omega_1 &= \frac{\kappa\eta}{\eta-1}\cdot\frac{\int Y(y(x))^{\kappa} g(x)dx}{Y^{\kappa}} - 1\\
w_{super}(\theta) &= \frac{W'(y(\theta))}{\int W'(y(\tilde{\theta}))f(\tilde{\theta})d\tilde{\theta}}\\
EEC_{GE}^{-1} &= \underbrace{\frac{1+\tilde{\epsilon}^W}{\tilde{\epsilon}^W}\cdot(\gamma+\nu)}_{=EEC^{-1}} \quad - \quad \Omega_{GE}\\
\Omega_{GE} &= \frac{1+\tilde{\epsilon}^W}{\tilde{\epsilon}^W}\cdot\epsilon^{W,o}_{\tilde{w},\tilde{y}}\cdot(\gamma+\nu)\,. \qquad (59)
\end{aligned}$$

The sign of additional term $-\Omega_{GE}$ in the effective efficiency cost is:

$$sign(-\Omega_{GE}) = -sign(\epsilon_{w,1-\tau_w}|_e).$$

If $\epsilon_{w,1-\tau_w}|_e > (<)0$, the introduction of the general equilibrium effects increases (decreases) effective efficiency cost of tax distortion (EEC_{GE}).

From Proposition 9, we can see that the linear tax formulas have similar representations to the nonlinear tax formulas. For worker's tax formula, the role of each model's component is captured by the additional terms as in the nonlinear tax formula. First, the role of the wage structure with endogenous human capital is reflected by $\gamma+\nu$ in the inverse of the effective efficiency (EEC) as in(33). Next, the role of complementarity in production and superstar effects are captured by Ω_1 and the weighting function $w_{super}(\theta)$. The term Ω_1 is exactly the same term as in the nonlinear tax. With a linear tax reform which changes tax rate of all workers, however, welfare effects of a tax reform are now integrated over all skill types with weights $\{w_{super}(\theta)\}$ increasing in skill type θ in the presence of the superstar effects ($\kappa > 1$).

More weight is put on the high skilled because a reduction in consumption is greater for the high skilled under the convex earnings schedule for the same tax rate change.

Our main analysis in this simple model is the role of the general equilibrium effects, which is captured by the term Ω_{GE} in the optimal tax formula (58). In the simple model with linear tax, the return to human capital ν does not depend on the tax rate, thus the first channel of general equilibrium ($\Omega_{GE,1}$) discussed in Section 3 does not show up in Ω_{GE} represented in (59). Thus the directional implication of the general equilibrium is entirely determined by sign of $\epsilon^{W,o}_{\tilde{w},\tilde{y}}$ which is essentially determined by the sign of $\epsilon_{w,1-\tau_w}|_e$ in this simple model. For the worker's optimal tax, if the sign of $\epsilon_{w,1-\tau_w}|_e$ is positive (negative), the introduction of the general equilibrium increases (decreases) the cost of tax distortion by making a tax distortion less (more) effective in relaxing the incentive constraints and thus optimal tax rate is lowered (raised).

From the closed forms (56) and (57), we already know that the sign of $\epsilon_{w,1-\tau_j}|_e$ is not deterministic. However, we can investigate under which condition each elasticity is positive or negative, and we explore this in the next proposition.

Proposition 10. *Consider the simple economy presented in this section.*

1. *The elasticity of wage with respect to worker's retention rate given human capital is*

$$\begin{aligned}
\epsilon_{w,1-\tau_w}|_e &= \frac{A_W}{1-B}, \quad \textit{where} \\
A_W &= A_{Wd} + A_{Ws}, \quad \textit{where} \\
&\quad A_{Wd} = \frac{\sigma(\eta-1)+1}{\eta} \cdot \frac{\sigma+\nu\zeta}{1-\kappa(\sigma+\nu\zeta)} > 0, \ \textit{and} \\
&\quad A_{Ws} = -\frac{1}{\kappa}\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)} < 0 \\
B &= B_1 + B_2, \quad \textit{where} \\
&\quad B_1 = \frac{1}{\eta} \cdot \frac{\sigma(\eta-1)+1}{1-\kappa(\sigma+\nu\zeta)} > 0 \\
&\quad B_2 = -\left(\frac{1}{\eta}+\kappa-1\right)\frac{\sigma-\gamma\zeta}{1-\kappa(\sigma-\gamma\zeta)} < 0
\end{aligned}$$

   (i) A_W can be positive or negative depending on the dominance between the demand shift A_{Wd} and supply shift A_{Ws}, and A_W is decreasing in η.

(ii) Assume that $\sigma < \frac{1}{2}$. B is increasing in ν and γ, and $B(\nu = 0, \gamma = 0) < 1$. Given $\gamma > 0$, there exists $\bar{nu}(\gamma)$ such that for $\nu < \bar{\nu}(\gamma)$, $B < 1$. We focus on the case with $1 - B > 0$ by assuming $\nu < \bar{\nu}(\gamma)$. Then $1 - B$ in the denominator of $\epsilon_{w,1-\tau_w}|_e$ either mitigates (if $B \leq 0$) or amplifies (if $0 < B < 1$) the effect of demand and supply shift represented by A_W.

2. The elasticity of wage with respect to entrepreneur's retention rate is

$$\begin{aligned}
\epsilon_{w,1-\tau_e} &= \frac{A_E}{1-B} \quad \textit{where} \\
A_E &= A_{E1} + A_{E2} = \frac{\sigma}{\kappa} \cdot \frac{\eta - 1}{\eta} > 0, \quad \textit{where} \\
&\quad A_{E1} = -\frac{\sigma}{(1-\sigma)\kappa - 1} \cdot \frac{\sigma(\eta - 1) + 1}{\eta} < 0, \ \textit{and} \\
&\quad A_{E2} = \frac{\sigma}{(1-\sigma)\kappa - 1} \frac{1}{\kappa} \left(\frac{1}{\eta} + \kappa - 1 \right) > 0
\end{aligned}$$

Positive A_E reflects that the labor demand curve shifts to the right when retention rate $1 - \tau_e$ increases. The direct shift of demand $A_{E2} > 0$ dominates $A_{E1} < 0$ which captures the indirect shift of demand through the increase in marginal cost of aggregate labor.

Proposition 10 shows that the signs of $\epsilon_{w,1-\tau_w}|_e$ and $\epsilon_{w,1-\tau_e}$ are determined by the signs of A_W and A_E which reflect the shift of labor demand and labor supply curves to the change in each retention rate respectively. The common denominator $1 - B$ of the both elasticities is positive if γ and ν are small enough, then this indirect general equilibrium force either amplifies or mitigates the direct effect by the shift of demand and supply curves without changing the direction.[9]

We further look into $\epsilon_{w,1-\tau_w}|_e$ which is the key price elasticity for the role of general equilibrium effects. The elasticity $\epsilon_{w,1-\tau_w}|_e$ can be either positive or negative because a rise in retention rate $1 - \tau_w$ shifts the demand curve to the right and shifts the supply curve to the right. Higher worker's retention rate obviously increases worker's labor supply without

[9] $B_1 > 0$ provides forces of amplification as the increase (decrease) in price driven by the shift of the demand and supply curves leads to decrease (increase) in the marginal cost of production and thus increases (decreases) demand which amplifies the price change. $B2 < 0$ provides forces of mitigation as the change in price also affects labor supply which counterbalances the initial price change.

income effects, but at the same time, increase in aggregate labor supply reduces the marginal cost of production (λ_c) and leads to higher labor demand of each e. The demand effect is stronger when labor inputs are less substitutable, and thus $A_W(\eta = 1) > 0$. As the degree of substitution (η) increases, the demand effect gets weaker and $A_W(\eta = \infty)$ can be negative.

Thus, the impact of the general equilibrium on the worker's optimal tax depends on the degree of substitution. With low degree of substitution, the demand shift in response to the change in worker's retention rate $1 - \tau_w$ is strong, and thus the $\epsilon_{w,1-\tau_w}|_e$ is positive. In this situation, if a planner increases worker's tax rate τ_w—decreasing retention rate $1 - \tau_w$—to relax the incentive constraints without considering the general equilibrium effects, the increase in tax rate has a counterbalancing impact of tightening the incentive constraints with decrease in wage rate. Then the introduction of general equilibrium reduces the effectiveness of tax distortion in relaxing the incentive constraints and thus increases the efficiency cost of tax distortion which lowers the optimal tax rate.

Next, we discuss the implication of the general equilibrium on the entrepreneur's optimal tax. Proposition 10 shows that the sign of $\epsilon_{w,1-\tau_e}$ is positive. This is because rise in entrepreneur's retention rate increases the entrepreneur's labor supply, and by the complementarity between entrepreneur's labor and worker's labor ($A_{yY} > 0$), demand for worker's labor increases.[10]

5 Quantitative Investigation - Need to Revise!!!!!!

- This section needs to be revised!!!!!

5.1 Comparative Statics with Exogenous Wage

We begin our comparative statics analysis in a model with exogenous wage as it allows clear decomposition of major forces of the effective efficiency cost. We assume a simple functional

[10]There is counterbalancing demand effect of $A_{E1} < 0$—through the change in marginal cost of production, but the direct demand effect of $A_{E2} > 0$ dominates.

form of wage schedule and we generate rise in inequality with exogenous shock to each return—return to skill, education, and effective labor.

Wage Function and Model Setup To control the return to each inequality factor with one parameter each, we consider the following functional forms for the wage schedule and productivity function. The wage schedule across effective labor 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.[11] We assume that the productivity function takes a form of $p(\theta, e) = \theta^{\gamma} \cdot e^{\nu}$. With this function, he 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. We also remark that under this Cobb-Douglas functional form, the Hicksian coefficient of complementarity between skill and education is one. Then a rise inequality—10% decrease in Pareto parameter of earnings distribution at the top—can be generated by increasing κ, γ, or ν, which represents each inequality factor.

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

$$u(c, \tilde{e}, n) = c - \zeta \cdot \tilde{e}^{\frac{1}{\zeta}} - \sigma \cdot n^{\cdot \frac{1}{\sigma}}, \tag{60}$$

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 baseline model, we take the HK-B model, in which the cost of human capital investment is decreasing in skill.

Although we will focus on the Pareto efficiency of the top income tax rate, we need to set the distribution of the entire skill types to calibrate all the necessary parameters. 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

[11] In section **??**, 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 $W(y) = y^{\kappa}$.

$\epsilon_N \sim N(\mu_\epsilon, \sigma_\epsilon^2)$ and $\epsilon_E \sim exp(\lambda_\epsilon)$ so that $\log\theta \sim EMG(\mu_\epsilon, \sigma_\epsilon^2, \lambda_\epsilon)$. Then the distribution of θ 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.

Calibration Since we calibrate parameters to match the target moments under the current US tax, we 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: $T(\omega) = \omega - \lambda^{US} \cdot \omega^{1-\tau^{US}}$.

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

$$e(\theta) = \left[\lambda^{US} \cdot (1-\tau^{US}) \cdot \kappa\right]^{\frac{\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \theta^{\frac{1-(1-\tau^{US})\kappa(\sigma-\gamma\zeta)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \nu^{\frac{\zeta(1-(1-\tau^{US})\kappa\sigma)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}}, \quad (61)$$

$$y(\theta) = \left[\lambda^{US} \cdot (1-\tau^{US}) \cdot \kappa\right]^{\frac{\sigma+\nu\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \theta^{\frac{\gamma+\nu}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \nu^{\frac{\nu\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}}, \quad (62)$$

$$\omega(\theta) = \left[\lambda^{US} \cdot (1-\tau^{US}) \cdot \kappa\right]^{\frac{\kappa(\sigma+\nu\zeta)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \theta^{\frac{\kappa(\gamma+\nu)}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}} \times \nu^{\frac{\kappa\nu\zeta}{1-(1-\tau^{US})\kappa(\sigma+\nu\zeta)}}. \quad (63)$$

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 baseline 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 of preferences (σ, ζ) and the parameters of the wage function (γ, ν, κ) we can infer the parameters of the skill distribution $(\sigma_\epsilon^2, \lambda_\epsilon)$ which matches the

estimated parameters of the earnings distribution $(\sigma_\omega^2, \lambda_\omega)$. We use the estimates of Cai and Heathcote (2022): $\sigma_\omega^2 = 0.548$ and $\lambda_\omega = 1.670$.[12] We then modify the tax schedule at the top earnings distribution, assuming that top income tax rate is constant. Then 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 then infer the Pareto parameter ϕ of the skill distribution to match the Pareto parameter $a_\omega^{top} = 1.4$ of the earnings distribution.[13]

We calibrate the return to skill parameter γ using the Mincerian wage regression. In the baseline calibration, we assume that the proxy of the skill variable θ in our model is the cognitive skill, and use the regression coefficient of AFQT test score from the NSLY data. In the benchmark, we use the estimates of Castex and Kogan Dechter (2014) and set $\gamma = 0.113$ to match the estimates controlling for schooling and work experiences. See the appendix for more details. 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 calibrate the 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 (61) $E[e_t] = \Psi_t \cdot \nu_t^{A_t(\zeta)}$, 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.[14] We then infer ζ given

[12] Cai and Heathcote (2022) estimate these parameters using the 2016 Survey of Consumer Finances data to include enough high income households.

[13] This benchmark a_ω^{top} is the estimation measured by the Adjusted Gross Income (AGI) data in the literature. Since the AGI includes capital gains which are not relevant for our earning tax, Badel, Huggett, and Luo (2020) calculates $a_\omega^{top} = 1.7$ by excluding dividends and long-term capital gain. We do sensitivity analysis for lower a_ω^{top} in the appendix.

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

Table 1: Exogenous Wage Case: Pareto Efficient Tax and Decomposition

$a_\omega^{US}=1.4$		$a_\omega^F=1.26$ (−10%)				
τ_{ub}^{US}	EEC	source	τ_{ub}^F	ΔEEC	$EEC=\tilde{\epsilon}\cdot\frac{1}{\frac{y'(\theta)\theta}{y(\theta)}}$	
					$\Delta\tilde{\epsilon}$	$\Delta\frac{1}{\frac{\tilde{y}'(\theta)\theta}{\tilde{y}(\theta)}}$
0.54	1.52	$\gamma\uparrow$	0.57 (+4.79%)	-10.0%	0.0%	-10.0%
		$\nu\uparrow$	0.56 (+2.52%)	-5.38%	+4.04%	-9.05%
		$\kappa\uparrow$	0.54 (0.0%)	0.0%	0.0%	0.0%

(ν, κ, τ^{US}).

Given ζ, the return to human capital ν and return to effective labor supply κ are determined to match the college premium and top income inequality. For the college premium, we use 1.9 which is 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). The target for the top income inequality is the Pareto parameter $a_\omega^{top}=1.4$. The resulting parameters are $\zeta=0.53$, $\nu=0.03$, and $\kappa=1.75$.

Numerical Results Table 1 presents the numerical results answering the main question of this paper—how does the Pareto efficient tax rate respond to rise in inequality? Recall that with exogenous wage, the upper bound of Pareto efficient top income tax rate is $\tau_{ub}=\frac{1}{1+EEC\cdot\phi}$. Given the Pareto parameter ϕ of the skill distribution fixed, the response of τ_{ub} is driven by the response of the effective efficiency cost (EEC) only.

We consider three cases—(i) 10% decrease in a_ω^F is driven rise in γ only, (ii) driven by rise in ν only, and (iii) driven by rise in κ only. Table 1 presents the response of the EEC to each source of inequality. When return to skill γ is the source of inequality, the EEC of tax distortion decreases by 10% as rise in γ increases the differential of effective labor supply across skill types which tightens the incentive constraint calling for the tax distortion. On the

other hand, when the return to effective labor κ (superstar effect) is the source of inequality, rise in inequality has no impact on the EEC, which confirms the theoretical result in Section **??**. Lastly, when return to education ν rises, Table 1 shows the two counterbalancing forces for the EEC. Higher ν increases the labor supply elasticity $\tilde{\epsilon}$—increasing the cost—and it increases the differential of effective labor across skills—decreasing the cost. Since γ is relatively small in our baseline calibration, the differential of y across skills is quite sensitive to the change in ν, and this force is dominating. Thus rise in ν results in decrease in EEC by 5.38%. As a result, the upper bound of the Pareto efficient top income tax rate increases in response to rise in inequality if either return to skill or return to human capital is the source of inequality, but the response is much larger when the return to skill rises.

5.2 Comparative Statics with Endogenous Wage

We now extend the investigation to the model with endogenous wage to quantitatively decompose the general equilibrium effect of the EEC and its response to each source of inequality. Given that the our analysis in Section **??** shows that the superstar effects can at best have indirect general equilibrium effects through its interaction with price of education, we focus on the model where only return to education is endogenized.

Model with Endogenous Price of Education We endogenize the price of education with imperfect substitution of labor input of different levels of education as in Section **??**. Different from the model in Section **??**, however, we now accommodate the (exogenous) superstar effects, which allows us to study the interaction between the superstar effects and return to education. As in Section 5.1, the wage schedule across effective labor has the form of $W(y) = y^{\kappa}$, where κ controls the superstar effects.

Given the price of effective labor $w(e)$ for each level of education e and the wage schedule $W(y) = y^{\kappa}$, to hire labor input $\tilde{\ell}(e)$ from a worker with education level e, a competitive firm should pay $W(y(e)) = y(e)^{\kappa}$ per worker, where $y(e) = w(e) \cdot \tilde{\ell(e)}$. Then the firm which produces aggregate effective labor Y solves the following cost-minimization problem given

the density of each education level $m(e)$:

$$C(Y) = \min_{\{\tilde{l}(e)\}} \int \left[w(e) \cdot \tilde{\ell}(e)\right]^{\kappa} \cdot m(e)de \tag{64}$$

$$(\lambda_c(Y)) \quad s.t. \quad Y = \left\{ \int \left[e^{\psi} \cdot \tilde{\ell}(e)^{\frac{\eta-1}{\eta}} \right]^{\kappa} \cdot m(e)de \right\}^{\frac{\eta}{\eta-\kappa}}, \tag{65}$$

which generates the cost function $C(Y)$. We denote the solution to (64) by $\tilde{\ell}_D(e|Y,w)$. From the first-order condition of this cost-minimization problem, the demand curve of $\tilde{\ell}_D(e|Y,w)$ should satisfy:

$$w(e) = \left[\lambda_c(Y) \cdot \frac{\eta-1}{\eta-\kappa}\right]^{\frac{1}{\kappa}} \times e^{\psi} \times \left(\frac{Y}{\tilde{\ell}_D(e)}\right)^{\frac{1}{\eta}}. \tag{66}$$

We remark that with the technology function in (65), the optimal ratio of $\tilde{\ell}_D(e|Y,w)$ across e's do not depend on Y, and this simplifies our analysis.[15]

The representative firm has a technology $A(Y)$ of producing aggregate output. The firm decides aggregate effective labor Y to maximize its profit:

$$\pi = \max_{Y} A(Y) - C(Y).$$

We assume that $A''(Y) \leq C''(Y)$ so that firm's profit maximization is well defined.

Worker's problem is exactly as in the model with exogenous wage in Section **??**, and we assume the same form of preferences (60) as in the model with exogenous wage in Section 5.1. Under the linear tax system with tax rate τ, consumption of the worker is $c(\theta) = (1-\tau) \cdot W(y(\theta)) + Tr = [\theta^{\gamma} \cdot w(e(\theta)) \cdot n(\theta)]^{\kappa} + Tr$, where Tr is a lumpsum transfer. Workers with skill level θ optimally choose education $e(\theta)$ and working hour $n(\theta)$, which yields the effective labor supply for every level of education e

$$\tilde{\ell}_S(e) = \left(e^{-1}(e)\right)^{\gamma} \times n(e^{-1}(e)).$$

[15]We also note that the since the firm chooses $\tilde{\ell}(e)$ given density $m(e)$, the relative price $w(e)$'s across different levels of e's do not depend on the mass of e, which is different from the model in Section **??** where the firm chooses $\ell(e) = \tilde{\ell}(e) \cdot m(e)$. We intentionally make this assumption for two purposes. First, with this setup, we can endogenize wage while keeping the wage structure exactly as in the model with exogenous wage. Second, it allows us to analyze the price elasticity independent of the change in skill distribution, so we focus on how different returns affect the price elasticity.

Table 2: Endogenous Wage Case: Pareto Efficient Tax and Decomposition

$a_\omega^{US}=1.4$		$a_\omega^F=1.26$ (−10%)				
τ_{ub}^{US}	EEC^{GE}	source	τ_{ub}^F	ΔEEC^{GE}	$EEC^{GE}=EEC\cdot\frac{1}{1-\epsilon_{w,1-\tau}\cdot\frac{1}{\tilde{e}}}$	
					ΔEEC	$\Delta\frac{1}{1-\epsilon_{w,1-\tau}\cdot\frac{1}{\tilde{e}}}$
0.38	1.80	$\gamma\uparrow$	0.41 (+8.51%)	-9.74%	-11.49%	+1.98%
		$\nu\uparrow$	0.39 (+3.40%)	-4.01%	-5.38%	+1.44%
		$\kappa\uparrow$	0.37 (−1.48%)	+1.80%	+0.38%	+1.42%

Then the equilibrium price $w(e)$ of every level of education e is endogenously determined to clear the labor market: $\tilde{\ell}_D(e)=\tilde{\ell}_S(e)$. In the appendix, we derive the equilibrium price $w(e)$ and show that it has a form of $w(e)=\Psi(\tau)\times e^\nu$, where ν is independent of tax rate.

Numerical Results Table 2 presents the numerical results for the endogenous wage case. The last column (C_3) of Table 2 shows that regardless of sources of inequality, when inequality increases, the general equilibrium effects with increased price elasticity provide additional forces of increasing EEC, which implies additional forces of decreasing τ_{ub}. As we discussed above, the price elasticity is positive under our baseline calibration, thus incorporating the endogenous price effect itself lowers τ_{ub} relative to the exogenous price case. This general equilibrium force is intensified with higher returns to each inequality factor. With higher returns, the scarcity of effective labor supply given tax schedule is lowered, which mitigates supply effects and increases price elasticity. Thus the general equilibrium forces are strengthened providing an additional force of increasing the EEC and decreasing τ_{ub}. The effect of rise in κ on the price elasticity, however, depends on the calibration as we can notice from Lemma **??**. If the return to skill γ is very big, the scarcity of effective labor would not be that sensitive to the change in κ and the price elasticity can be decreasing in κ.

6 Conclusion

- need to fill out

References

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

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

Cai, Z., and J. Heathcote (2022): "College Tuition and Income Inequality," *American Economic Review*, 112(1), 81–121.

Castex, G., and E. Kogan Dechter (2014): "The Changing Roles of Education and Ability in Wage Determination," *Journal of Labor Economics*, 32(4), 685–710.

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.

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

Piketty, T., E. Saez, and S. Stantcheva (2014): "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, 6(1), 230–271.

Sachs, D., A. Tsyvinski, and N. Werquin (2020): "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," *Econometrica*.

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 I. Werning (2017): "The Taxation of Superstars," *The Quarterly Journal of Economics*, 132(1), 211–270.

Stantcheva, S. (2017): "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*, 125(6), 1931–1990.

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

Appendix (For Online Publication)

A Proofs of Equations in Section ??

We first derive the first order conditions of the planning problem in a simple framework with exogenous returns to inequality factors in section **??**. We present the planning problem (19) of this simple framework and Lagrangian of this planning problem is given by

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

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 (**??**) and (**??**) 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 (two-step problem provided in Section **??**). Given the value function $U(c, y, \theta)$ associated with worker's sub-problem (3) and the wage schedule $W(\cdot)$, worker 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:

$$MRS(c(\theta), y(\theta), \theta) = (1 - T'(\omega))W'(y(\theta)), \tag{67}$$

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 (**??**) can be rewritten as

$$\frac{\frac{\partial\Phi}{\partial y(\theta)} \cdot \frac{1}{f(\theta)} - C_y(\theta)}{MRS(\theta)} = \hat{\mu}(\theta) \times \frac{U_{\theta c}(\theta)C_y(\theta) + U_{\theta y}(\theta)}{U_c(\theta)MRS(\theta)}.$$

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 (**??**), we obtain

$$\frac{\frac{\partial \Phi}{\partial y(\theta)} \cdot \frac{1}{f(\theta)} - C_y(\theta)}{MRS(\theta)} = \frac{T'(\omega)}{1 - T'(\omega)}.$$

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 of the planning problem with respect to $y(\theta)$ becomes

$$\hat{\mu}(\theta) = \frac{T'(\omega(\theta))}{1 - T'(\omega(\theta))} \theta f(\theta) \left(-\frac{MRS_\theta \cdot \theta}{MRS(\theta)} \right)^{-1}. \tag{68}$$

Next, the first-order condition (**??**) of the planning problem with respect to $v(\theta)$ 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 $v(\theta)$, we obtain

$$-\hat{\mu}'(\theta) - \hat{\mu}(\theta) \cdot MRS_c \cdot y'(\theta) = f(\theta) - \frac{\lambda(\theta) U_c(\theta)}{\eta}. \tag{69}$$

B Proof of Propositions in 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 **??**.

Combining (**??**) 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 (**??**) 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 (**??**), we replace $\frac{d\log\hat{\mu}(\theta)}{d\log\theta}$ to yield

$$\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\left[\begin{array}{l}-\frac{d}{d\log\theta}\log\left(\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\right)\\ -MRC_c(\theta)\cdot y'(\theta)\cdot\theta\end{array}\right]\leq 1, \tag{70}$$

which is the test formula in terms of skill distribution and marginal rate of substitution. We can also rewrite (70) to express the equivalent formula in an alternative way:

$$\begin{array}{lll}
 & (70) & \\
\Leftrightarrow & -\theta\cdot\frac{1}{\theta f(\theta)}\cdot\frac{d}{d\theta}\left[\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\right] & \\
 & \quad -\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\cdot MRS_c(\theta)\cdot y'(\theta)\cdot\theta \quad\leq\quad 1 \\
\Leftrightarrow & -f(\theta)-\frac{d}{d\theta}\left[\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\right] & \\
 & \quad -\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\cdot MRS_c(\theta)\cdot y'(\theta)\cdot\theta f(\theta) \quad\leq\quad 0 \\
\Leftrightarrow & \frac{d}{d\theta}\left[\begin{array}{l}1-F(\theta)-\frac{T'(\omega(\theta))}{1-T'(\omega(\theta))}\theta f(\theta)\left(-\frac{MRS_\theta\cdot\theta}{MRS(\theta)}\right)^{-1}\\ \quad+\int_\theta^{\overline{\theta}}\frac{T'(\omega(\tilde{\theta}))}{1-T'(\omega(\tilde{\theta}))}\left(-\frac{MRS_\theta\cdot\tilde{\theta}}{MRS(\tilde{\theta})}\right)^{-1}\cdot MRS_c(\tilde{\theta})\cdot y'(\tilde{\theta})\cdot\tilde{\theta}f(\tilde{\theta})d\tilde{\theta}\end{array}\right] \quad\leq\quad 0.
\end{array}$$

This alternative condition can be stated as

$$\frac{d}{d\theta}R(\theta)\leq 0,\quad\forall\theta, \tag{71}$$

where

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

B.1 Test Formulas in terms of Earnings Distribution

We now derive the test formulas in terms of earnings distribution (**??**) in Proposition **??** and alternative representation (**??**).

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

$$\epsilon^u_\omega \equiv \left.\frac{\partial \omega}{\partial(1-\tau)}\right|_{\tau=I=0} \frac{1-T'}{\omega} = \frac{\frac{W'}{\omega} - MRS_c}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'} \tag{72}$$

$$\eta_\omega \equiv -(1-T')\left.\frac{\partial \omega}{\partial I}\right|_{\tau=I=0} = \frac{MRS_c}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'} \tag{73}$$

$$\epsilon^c_\omega \equiv \epsilon^u_\omega + \eta_\omega = \frac{\frac{W'}{\omega}}{MRS_c + \frac{MRS_y}{MRS} - \frac{W''}{W'} + \frac{T''}{1-T'}W'}. \tag{74}$$

For later use, the ratio between η_ω and ϵ^c_ω is

$$\frac{\eta_\omega}{\epsilon^c_\omega} = MRS_c \frac{\omega}{W'(y)} = MRS_c \cdot y \cdot \frac{1}{\frac{W'(y)y}{W(y)}}. \tag{75}$$

Proof of Proposition ?? We now derive the Pareto efficiency test formulas in terms of earnings distribution. To relate the test formulas (70) and (71) 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

$$f(\theta) = h(\omega(\theta))\omega'(\theta). \tag{76}$$

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

$$\left(-\frac{MRS_\theta \theta}{MRS}\right)^{-1} \frac{\omega'(\theta)\theta}{\omega(\theta)} = \epsilon^c_\omega(\omega) \tag{77}$$

Also, note that from (75)

$$MRS_c(\theta) y'(\theta)\theta = \frac{\eta_\omega(\omega)}{\epsilon^c_\omega(\omega)} \cdot \frac{y'(\theta)\theta}{y(\theta)} \cdot \frac{W'(y(\theta))y(\theta)}{W(y(\theta))} = \frac{\eta_\omega(\omega)}{\epsilon^c_\omega(\omega)} \cdot \frac{\omega'(\theta)\theta}{\omega(\theta)}. \tag{78}$$

Using (76), (77), and (78), we can rewrite (70) 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 (71). 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 (**??**) in Proposition **??**: $\frac{dR^\omega(\omega)}{d\omega} \leq 0$, 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

B.2.1 Proof of Proposition ??

We now derive the test formulas in terms of skill distribution in a simple baseline model with exogenous returns to inequality factors. Note that we have already expressed the formulas in terms

of skill distribution and marginal rate of substitutions in equation (71) above. The test formula (**??**) in Proposition **??** is obtained simply by defining the effective efficiency cost (EEC) by

$$EEC(\theta) = \left(-\frac{MRS_\theta(\theta)\cdot\theta}{MRS(\theta)}\right)^{-1}. \tag{79}$$

We have shown the relationship between the $\frac{MRS_\theta\theta}{MRS}$ and the earnings elasticity in equation (77). Then defining the EEC as in equation (79) yields the equation (**??**) in Proposition **??**.

B.2.2 Proof of Proposition ??

We now characterize the EEC in terms of ealsticities. We begin with the derivation of the elasticities of effective labor supply.

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

$$\tilde{\epsilon}^u \equiv \left.\frac{\partial y}{\partial(1-\tau)}\right|_{\tau=I=0}\frac{1-T'}{y} = \frac{\frac{1}{y} - MRS_c}{MRS_c + \frac{MRS_y}{MRS}} \tag{80}$$

$$\tilde{\eta} \equiv -(1-T')\left.\frac{\partial\omega}{\partial I}\right|_{\tau=I=0} = \frac{MRS_c}{MRS_c + \frac{MRS_y}{MRS}} \tag{81}$$

$$\tilde{\epsilon}^c \equiv \tilde{\epsilon}^u + \tilde{\eta} = \frac{\frac{1}{y}}{MRS_c + \frac{MRS_y}{MRS}}. \tag{82}$$

We also note that

$$\frac{\tilde{\epsilon}^c}{1+\tilde{\epsilon}^u} = \frac{\frac{1}{y}}{\frac{1}{y}+\frac{MRS_y}{MRS}} = \left(1+\frac{MRS_y \cdot y}{MRS}\right)^{-1} \tag{83}$$

$$\frac{\tilde{\eta}}{\tilde{\epsilon}^c} = MRS_c \cdot y \tag{84}$$

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). Thus, the shape of the utility over (e, h) matters for the elasticity, although the expressions of the elasticities (83) and (83) are the same as in the standard Mirrlees model.

Characterization of EEC in terms of Elasticities We now express the EEC in terms of elasticities. Given that the elasticities are expressed in terms of $\frac{MRS_y \cdot y}{MRS}$ but that the EEC is defined as the inverse of $-\frac{MRS_\theta \cdot \theta}{MRS}$, we identify the relationship between MRS_θ and MRS_y. This relationship (and thus the EEC) does depend on the wage structure $p(\theta, e)$ and primitive utility $u(c, \tilde{e}, n)$.

First, we define the notation of derivatives. We denote the second-order derivatives of the value function U holding the productivity function $p(\cdot, \cdot)$ fixed (equivalently, holding the human capital price function $w(\cdot)$ fixed) by $U_{yy}, U_{cy}, U_{y\theta}$, and $U_{c\theta}$. Similarly, we use the notation of MRS_y, MRS_c, and MRS_θ for the derivatives holding the function $p(\cdot, \cdot)$ fixed. In this model with exogenous returns, there is no confusion regarding these notations, but in the model with endogenous returns, we need to make the notation explicit.

By definition, $MRS = -\frac{U_y}{U_c}$. We then obtain

$$MRS_y = -\frac{U_{yy} \cdot U_c - U_y \cdot U_{cy}}{U_c^2},$$

and thus

$$\frac{MRS_y \cdot y}{MRS} = \frac{U_c \cdot y}{U_y} \times \frac{U_{yy} \cdot U_c - U_y \cdot U_{cy}}{U_c^2} = \left(\frac{U_{yy}}{U_y} - \frac{U_{cy}}{U_c}\right) \cdot y. \tag{85}$$

Similarly, we obtain

$$MRS_\theta = -\frac{U_{y\theta} \cdot U_c - U_y \cdot U_{c\theta}}{U_c^2},$$

and thus

$$\frac{MRS_\theta \cdot \theta}{MRS} = \frac{U_c \cdot \theta}{U_y} \times \frac{U_{y\theta} \cdot U_c - U_y \cdot U_{c\theta}}{U_c^2} = \left(\frac{U_{y\theta}}{U_y} - \frac{U_{c\theta}}{U_c}\right) \cdot \theta. \tag{86}$$

Note that worker's sub-problem in the second step can be written as

$$U(c, y, \theta) = \max_e u\left(c, \tilde{e}(e, \theta), \frac{y}{p(\theta, e)}\right),$$

and we obtain the first-order condition $u_{\tilde{e}} \cdot \tilde{e}_e = u_n \cdot \frac{y \cdot p_e}{p^2}$. Using the envelope theorem, we obtain $U_y = u_n \cdot \frac{1}{p(\theta,e)}$ and $U_\theta = -u_{\tilde{e}} \cdot \frac{e}{\theta^2} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} - u_n \cdot \frac{y}{p^2} \cdot p_\theta$. Combining the envelope conditions and the first-order condition yields

$$\begin{aligned} U_\theta &= -\frac{u_n}{p} \cdot \frac{y}{\theta} \times \left(\frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_\theta \cdot \theta}{p} \right) \\ &= -U_y \cdot \frac{y}{\theta} \times \left(\frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_\theta \cdot \theta}{p} \right). \end{aligned} \tag{87}$$

Using equation (87), we obtain the derivatives of U_θ with respect to c and y:

$$\begin{aligned} U_{\theta y} &= -\left(U_{yy} \cdot \frac{y}{\theta} + U_y \cdot \frac{1}{\theta} \right) \times \left(\frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_\theta \cdot \theta}{p} \right) \\ &\quad -U_y \cdot \frac{y}{\theta} \times \left[\left(\frac{p_{ee} \cdot p - p_e^2}{p^2} e + \frac{p_e}{p} \right) \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_{\theta e} \cdot p - p_\theta \cdot p_e}{p^2} \theta \right] \times \frac{\partial e}{\partial y}, \\ U_{\theta c} &= -U_{cy} \cdot \frac{y}{\theta} \times \left(\frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_\theta \cdot \theta}{p} \right) \\ &\quad -U_y \cdot \frac{y}{\theta} \times \left[\left(\frac{p_{ee} \cdot p - p_e^2}{p^2} e + \frac{p_e}{p} \right) \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_{\theta e} \cdot p - p_\theta \cdot p_e}{p^2} \theta \right] \times \frac{\partial e}{\partial c}. \end{aligned}$$

Replacing $U_{\theta y}$ and $U_{\theta c}$ in equation 86 yields

$$\begin{aligned} \frac{MRS_\theta \cdot \theta}{MRS} &= -\left(1 + \left(\frac{U_{yy}}{U_y} - \frac{U_{cy}}{U_c} \right) \cdot y \right) \times \left(\frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_\theta \cdot \theta}{p} \right) \\ &\quad + y \times \left[\left(\frac{p_{ee} \cdot p - p_e^2}{p^2} e + \frac{p_e}{p} \right) \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_{\theta e} \cdot p - p_\theta \cdot p_e}{p^2} \theta \right] \times \left(-\frac{\partial e}{\partial y} + \frac{U_y}{U_c} \cdot \frac{\partial e}{\partial c} \right) \end{aligned} \tag{88}$$

Replacing equation (85) into the right-hand side of (88) yields

$$-\frac{MRS_\theta \cdot \theta}{MRS} = \left(1 + \frac{MRS_y \cdot y}{MRS} \right) \times \left(\frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_\theta \cdot \theta}{p} \right) + \Omega \tag{89}$$

where

$$\Omega = U_y \cdot y \times \left[\frac{p_e}{p} \cdot \left(1 + \frac{p_{ee}}{p_e} e - \frac{p_e}{p} e \right) \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_{\theta e} \cdot p - p_\theta \cdot p_e}{p^2} \theta \right] \times \left(\frac{1}{U_y} \frac{\partial e}{\partial y} - \frac{1}{U_c} \cdot \frac{\partial e}{\partial c} \right). \tag{90}$$

Replacing (83) into (89) yields

$$\begin{aligned} EEC &= \left(-\frac{MRS_\theta \cdot \theta}{MRS} \right)^{-1} \\ &= \left[\left(\frac{\tilde{\epsilon}^c(\theta)}{1 + \tilde{\epsilon}^u(\theta)} \right)^{-1} \times \left(\frac{p_e \cdot e}{p} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} + \frac{p_\theta \cdot \theta}{p} \right) + \Omega \right]^{-1}, \end{aligned} \tag{91}$$

which provides the equation (**??**) in Proposition **??**

Finally, Proposition **??** also states that $\Omega = 0$ when $\rho_{\theta e} = 1$. We show this in Section **??**.

B.3 Analysis of an Extended Model in Section ??

In Section **??** of the main text, we provided an extended model which endogenizes returns to human capital only. This simple extension allows us to clearly analyze the general equilibrium effects through the human capital prices. We first derive the Pareto efficient tax formula given the incidence on prices of a tax perturbation, then using a special example, we will provide a sharp characterization of the relevant EEC of tax distortion accommodating the price elasticities.

We assume that a representative firm has a CES technology in the form of (24), which allows imperfect substitution across labor inputs $\ell(e)$'s of different human capital level e's, where the elasticity of substitution is constant η. Under this CES technology, the demand curve of effective labor $\ell(e)$ of each human capital level is provided by (**??**), which shows that the ratio of prices across different levels of human capital is independent of aggregate labor Y, although the level of price depends on Y. As the incentive constraints essentially depends on the relative prices, it is convenient to normalize the return to human capital and efficiency unit of labor supply by $Y^{\frac{1}{\eta}}$ and denote them by $\tilde{w}(e) = \frac{w(e)}{Y^{\frac{1}{\eta}}}$ and $\tilde{y}(\theta) = \frac{y(\theta)}{Y^{\frac{1}{\eta}}}$ respectively. We also denote the normalized productivity function by $\tilde{p}(\theta) = \theta^{\gamma} \cdot \frac{w(e(\theta))}{Y^{\frac{1}{\eta}}}$

Both workers and firms solve their problems taking prices as given. Thus worker's problem and its solution are the same as in the economy with exogenous returns if the prices are the same in two economies. As a result, the elasticities of earnings and the elasticities of effective labor supply which we defined in Section **??** of the main text do not change (except that we now assume $W(y) = y$ and $\omega(\theta) = y(\theta)$). We remark that these are micro-elasticities holding the prices constant (not macro-elasticities which takes into account change in prices). The planner, however, considers how the prices respond when she changes the tax rates, and thus in the planning problem, we denote the value function of workers by $U(c, y, \theta; \tilde{p})$ to show its dependence on prices $\tilde{p}$ explicitly. As a result, the formulas will be expressed in terms of modified elasticity which accommodates the price elasticity.

We rewrite the incentive constraints in terms of normalized variables. The envelope conditions of the worker's sub-problem (**??**) yield $U_{\tilde{y}} = u_n \cdot \frac{1}{\tilde{p}}$ and $U_\theta = -u_{\tilde{e}} \cdot \frac{e}{\theta^2} \cdot \mathbb{1}_{\left\{\tilde{e}=\frac{e}{\theta}\right\}} - u_n \cdot \frac{\tilde{y}}{\tilde{p}^2} \cdot \tilde{p}_\theta$. Combining these with the first-order condition of (**??**) $u_{\tilde{e}} \cdot \tilde{e}_e = u_n \cdot \frac{\tilde{y} \cdot \tilde{p}_e}{\tilde{p}^2}$, we obtain $U_\theta = -U_{\tilde{y}} \cdot \frac{\tilde{y}}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right)$. Thus the incentive constraints of the planning problem (**??**) can be rewritten as

$$v'(\theta) = -U_{\tilde{y}(\theta)} \frac{\tilde{y}(\theta)}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right).$$

We can write the Lagrangian of the part of the planning problem as follows:

$$\begin{aligned} L &= \int \Lambda(\theta) v(\theta) dF(\theta) + \lambda \left[\int \tilde{y}(\theta) dF(\theta)\right]^{\frac{\eta}{\eta-1}} - \int C(v(\theta), \tilde{y}(\theta), \theta) dF(\theta) \\ &\quad - \int \mu'(\theta) v(\theta) d\theta + \int \mu(\theta) U_{\tilde{y}}(\theta) \cdot \frac{\tilde{y}(\theta)}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) d\theta, \end{aligned}$$

where we used integral by parts and $\mu(\bar{\theta}) = \mu(\underline{\theta}) = 0$. Then the planner solves this Lagrangian subject to the constraints which determine the equilibrium price schedule $w(e)$.

Instead of deriving the first-order conditions of the full planning problem—which includes all the constraints related to the general equilibrium, we derive the first-order condition *given the incidences of perturbation on prices*. These incidences on prices are highly endogenous objects, which depends on the structure of the labor market, and we do not attempt to characterize these price elasticities in a general set up. After providing the formula given the price elasticity, we characterize the role of these price elasticity with a simple example.

We denote the second-order derivatives of the value function U holding the productivity function $\tilde{p}(\cdot,\cdot)$ fixed by $U_{\tilde{y}\tilde{y}}$ and $U_{c\tilde{y}}$. We also use the notation of MRS_c and $MRS_{\tilde{y}}$ for the derivatives of MRS holding the function $\tilde{p}(\cdot,\cdot)$ fixed. On the other hand, we denote the elasticity of $\tilde{p}(e,\theta)$ with respect to $\tilde{y}$ holding the human capital level e fixed by $\epsilon_{\tilde{p},\tilde{y}}(\theta)$. Then the first order condition of this Lagrangian with respect to $\tilde{y}(\theta)$ yields

$$\begin{aligned}
&\lambda \cdot \left(\frac{\eta}{\eta-1} \cdot Y^{\frac{1}{\eta}} - C_{\tilde{y}}(\theta)\right) \cdot f(\theta) \\
&= -\mu(\theta) \times \left[\begin{array}{l} \left\{U_{\tilde{y}c}(\theta) \cdot C_{\tilde{y}}(\theta) + U_{\tilde{y}\tilde{y}}(\theta) + \frac{\partial U_{\tilde{y}}(\theta)}{\partial \tilde{p}(\theta)} \epsilon_{\tilde{p}\tilde{y}}(\theta) \frac{\tilde{p}(\theta)}{\tilde{y}(\theta)}\right\} \cdot \frac{\tilde{y}(\theta)}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) \\ + U_{\tilde{y}}(\theta) \cdot \frac{1}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) + U_{\tilde{y}}(\theta) \cdot \frac{\tilde{y}(\theta)}{\theta} \cdot \frac{\partial \nu(\theta)}{\partial \tilde{y}(\theta)} \end{array}\right], \quad (92)
\end{aligned}$$

where $\nu(\theta) = \frac{\tilde{p}_e e}{\tilde{p}}$, and $\frac{\partial \nu(\theta)}{\partial \tilde{y}(\theta)}$ denotes the total incidence of the return to e considering the change in $\nu(\theta)$ both through the change in human capital level and change in price function $w(e)$. Since we have already normalized the prices by denominating with $Y^{\frac{1}{\eta}}$, the first order condition does not include the cross derivatives of the prices.

Define $\hat{\mu}(\theta) = \frac{\mu(\theta) U_c}{\lambda}$ and $MRS(c,y,\theta) = -\frac{U_{\tilde{y}}(\theta)}{U_c(\theta)} \times \frac{1}{Y^{\frac{1}{\eta}}}$. From the first-order condition of the worker's sub-problem (**??**), we have $MRS(c,\tilde{y},\theta) = (1 - T'(y))$. Then $C_{\tilde{y}}(\theta) = -\frac{U_{\tilde{y}}(\theta)}{U_c(\theta)} = MRS(\theta) \times Y^{\frac{1}{\eta}} = (1 - T'(y(\theta))) \cdot Y^{\frac{1}{\eta}}$ and thus $\frac{Y^{\frac{1}{\eta}} - C_{\tilde{y}}(\theta)}{MRS(\theta)} = \frac{T'(y(\theta))}{1 - T'(y(\theta))} \cdot Y^{\frac{1}{\eta}}$. Replacing these terms in (92) yields

$$\left(\frac{T'(y(\theta))}{1 - T'(y(\theta))} + \frac{\eta}{\eta - 1} \cdot \frac{1}{T'(y(\theta))}\right) \times \theta f(\theta) = \hat{\mu}(\theta) \times \left[\begin{array}{l} \left\{\left(-\frac{U_{\tilde{y}c}(\theta)}{U_c(\theta)} + \frac{U_{\tilde{y}\tilde{y}}(\theta)}{U_{\tilde{y}}(\theta)}\right) \cdot \tilde{y}(\theta) + 1\right\} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) \\ + \frac{\partial \nu(\theta)}{\partial \tilde{y}(\theta)} \cdot \tilde{y}(\theta) \\ + \frac{\tilde{p}(\theta)}{U_{\tilde{y}}(\theta)} \cdot \frac{\partial U_{\tilde{y}}(\theta)}{\partial \tilde{p}(\theta)} \cdot \epsilon_{\tilde{p}\tilde{y}}(\theta) \times \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) \end{array}\right]. \quad (93)$$

To express the right-hand side of (93) in terms of elasticity, note that $\frac{MRS_{\tilde{y}} \cdot \tilde{y}}{MRS} = \frac{U_c \cdot \tilde{y}}{U_{\tilde{y}}} \times \frac{U_{\tilde{y}\tilde{y}} \cdot U_c - U_{\tilde{y}} \cdot U_{c\tilde{y}}}{U_c^2} = \left(\frac{U_{\tilde{y}\tilde{y}}}{U_{\tilde{y}}} - \frac{U_{c\tilde{y}}}{U_c}\right) \cdot \tilde{y}$. By using $\frac{MRS_{\tilde{y}} \cdot \tilde{y}}{MRS} = \frac{MRS_y \cdot y}{MRS}$, (83) and (85), we have

$$1 + \left(\frac{U_{\tilde{y}\tilde{y}}(\theta)}{U_{\tilde{y}}(\theta)} - \frac{U_{\tilde{y}c}(\theta)}{U_c(\theta)}\right) \cdot \tilde{y}(\theta) = \frac{1 + \tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)}. \quad (94)$$

We also note that $U_{\tilde{y}} = \frac{1}{\tilde{p}} u_n$ and thus $\frac{\partial U_{\tilde{y}}}{\partial \tilde{p}} = -\frac{\tilde{y}}{\tilde{p}} \cdot \left(\frac{U_{\tilde{y}}}{\tilde{y}} + U_{\tilde{y}\tilde{y}} \right)$. Thus we can express $\frac{\tilde{p}(\theta)}{U_{\tilde{y}}(\theta)} \cdot \frac{\partial U_{\tilde{y}}(\theta)}{\partial \tilde{p}(\theta)}$ by

$$\begin{aligned} \frac{\tilde{p}}{U_{\tilde{y}}} \cdot \frac{\partial U_{\tilde{y}}}{\partial \tilde{p}} &= -\frac{\tilde{p}}{U_{\tilde{y}}} \cdot \frac{\tilde{y}}{\tilde{p}} \cdot \left(\frac{U_{\tilde{y}}}{\tilde{y}} + U_{\tilde{y}\tilde{y}} \right) \\ &= -\left(1 + \frac{U_{\tilde{y}\tilde{y}} \cdot \tilde{y}}{U_{\tilde{y}}} - \frac{U_{c\tilde{y}} \cdot \tilde{y}}{U_c} \right) - \frac{U_{c\tilde{y}}}{U_c} \cdot \tilde{y}. \end{aligned} \tag{95}$$

For simplicity of the formula, we assume that workers' preferences are additively separable.

Assumption 11. *Workers' preferences $u(c, \tilde{e}, n)$ satisfies $u_{c\tilde{e}} = u_{cn} = 0$.*

Under this assumption, the second-order derivative of the value function satisfies $U_{c\tilde{y}} = 0$. By combining (94) and (95), we can rewrite (93):

$$\left(\frac{T'(y(\theta))}{1 - T'(y(\theta))} + \frac{\eta}{\eta - 1} \cdot \frac{1}{T'(y(\theta))} \right) \times \theta f(\theta) = \hat{\mu}(\theta) \times \begin{bmatrix} \frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) + \frac{\partial \nu(\theta)}{\partial \tilde{y}(\theta)} \cdot \tilde{y}(\theta) \\ -\frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)} \epsilon_{\tilde{p}\tilde{y}}(\theta) \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) \end{bmatrix},$$

which yields

$$\hat{\mu}(\theta) = \left(\frac{T'(y(\theta))}{1 - T'(y(\theta))} + \frac{\eta}{\eta - 1} \cdot \frac{1}{T'(y(\theta))} \right) \times \theta f(\theta) \times EEC^{GE}(\theta), \tag{96}$$

where

$$EEC^{GE}(\theta) = \left[\frac{1 + \tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)} \cdot (1 - \epsilon_{\tilde{p}\tilde{y}}(\theta)) \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) + \frac{\partial \nu(\theta)}{\partial \tilde{y}(\theta)} \cdot \tilde{y}(\theta) \right]^{-1}. \tag{97}$$

On the other hand, the first-order condition of the Lagrangian of the planning problem with respect to $v(\theta)$ is the same as in the baseline model with exogenous returns, which yields

$$-f(\theta) - \hat{\mu}'(\theta) - MRS_c(\theta) \cdot y'(\theta) \cdot \hat{\mu}(\theta) \leq 0.$$

Using (96), we replace $\hat{\mu}(\theta)$ in this inequality to obtain the formula:

$$\frac{d}{d\theta} R(\theta) \leq 0, \quad \forall \theta, \tag{98}$$

where

$$\begin{aligned} R(\theta) &= 1 - F(\theta) - \left(\frac{T'(y(\theta))}{1 - T'(y(\theta))} + \frac{1}{\eta - 1} \cdot \frac{1}{1 - T'(y(\theta))} \right) \cdot \theta f(\theta) \cdot EEC^{GE}(\theta) \\ &\quad + \int_\theta^{\bar{\theta}} \left(\frac{T'(y(\tilde{\theta}))}{1 - T'(y(\tilde{\theta}))} + \frac{1}{\eta - 1} \cdot \frac{1}{1 - T'(y(\tilde{\theta}))} \right) EEC^{GE}(\tilde{\theta}) MRS_c(\tilde{\theta}) y'(\tilde{\theta}) \tilde{\theta} dF(\tilde{\theta}) \end{aligned}$$

Suppose that there is no income effect in labor supply, which implies that $MRC_c = 0$. In addition, if the marginal tax rate T' and EEC^{GE} are locally constant at θ, the test formula (98) can be simplified as follows:

$$\left(\frac{T'(y(\theta))}{1 - T'(y(\theta))} - \frac{1}{\eta - 1} \cdot \frac{1}{1 - T'(y(\theta))} \right) \cdot EEC^{GE}(\theta) \cdot \phi(\theta) \leq 1, \tag{99}$$

where $\phi(\theta) = -\frac{d \log(\theta f(\theta))}{d \log \theta}$.

B.3.1 Special Example in Section ??

We now proceed further using the simple example with linear tax provided in Section **??**. Assuming the special utility function (**??**), linear tax rate τ, and Pareto distribution of skill with $F(\theta) = 1 - (\underline{\theta}/\theta)^{\phi}$, we can analytically solve for the price of human capital.

Derivation of Analytical Expression of $w(e)$ We solve for the price function $w(e)$ using guess and verify method. We begin with a guess for the price function $w(e) = \Psi \cdot e^{\nu}$. which implies that the worker's productivity function has the form of $p(\theta, e) = \theta^{\gamma} \times \Psi \cdot e^{\nu}$. Given $p(\theta, e)$, worker with skill level θ solves the following problem:

$$\max_{e,n}(1-\tau) \cdot \Psi \cdot \theta^{\gamma} \cdot e^{\nu} \cdot n - \zeta \cdot \left(\frac{e}{\theta}\right)^{\frac{1}{\zeta}} - \sigma \cdot n^{\frac{1}{\sigma}}. \tag{100}$$

Combining the first-order conditions with respect to e and n, we can solve for the optimal decision of human capital e and working hours n:

$$\begin{aligned} e(\theta) &= A \times \theta^{\frac{1-(\sigma-\gamma\zeta)}{1-(\sigma+\nu\zeta)}}, & (101)\\ n(\theta) &= (1-\tau)^{\frac{\sigma}{1-\sigma}} \cdot \Psi^{\frac{\sigma}{1-\sigma}} \cdot A^{\frac{\nu\sigma}{1-\sigma}} \cdot \theta^{\frac{\sigma}{1-\sigma}\cdot\left(\gamma+\nu\frac{1-(\sigma-\gamma\zeta)}{1-(\sigma+\nu\zeta)}\right)}, & (102) \end{aligned}$$

where

$$A = (1-\tau)^{\frac{\zeta}{1-(\sigma+\nu\zeta)}} \cdot \Psi^{\frac{\zeta}{1-(\sigma+\nu\zeta)}} \cdot \nu^{\frac{\zeta(1-\sigma)}{1-(\sigma+\nu\zeta)}}.$$

We then calculate the effective labor supply of each human capital level $\ell_S(e) = e^{-1}(e) \cdot n(e) \cdot m(e)$ where $m(e)$ is density of human capital level e. From the analytical solution of human capital (101), we can obtain the inverse of $e(\theta)$ by

$$\theta = e^{-1}(e) = \left(\frac{e}{A}\right)^{\gamma\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}}.$$

The working hours (102) can be re rewritten using change of variables:

$$\begin{aligned} n(e) &= (1-\tau)^{\frac{\sigma}{1-\sigma}} \cdot \Psi^{\frac{\sigma}{1-\sigma}} \cdot A^{\frac{\nu\sigma}{1-\sigma}} \cdot \left(e^{-1}(e)\right)^{\frac{\sigma}{1-\sigma}\cdot\left(\gamma+\nu\frac{1-(\sigma-\gamma\zeta)}{1-(\sigma+\nu\zeta)}\right)} \\ &= (1-\tau)^{\frac{\sigma}{1-\sigma}} \cdot \Psi^{\frac{\sigma}{1-\sigma}} \cdot A^{-\frac{\sigma\gamma}{1-\sigma}\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}} \cdot e^{\frac{\sigma}{1-\sigma}\cdot\left(\nu+\gamma\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\right)}. \end{aligned}$$

From the distribution of skill, we have its density function $f(\theta) = \frac{\phi\cdot\underline{\theta}^{\phi}}{\theta^{1+\phi}}$, and $m(e) = f(e^{-1}(e)) \cdot \frac{1}{e'(e^{-1}(e))}$ yields

$$m(e) = \phi \cdot \frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\underline{\theta}^{\phi} \cdot A^{\phi\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}} \cdot e^{-1-\phi\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}}.$$

In sum, the effective labor supply yields

$$\ell_S(e) = B \times e^{-1+\frac{\nu\sigma}{1-\sigma}+\left(\frac{\gamma}{1-\sigma}-\phi\right)\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}}, \tag{103}$$

where

$$B = (1-\tau)^{\frac{\sigma}{1-\sigma}} \cdot \Psi^{\frac{\sigma}{1-\sigma}} \cdot A^{\left(\phi-\frac{\gamma}{1-\sigma}\right)\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}} \cdot \phi \cdot \frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)} \cdot \underline{\theta}^{\phi}.$$

On the other hand, the demand of effective labor $\ell_D(e)$ should satisfy (**??**). Then using the market clearing condition $\ell_D(e) = \ell_S(e)$, we obtain

$$\begin{aligned}
w(e) &= Y^{\frac{1}{\eta}} \cdot e^{\psi} \cdot \ell_S(e)^{-\frac{1}{\eta}} \\
&= Y^{\frac{1}{\eta}} \cdot B^{-\frac{1}{\eta}} \cdot e^{\psi-\frac{1}{\eta}\left(-1+\frac{\nu\sigma}{1-\sigma}+\left(\frac{\gamma}{1-\sigma}-\phi\right)\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\right)} \\
&= Y^{\frac{1}{\eta}} \cdot [(1-\tau)\cdot\Psi]^{-\frac{1}{\eta}\cdot\frac{\sigma}{1-\sigma}} \cdot A^{-\frac{1}{\eta}\cdot\left(\phi-\frac{\gamma}{1-\sigma}\right)\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}} \\
&\qquad \times \left[\phi\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\cdot\underline{\theta}^{\phi}\right]^{-\frac{1}{\eta}} \cdot e^{\psi-\frac{1}{\eta}\left(-1+\frac{\nu\sigma}{1-\sigma}+\left(\frac{\gamma}{1-\sigma}-\phi\right)\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\right)} \\
&= Y^{\frac{1}{\eta}} \cdot [(1-\tau)\cdot\Psi]^{-\frac{1}{\eta}\cdot\frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}} \cdot \nu^{-\frac{1}{\eta}\cdot\frac{\zeta\cdot(\phi(1-\sigma)-\gamma)}{1-(\sigma-\gamma\zeta)}} \\
&\qquad \times \left[\phi\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\cdot\underline{\theta}^{\phi}\right]^{-\frac{1}{\eta}} \cdot e^{\psi-\frac{1}{\eta}\left(-1+\frac{\nu\sigma}{1-\sigma}+\left(\frac{\gamma}{1-\sigma}-\phi\right)\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\right)}, \qquad (104)
\end{aligned}$$

whose right-hand side still depends on the endogenous aggregate output Y.

Note that market clearing implies that $Y = \left\{\int e^{\psi}\ell_S(e)de\right\}^{\frac{\eta}{1-\eta}}$. From (**??**) and $\ell_D(e) = \ell_S(e)$, we have $e^{\psi}\cdot\ell_S(e)^{-\frac{1}{\eta}} = w(e)\cdot Y^{-\frac{1}{\eta}}$, and thus

$$\begin{aligned}
Y &= Y^{-\frac{1}{\eta-1}} \times \{w(e)\ell(e)de\}^{\frac{\eta}{\eta-1}} \\
&= Y^{-\frac{1}{\eta-1}} \times \left\{w(e)\cdot\left(e^{-1}(e)\right)^{\gamma}\cdot n(e)\cdot m(e)de\right\}^{\frac{\eta}{\eta-1}} \\
&= Y^{-\frac{1}{\eta-1}} \times \{\theta^{\gamma}\cdot w(e(\theta))\cdot n(\theta)\cdot f(\theta)d\theta\}^{\frac{\eta}{\eta-1}},
\end{aligned}$$

which implies

$$Y = \int y_s(\theta)f(\theta)d\theta,$$

where $y_s(\theta) = \theta^{\gamma}\cdot w(e(\theta))\cdot n(\theta)$. Combining (101), (102), and $w(e(\theta)) = \Psi\cdot e(\theta)^{\nu}$ yields

$$y_s(\theta) = (1-\tau)^{\frac{\sigma+\nu\zeta}{1-(\sigma+\nu\zeta)}} \cdot \Psi^{\frac{1}{1-(\sigma+\nu\zeta)}} \cdot \nu^{\frac{\nu\zeta}{1-(\sigma+\nu\zeta)}} \cdot \theta^{\frac{\gamma+\nu}{1-(\sigma+\nu\zeta)}}.$$

Thus, the aggregate output is

$$Y = (1-\tau)^{\frac{\sigma+\nu\zeta}{1-(\sigma+\nu\zeta)}} \cdot \Psi^{\frac{1}{1-(\sigma+\nu\zeta)}} \cdot \nu^{\frac{\nu\zeta}{1-(\sigma+\nu\zeta)}} \times \int \theta^{\frac{\gamma+\nu}{1-(\sigma+\nu\zeta)}} f(\theta)d\theta. \qquad (105)$$

By replacing (105) into (104),

$$\begin{aligned}
w(e) &= (1-\tau)^{\frac{1}{\eta}\cdot\left(\frac{\sigma+\nu\zeta}{1-(\sigma+\nu\zeta)}-\frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)} \cdot \Psi^{\frac{1}{\eta}\cdot\left(\frac{1}{1-(\sigma+\nu\zeta)}-\frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)} \cdot \nu^{\frac{1}{\eta}\cdot\left(\frac{\nu\zeta}{1-(\sigma+\nu\zeta)}-\frac{\zeta(\phi(1-\sigma)-\gamma)}{1-(\sigma\gamma\zeta)}\right)} \\
&\quad \times \left[\int \theta^{\frac{\gamma+\nu}{1-(\sigma+\nu\zeta)}} f(\theta)d\theta\right]^{\frac{1}{\eta}} \cdot \left[\phi\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\cdot\underline{\theta}^{\phi}\right]^{-\frac{1}{\eta}} \times e^{\psi-\frac{1}{\eta}\left(-1+\frac{\nu\sigma}{1-\sigma}+\left(\frac{\gamma}{1-\sigma}-\phi\right)\cdot\frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)}\right)} \qquad (106)
\end{aligned}$$

Finally, we verify the guess that the price function has the form of $w(e) = \Psi \cdot e^{\nu}$ with

$$\Psi = (1-\tau)^{\frac{\frac{1}{\eta}\left(\frac{\sigma+\nu\zeta}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}{1-\frac{1}{\eta}\left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}} \times D, \text{ and} \tag{107}$$

$$\nu = \frac{\psi - \frac{1}{\eta} \cdot \frac{\gamma(1-\zeta)-(1+\phi)(1-\sigma)}{1-(\sigma-\gamma\zeta)}}{1-\frac{1}{\eta}\left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}, \tag{108}$$

where

$$D = \nu^{\frac{\frac{1}{\eta}\cdot\left(\frac{\nu\zeta}{1-(\sigma+\nu\zeta)} - \frac{\zeta(\phi(1-\sigma)-\gamma)}{1-(\sigma\gamma\zeta)}\right)}{1-\frac{1}{\eta}\left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}} \times \left[\int \theta^{\frac{\gamma+\nu}{1-(\sigma+\nu\zeta)}} f(\theta) d\theta\right]^{\frac{\frac{1}{\eta}}{1-\frac{1}{\eta}\left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}}$$
$$\times \left[\phi \cdot \frac{1-(\sigma+\nu\zeta)}{1-(\sigma-\gamma\zeta)} \cdot \underline{\theta}^{\phi}\right]^{\frac{-\frac{1}{\eta}}{1-\frac{1}{\eta}\left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}}.$$

Proof of Proposition ?? We denote the elasticity of human capital price $w(e)$ with respect to tax rate $(1-\tau)$ holding the education level e fixed by $\epsilon_{w,1-\tau} \equiv \frac{\partial \log w}{\partial \log(1-\tau)}\Big|_e$, and in this simple example, the elasticity is

$$\epsilon_{w,1-\tau} = \frac{d\log\Psi}{d\log(1-\tau)} = \frac{\frac{1}{\eta}\left(\frac{\sigma+\nu\zeta}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}{1-\frac{1}{\eta}\left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)}.$$

We also remark that in this example with no income effects in labor supply, the elasticity of effective labor supply y (or equivalently ω) with respect to retention rate $1-\tau$ is

$$\tilde{\epsilon}^u(\theta) = \tilde{\epsilon}^c(\theta) \equiv \tilde{\epsilon} = \frac{\sigma+\nu\zeta}{1-(\sigma+\nu\zeta)}, \quad \forall\theta.$$

From (108), note that ν does not depend on tax rate τ and only depends on the fixed parameters. Thus, in this simple example, $\frac{\partial\nu}{\partial\tilde{y}}\tilde{y}$ term in the EEC^{GE} equation (97) disappears. On the other hand, $\tilde{\epsilon}_{\tilde{p}\tilde{y}}$ term—derivative of price function $\tilde{p}$ with respect to $\tilde{y}$ holding e constant—can be rewritten as

$$\tilde{\epsilon}_{\tilde{p}\tilde{y}} = \frac{\partial\log\tilde{p}}{\partial\log\tilde{y}}\Big|_e = \frac{\partial\log p}{\partial\log y}\Big|_e = \frac{\partial\log w}{\partial\log(1-\tau)}\Big|_e \times \frac{\partial\log(1-\tau)}{\partial\log y} = \epsilon_w^d \times \frac{1}{\tilde{\epsilon}}.$$

It is then straightforward to show that $EEC^{GE} = \frac{\epsilon}{1+\tilde{\epsilon}} \times \frac{1}{\gamma+\nu} \times \frac{1}{(1-\epsilon_{w,1-\tau}\cdot\frac{1}{\tilde{\epsilon}})}$.

We have shown that $EEC^{GE}(\theta)$ is constant in this example. By assumption, the Pareto parameter is also constant with $-\frac{d\log(\theta f(\theta))}{d\log\theta} = \phi$. Then the Pareto efficiency test formula (99) for the tax rate τ is given by

$$\left(\frac{\tau}{1-\tau} - \frac{1}{\eta}\cdot\frac{1}{1-\tau}\right)\cdot EEC^{GE}\cdot\phi \leq 1.$$

This can be rewritten as

$$\tau \leq \frac{1-\frac{1}{\eta-1}\cdot\phi\cdot EEC^{GE}}{1+\phi\cdot EEC^{GE}},$$

which finishes the proof of Proposition **??**.

In this special example, the effective efficiency cost considering the general equilibrium effects is

$$EEC^{GE} = EEC \times \frac{1}{(1 - \epsilon_{w,1-\tau} \cdot \frac{1}{\tilde{\epsilon}})},$$

where $EEC = \frac{\tilde{\epsilon}}{1+\tilde{\epsilon}} \cdot \frac{1}{\gamma+\nu}$ is the effective efficiency cost given exogenous return. As we discuss in the main text, incorporating the endogenous prices makes the EEC of tax distortio either smaller or larger depending on the existing scarcity of effective labor.

Proof of Proposition ?? What we are interest in this paper is how this EEC^{GE} is affected by different sources of inequality. The response of EEC^{GE} to increase in each inequality factor can be decomposed into the direct force and the general equilibrium force:

$$\frac{dEEC^{GE}}{d\nu} = \underbrace{\frac{\partial EEC}{\partial \nu} \cdot \frac{1}{1 - \epsilon_{w,1-\tau} \cdot \frac{1}{\tilde{\epsilon}}}}_{\text{direct effect}} \underbrace{-EEC \cdot \frac{1}{\left(1 - \epsilon_{w,1-\tau} \cdot \frac{1}{\tilde{\epsilon}}\right)^2} \cdot \left\{ -\frac{\partial \tilde{\epsilon}_{w,1-\tau}}{\partial \nu} \cdot \frac{1}{\tilde{\epsilon}} + \tilde{\epsilon}_{w,1-\tau} \cdot \frac{1}{\tilde{\epsilon}^2} \cdot \frac{\partial \tilde{\epsilon}}{\partial \nu} \right\}}_{\text{GE effect}}, \text{ and}$$

$$\frac{dEEC^{GE}}{d\gamma} = \underbrace{\frac{\partial EEC}{\partial \gamma} \cdot \frac{1}{1 - \epsilon_{w.1-\tau} \cdot \frac{1}{\tilde{\epsilon}}}}_{\text{direct effect}} \underbrace{-EEC \cdot \frac{1}{\left(1 - \epsilon_{w,1-\tau} \cdot \frac{1}{\tilde{\epsilon}}\right)^2} \cdot \left\{ -\frac{\partial \tilde{\epsilon}_{w1-\tau}}{\partial \gamma} \cdot \frac{1}{\tilde{\epsilon}} \right\}}_{\text{GE effect}}.$$

In each equation of $\frac{dEEC^{GE}}{d\nu}$ and $\frac{dEEC^{GE}}{d\gamma}$, the first term reflects the direct effect of increase in returns (ν and γ respectively) holding the human capital price schedule $w(\cdot)$ fixed, and the second term reflects the general equilibrium effects due to change in price schedule $w(\cdot)$.

As we already discussed in Section **??**, the dependence of EEC on the sources of inequality still applies—$EEC(= \frac{\tilde{\epsilon}}{1+\tilde{\epsilon}} \cdot \frac{1}{\gamma+\nu})$ is decreasing in γ, while the EEC can increase/decrease in ν if γ is high/small enough (increase in ν if $\gamma > \frac{\sigma}{\zeta}$). With endogenous returns, not only EEC but also additional general equilibrium channel in the EEC^{GE} depends on different sources of inequality.

To see how the general equilibrium channel responds to each source, Proposition **??** in the main text provides comparative statics of the price elasticity $\epsilon_{w,1-\tau}$. With the analytical expression of $\frac{\partial \log \Psi}{\partial \log(1-\tau)}$ we have derived, it is straightforward to obtain

$$\frac{\partial \epsilon_{w,1-\tau}}{\partial \nu} = \frac{\partial}{\partial \nu}\left(\frac{\partial \log \Psi}{\partial \log(1-\tau)}\right) = \frac{\frac{\zeta}{(1-(\sigma+\nu\zeta))^2} \cdot \frac{1}{\eta}\left(1 - \frac{1}{\eta}\right)}{\left[1 - \frac{1}{\eta} \cdot \left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)\right]^2} > (<)0 \quad \text{if } \eta > (<)1, \text{ and}$$

$$\frac{\partial \epsilon_{w,1-\tau}}{\partial \gamma} = \frac{\partial}{\partial \gamma}\left(\frac{\partial \log \Psi}{\partial \log(1-\tau)}\right) = \frac{\frac{\zeta(1+\phi\zeta)}{(1-(\sigma-\gamma\zeta))^2} \cdot \frac{1}{\eta}\left(1 - \frac{1}{\eta}\right)}{\left[1 - \frac{1}{\eta} \cdot \left(\frac{1}{1-(\sigma+\nu\zeta)} - \frac{(\sigma-\gamma\zeta)+\phi\zeta}{1-(\sigma-\gamma\zeta)}\right)\right]^2} > (<)0 \quad \text{if } \eta > (<)1.$$

In addition, we have $\frac{\partial \tilde{\epsilon}}{\partial \nu} = \frac{\zeta}{(1-(\sigma+\nu\zeta))^2} > 0$. Combining these results with the assumption $\eta > 1$ implies that either increase in ν or increase in γ provides an additional force of raising the EEC^{GE}

through the general equilibrium effects. For expositional simplicity, suppose that $\epsilon_{w,1-\tau}$ is positive before a tax reform which is the case under the baseline calibration of our numerical analysis. In this case, introduction of general equilibrium effect itself increases the effective efficiency cost relative to the cost under the exogenous price. Proposition **??** implies that when either return to human capital increases or return to skill increases, a tax reform leads to a larger increase in prices amplifying the positive GE effects. This is because with higher return on the human capital or skill, the relative scarcity of effective labor of high e's becomes smaller leading to more dominant demand effects of a tax reform.

We also remark that this impact of inequality on the EEC^{GE} through the price elasticity does not depend on the initial sign of the price elasticity. Regardless of the sign of $\epsilon_{w,1-\tau}$, its change in response to the increase in γ or ν is positive, which provides additional force of raising the EEC^{GE} and decreasing τ_{ub}.

B.4 Analysis of an Extended Model in Section ??

We now analyze a model provided in Section **??** of the main text where both returns to human capital and effective labor supply are endogenized. As in the simple model in Section **??**, we endogenize the price of human capital with imperfect substitution of effective labor inputs of different human capital. In addition, we now incorporate the endogenous superstar effects. Do we still have the neutrality of the superstar effects? If there is any interaction between the return to human capital and superstar effects, then the neutrality result would not hold any more, but what is the key for this interaction? To answer this question, we investigate the extended model in Section **??**.

Entrepreneur's Problem and Proof of Lemma ?? In Section **??**, we have shown that the entrepreneur's earnings schedule $W(y|w)$ is determined by the entrepreneur's profit maximization (**??**), where the cost function $c(Y|w)$ is given by

$$c(Y|w) = \min_{\{\ell(e)\}} \int w(e) \cdot \ell(e) de \qquad s.t. \quad Y = \left\{ \int e^{\psi} \cdot \ell(e)^{\frac{\eta-1}{\eta}} \right\}^{\frac{\eta}{\eta-1}}, \tag{109}$$

and we have denoted the Lagrange multiplier of the production constraint by $\lambda_c(Y)$ and the optimal solution by $\{\ell(e|Y,w)\}$. We also denote the solution of the entrepreneur's profit maximization (**??**) by $Y(y(x)|w)$ and we denote the labor demand associated with $y(x)$ by $\ell_D(e|y(x),w) \equiv \ell(e|Y(y(x)|w),w)$. The aggregate effective labor of workers in the entire economy is denoted by $\bar{Y} = \int Y(y(x)|w)g(x)dx$ and the labor demand of human capital level e is $\ell_D(e) = \int \ell_D(e|y(x),w)g(x)dx$.

Lemma **??** states that $\lambda_c(Y) = 1$ for all Y. We prove this in two steps. We first show that the Lagrange multiplier $\lambda_c(Y)$ does not depend on the level of Y, and then we prove that $\lambda_c = 1$.

First, we show that under the CES aggregator of labor inputs, $\lambda_c(Y) = \lambda_c$ for $\forall Y > 0$ and the cost function is linear $c(Y|w) = \lambda_c \cdot Y$ with the constant marginal cost λ_c of aggregate labor Y. Consider a level of aggregate labor input $\hat{Y}$. We can calculate the cost function:

$$\begin{aligned} c(\hat{Y}|w) &= \int w(e) \cdot \ell(e|\hat{Y}, w) de \\ &= \int \lambda_c(\hat{Y}) \cdot \hat{Y}^{\frac{1}{\eta}} \cdot e^{\psi} \cdot \ell(e|\hat{Y}, w)^{\frac{\eta-1}{\eta}} de \\ &= \lambda_c(\hat{Y}) \cdot \hat{Y}^{\frac{1}{\eta}} \cdot \int e^{\psi} \cdot \ell(e|\hat{Y}, w)^{\frac{\eta-1}{\eta}} de \\ &= \lambda_c(\hat{Y}) \cdot \hat{Y}, \end{aligned}$$

where the second equality is obtained by using the first-order conditions of the cost minimization (109), $w(e) = \lambda_c(Y) \cdot Y^{\frac{1}{\eta}} \cdot e^{\psi} \cdot \ell(e|Y, w)$. Then for any level of aggregate labor $\tilde{Y} = k \cdot \hat{Y}$ for some $k > 0$, the cost of production is

$$c(\tilde{Y}|w) = \lambda_c(\tilde{Y}) \cdot \tilde{Y} = k \cdot \lambda_c(\tilde{Y}) \cdot \hat{Y}. \tag{110}$$

On the other hand, the ratio of optimal $\ell(e)$'s across e's does not depend on Y and we can compute the cost $c(\tilde{Y}|w)$ using this fact. Consider two different levels of human capital $\hat{e}$ and $\tilde{e}$. Combining the first-order conditions of (109) with respect to $\ell(\hat{e})$ and $\ell(\tilde{e})$ yields

$$\frac{\ell(\hat{e}|Y, w)}{\ell(\tilde{e}|Y, w)} = \left(\frac{\hat{e}^{\psi}}{\tilde{e}^{\psi}}\right)^{\eta} \times \left(\frac{w(\tilde{e})}{w(\hat{e})}\right)^{\eta},$$

which shows that the ratio of $\ell(e)$'s is independent of the level of aggregate labor Y under the CES aggregator. Then the optimal solution to (109) associated with $\tilde{Y} = k\hat{Y}$ should satisfy $\ell(e|\tilde{Y}, w) = \alpha \cdot \ell(e|\hat{Y}, w)$ for some constant $\alpha > 0$, for $\forall e$, and $\alpha = k$ because

$$\tilde{Y} = k\hat{Y} = \left[\int e^{\psi} \cdot \left(\alpha \cdot \ell(e|\hat{Y})\right)^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} = \alpha \times \left[\int e^{\psi} \cdot \ell(e|\hat{Y})^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} = \alpha \cdot \hat{Y}.$$

Thus,

$$c(\tilde{Y}|w) = \int w(e) \cdot \ell(e|\tilde{Y}, w) de = \int w(e) \cdot k \cdot \ell(e|\hat{Y}, w) de = k \cdot \lambda_c(\hat{Y}) \cdot \hat{Y}. \tag{111}$$

Comparing (110) and (111) yields $\lambda_c(\tilde{Y}) = \lambda_c(\hat{Y})$. This implies that the Lagrange multiplier $\lambda_c(Y)$ does not depends on Y and we denote the constant multiplier by λ_c.

Next, we prove that $\lambda_c = 1$. By integrating the cost of all firms, the aggregate cost is

$$\begin{aligned} \int c(Y(y(x)|w)|w) \cdot g(x) dx &= \iint w(e) \cdot \ell_D(e|y(x), w) \cdot g(x) de dx \\ &= \int w(e) \cdot \left(\int \ell_D(e|y(x), w) \cdot g(x) dx\right) de \\ &= \int w(e) \cdot \ell_D(e) de \end{aligned} \tag{112}$$

On the other hand, using the result that cost function is linear with constant marginal cost λ_c, the aggregate cost is

$$\begin{aligned}
\int c(Y(y(x)|w)|w) \cdot g(x)dx &= \lambda_c \times \int Y(y(x)|w) \cdot g(x)dx \\
&= \lambda_c \times \int y(\theta) f(\theta) d\theta \\
&= \lambda_c \times \int w(e(\theta)) \cdot \theta^\gamma \cdot n(\theta) f(\theta) d\theta \\
&= \lambda_c \times \int w(e) \cdot \left(e^{-1}(e)\right)^\gamma \cdot n(e) m(e) de \\
&= \lambda_c \times \int w(e) \cdot \ell_S(e) de \qquad (113)
\end{aligned}$$

By equating (112) and (113) and using the market clearing $\ell_D(e) = \ell_S(e)$, we obtain $\lambda_c = 1$. The remaining proof of Lemma **??** is provided below.

Planner's Problem and Pareto Efficient Tax Formula We first express the labor demand curve with respect to $\ell_D(e)$. From the first order condition of (109), we have $\left(\frac{w(e)}{e^\psi}\right)^\eta = \frac{Y(y(x)|w)}{\ell_D(e|y(x),w)}$ for all x, which implies that the ratio $\frac{\ell_D(e|y(x),w)}{Y(y(x)|w)}$ is independent of x and is equal to some constant $k(e)$. Then $\frac{\ell_D(e|y(x),w)}{Y(y(x)|w)} = k(e) = \frac{\ell_D(e)}{\bar{Y}}$, and thus the labor demand curve $\ell_D(e)$ satisfies $w(e) = e^\psi \times \left(\frac{Y}{\ell_D(e)}\right)^{\frac{1}{\eta}}$. On the other hand, labor supply curve is expressed as $\ell_S(e) = \left(e^{-1}(e)\right)^\gamma \cdot n(e^{-1}(e)) \cdot f(e^{-1}(e)) \cdot \frac{1}{e'(e^{-1}(e))}$, where $e(\theta)$ and $n(\theta)$ is the optimal solution of the worker's problem.

We denote the utility of entrepreneur and worker by

$$v(\theta) = U(c(\theta), \tilde{y}(\theta), \theta; \tilde{p}) \quad \text{and} \quad \tilde{v}(x) = \tilde{U}(c(x), y(x), x; \tilde{p}).$$

Then the planner solves the following problem:

$$\begin{aligned}
\max_{v(\theta), \tilde{v}(x), \tilde{y}(\theta), y(x)} \quad & \int \Lambda_\theta(\theta) v(\theta) dF(\theta) + \int \Lambda_x(x) \tilde{v}(x) dG(x) \\
s.t. \quad & v'(\theta) = U_\theta(c(\theta), \tilde{y}(\theta), \theta; \tilde{p}) \\
& \tilde{v}'(x) = \tilde{U}_x(c(x), y(x), x; \tilde{p}) \\
& \int C(v(\theta), \tilde{y}(\theta) dF(\theta) + \int C(\tilde{v}(x), y(x), x) dG(x) = \int A(y(x), Y(y(x)|w)) g(x) dx \qquad (114) \\
& \int Y(y(x)|w) g(x) dx = \left\{ \int \tilde{y}(\theta) f(\theta) d\theta \right\}^{\frac{\eta}{\eta-1}}, \qquad (115)
\end{aligned}$$

where $\tilde{p}(\theta, e) = \theta^\gamma \cdot \frac{w(e)}{\bar{Y}^{\frac{1}{\eta}}}$, and $w(e)$ schedule is determined to satisfy $\ell_D(e) = \ell_S(e)$.

Observe that the aggregate labor $Y(y(x)|w)$ only appears in the constraints (114) and (115).

Then we can rewrite the planning problem in the following way:

$$\begin{aligned}
\max_{v(\theta),\tilde{v}(x),\tilde{y}(\theta),y(x)} \quad & \int \Lambda_\theta(\theta)v(\theta)dF(\theta) + \int \Lambda_x(x)\tilde{v}(x)dG(x) \\
(\mu(\theta)) \quad s.t. \quad & v'(\theta) = U_\theta(c(\theta), \tilde{y}(\theta), \theta; \tilde{p}) \\
(\tilde{\mu}(x)) \quad & \tilde{v}'(x) = \tilde{U}_x(c(x), y(x), x; \tilde{p}) \\
(\lambda) \quad & \int C(v(\theta), \tilde{y}(\theta)dF(\theta) + \int C(\tilde{v}(x), y(x), x)dG(x) = \Phi\left(\{y(x)\}, \{\tilde{y}(\theta)\}\right)
\end{aligned}$$

where

$$\begin{aligned}
\Phi\left(\{y(x)\}, \{\tilde{y}(\theta)\}\right) \quad = \quad & \max_{Y(y(x))} A(y(x), Y((y(x))))g(x)dx \qquad (116) \\
(\lambda_\Phi) \quad s.t. \quad & \int Y(y(x))g(x)dx = \left\{\int \tilde{y}(\theta)f(\theta)d\theta\right\}^{\frac{\eta}{\eta-1}},
\end{aligned}$$

and $w(e)$ satisfies $\ell_D(e) = \ell_S(e)$. The first-order condition of (116) is

$$A_Y(y(x), Y(y(x))) = \lambda_\Phi,$$

and from the the first-order condition (**??**) of the entreprenuer's profit maximization, we have $A_Y(y(x), Y(y(x))) = \lambda_c = 1$, and thus $\lambda_\Phi = 1$.

Before solving the planning problem, we finish the proof of Lemma **??** by deriving the derivatives of the aggregate production function. The envelope condition of (116) yields

$$\frac{\partial \Phi}{\partial y(x)} = A_y(y(x), Y(y(x))) \cdot g(x) = W'(y(x)|w) \cdot g(x),$$

where the second equality is from the envelope condition of (**??**). Another envelope condition is

$$\frac{\partial \Phi}{\partial \tilde{y}(\theta)} = \lambda_\Phi \cdot \frac{\eta}{\eta-1} \cdot Y^{\frac{1}{\eta}} \cdot f(\theta) = A_Y(y(x), Y(y(x))) \cdot \frac{\eta}{\eta-1} \cdot Y^{\frac{1}{\eta}} \cdot f(\theta) = \frac{\eta}{\eta-1} \cdot Y^{\frac{1}{\eta}} \cdot f(\theta).$$

Using these expressions, we will replace $\frac{\partial \Phi}{\partial y(x)}$ and $\frac{\partial \Phi}{\partial \tilde{y}(\theta)}$ terms after deriving the first order condition of (the part of) the planning problem given the incidence on the production function to obtain the Pareto efficient tax formula.

As in Appendix B.3, it is convenient to rewrite the incentive constraints. Following the same steps as in Appendix B.3, we can rewrite the worker's incentive constraints as

$$v'(\theta) = -U_{\tilde{y}}(\theta)\frac{\tilde{y}(\theta)}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right), \quad \forall \theta$$

We also rewrite the entrepreneur's incentive constraints. Since $\tilde{U}(c, y, x) = \tilde{u}\left(c, \frac{y}{x}\right)$, the derivative of $\tilde{U}$ satisfies $\tilde{U}_x = -\tilde{u}_n \cdot \frac{y}{x^2} = -\tilde{U}_y \frac{y}{x}$, and thus the incentive constraints of the entrepreneurs are

$$\tilde{v}'(x) = -\tilde{U}_y(c(x), y(x), x) \cdot \frac{y(x)}{x}, \quad \forall x.$$

We then write the Lagrangian of the part of the planner's problem using these incentive constraints:

$$\begin{aligned} L &= \int \Lambda_\theta v(\theta) dF(\theta) + \int \Lambda_x(x)\tilde{v}(x) dG(x) \\ &\quad + \lambda \times \left[\Phi(\{y(x)\},\{\tilde{y}(\theta)\}) - \int C(v(\theta),\tilde{y}(\theta),\theta) f(\theta) d\theta - \int C(\tilde{v}(x),y(x),x) g(x) dx \right] \\ &\quad - \int \mu'(\theta) v(\theta) d\theta + \int \mu(\theta) U_{\tilde{y}}(\theta) \frac{\tilde{y}(\theta)}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) \\ &\quad - \int \tilde{\mu}'(x)\tilde{v}(x) dx + \int \tilde{\mu}(x) \tilde{U}_y(x) \frac{y(x)}{x} dx. \end{aligned}$$

The planner maximizes the Lagrangian L subject to the remaining constraints (116) and $\ell_D(e) = \ell_S(e)$.

We again denote the derivatives holding the productivity function $\tilde{p}(\cdot,\cdot)$ fixed by $U_{\tilde{y}\tilde{y}}$, $U_{c\tilde{y}}$, MRC_c and MRS_y. We also use the same notation $\epsilon_{\tilde{p},\tilde{y}}$ as in Appendix B.3 for the elasticity of $\tilde{p}$ with respect to $\tilde{y}$ holding e constant. The first-order condition of the Lagrangian with respect to $\tilde{y}(\theta)$ given the derivatives of the production function Φ and the endogenous price $\tilde{p}$ is

$$\begin{aligned} &\lambda \cdot \left(\frac{\partial \Phi}{\partial \tilde{y}(\theta)} - C_{\tilde{y}}(\theta) f(\theta) \right) \\ &= -\mu(\theta) \times \left[\begin{array}{l} \left\{ U_{\tilde{y}c}(\theta) \cdot C_{\tilde{y}}(\theta) + U_{\tilde{y}\tilde{y}}(\theta) + \frac{\partial U_{\tilde{y}}(\theta)}{\partial \tilde{p}(\theta)} \epsilon_{\tilde{p},\tilde{y}}(\theta) \frac{\tilde{p}(\theta)}{\tilde{y}(\theta)} \right\} \cdot \frac{\tilde{y}(\theta)}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) \\ + U_{\tilde{y}}(\theta) \cdot \frac{1}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) + U_{\tilde{y}}(\theta) \cdot \frac{\tilde{y}(\theta)}{\theta} \cdot \frac{\partial \nu(\theta)}{\partial \tilde{y}(\theta)} \end{array} \right], \end{aligned}$$

where $\nu(\theta) = \frac{\tilde{p}_e e}{\tilde{p}}$.

We then use Lemma **??** to replace the derivatives of production function $\frac{\partial \Phi}{\partial \tilde{y}(\theta)}$. Define $\hat{\mu}(\theta) = \frac{\mu(\theta) U_c}{\lambda}$ and $MRS(c,y,\theta) = -\frac{U_{\tilde{y}}(\theta)}{U_c(\theta)} \times \frac{1}{\bar{Y}^{\frac{1}{\eta}}}$. Following the same process as in Appendix B.3, we can rewrite the first-order condition as follows:

$$\left(\frac{T'(\omega(\theta))}{1 - T'(\omega(\theta))} + \frac{1}{\eta - 1} \cdot \frac{1}{1 - T'(\omega(\theta))} \right) \times \theta f(\theta) = \hat{\mu}(\theta) \times \left[\begin{array}{l} \frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) + \frac{\partial \nu(\theta)}{\partial \tilde{y}(\theta)} \cdot \tilde{y}(\theta) \\ - \frac{1+\tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)} \epsilon_{\tilde{p}\tilde{y}}(\theta) \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) \end{array} \right],$$

which yields EEC^{GE} exactly as in Appendix B.3 (Equations (96) and (97), and the tax formula for workers also does not change from the one in Appendix B.3.)

On the other hand, the first-order conditions of the Lagrangian with respect to $y(x)$ yields

$$\begin{aligned} \lambda \cdot \left(\frac{\partial \Phi}{\partial y(x)} - C_y(x) \cdot g(x) \right) &= -\tilde{\mu}(x) \times \left[\left\{ \tilde{U}_{yc}(x) \cdot C_y(x) + \tilde{U}_{yy}(x) \right\} \cdot \frac{y(x)}{x} + \tilde{U}_y(x) \cdot \frac{1}{x} \right] \\ &\quad - \int \mu(\theta) \times \left[\begin{array}{l} \frac{\partial U_{\tilde{y}}(\theta)}{\partial \tilde{p}(e(\theta))} \epsilon_{\tilde{p},y(x)}(\theta) \cdot \frac{\tilde{p}(e(\theta))}{y(x)} \cdot \frac{\tilde{y}(\theta)}{\theta} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}} \right) \\ + U_{\tilde{y}}(\theta) \cdot \frac{\tilde{y}(\theta)}{\theta} \cdot \frac{\partial \nu(\theta)}{\partial y(x)} \end{array} \right] d\theta, \end{aligned}$$

where $\frac{\partial \nu(\theta)}{\partial y(x)}$ again denotes the endogenous incidence on $\nu(\theta)$. Define $\hat{\tilde{\mu}}(x) = \frac{\tilde{\mu}(x)\tilde{U}_c(x)}{\lambda}$ and $MRS(x) = -\frac{\tilde{U}_y(x)}{\tilde{U}_c(x)}$. Using Lemma **??** and $C_y(x) = -\frac{\tilde{U}_y(x)}{\tilde{U}_c(x)} = MRS(x)$, we obtain

$$\begin{aligned}\frac{A_y(x) - MRS(x)}{MRS(x)} \times xg(x) \quad &= \quad \hat{\tilde{\mu}}(x) \times \left[1 + \left(\frac{\tilde{U}_{yy}(x)}{\tilde{U}_y(x)} - \frac{\tilde{U}_{yc}(x)}{\tilde{U}_c(s)}\right) \cdot y(x)\right] \\ &+ \frac{x}{MRS(x)} \times \int \hat{\mu}(\theta) \cdot \frac{MRS(\theta)}{\theta} \cdot \left[\begin{array}{l} \frac{\bar{Y}^{\frac{1}{\eta}} \cdot \tilde{p}(\theta)}{U_{\tilde{y}}(\theta)} \cdot \frac{\partial \tilde{U}_{\tilde{y}}(\theta)}{\partial \tilde{p}(\theta)} \cdot \epsilon_{\tilde{p},y(x)}(\theta) \cdot \frac{\tilde{y}(\theta)}{y(x)} \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) \\ + \bar{Y}^{\frac{1}{\eta}} \cdot \tilde{y}(\theta) \cdot \frac{\partial \nu(\theta)}{\partial y(x)} \end{array}\right] d\theta.\end{aligned}$$

First-order condition of the entrepreneur's problem (11) can be written as $(1 - T'(\omega(x))) \times W'(y(x)|w) = MRS(x)$. Combining this with $W'(y(x)|w) = A_y(x)$, we have $\frac{A_y(x) - MRS(x)}{MRS(x)} = \frac{T'(\omega(x))}{1 - T'(\omega(x))}$. Note that (83) also holds for the elasticity and MRS of the entrepreneurs, and $\frac{MRS_y(x) \cdot y(x)}{MRS(x)} = \left(\frac{\tilde{U}_{yy}(x)}{\tilde{U}_y(x)} - \frac{\tilde{U}_{cy}(x)}{\tilde{U}_c(x)}\right) \cdot y(x)$, which yields $1 + \left(\frac{\tilde{U}_{yy}(x)}{\tilde{U}_y(x)} - \frac{\tilde{U}_{yc}(x)}{\tilde{U}_c(x)}\right) \cdot y(x) = \frac{1 + \tilde{\epsilon}^u(x)}{\tilde{\epsilon}^c(x)}$. Combining these equations with (94) and (95), the first-order condition of the planner's problem with respect to $y(x)$ can be written as:

$$\begin{aligned}\frac{T'(\omega(\theta))}{1 - T'(\omega(\theta))} \times xg(x) \quad &= \quad \hat{\tilde{\mu}}(x) \times \frac{1 + \tilde{\epsilon}^u(x)}{\tilde{\epsilon}^c(x)} \\ &+ \int \hat{\mu}(\theta) \frac{x}{MRS(x)} \frac{MRS(\theta)}{\theta} \cdot \bar{Y}^{\frac{1}{\eta}} \cdot \frac{\tilde{y}(\theta)}{y(x)} \left[\begin{array}{l} -\left(\frac{1 + \tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)}\right) \cdot \epsilon_{\tilde{p},y(x)}(\theta) \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) \\ + \frac{\partial \nu(\theta)}{\partial y(x)} y(x) \end{array}\right] d\theta.\end{aligned}$$

Finally, in this economy, we can show that $\frac{\partial \nu(\theta)}{\partial y(x)} = 0$. Recall that aggregate labor demand curve is expressed as $w(e) = e^\psi \times \left(\frac{\bar{Y}}{\ell_D(e)}\right)^{\frac{1}{\eta}}$, and thus the ratio of relative prices between two different levels of human capital is $\frac{w(\hat{e})}{w(\tilde{e})} = \frac{\hat{e}^\psi}{\tilde{e}^\psi} \times \left(\frac{\ell_D(\tilde{e})}{\ell_D(\hat{e})}\right)^{\frac{1}{\eta}}$. As we showed in the proof of Lemma **??**, the ratio $\frac{\ell_D(\tilde{e})}{\ell_D(\hat{e})}$ does not depend on $y(x)$, which implies that the return to human capital $\nu(\theta)$ is independent of $y(x)$. Thus, the first-order condition yields

$$\begin{aligned}\hat{\tilde{\mu}}(x) \quad &= \quad \frac{T'(\omega(\theta))}{1 - T'(\omega(\theta))} \times xg(x) \frac{\tilde{\epsilon}^c(x)}{1 + \tilde{\epsilon}^u(x)} \\ &- \frac{\tilde{\epsilon}^c(x)}{1 + \tilde{\epsilon}^u(x)} \int \hat{\mu}(\theta) \frac{x}{MRS(x)} \frac{MRS(\theta)}{\theta} \cdot \bar{Y}^{\frac{1}{\eta}} \cdot \frac{\tilde{y}(\theta)}{y(x)} \left[\ \left(\frac{1 + \tilde{\epsilon}^u(\theta)}{\tilde{\epsilon}^c(\theta)}\right) \cdot \epsilon_{\tilde{p},y(x)}(\theta) \cdot \left(\frac{\tilde{p}_e \cdot e}{\tilde{p}} + \frac{\tilde{p}_\theta \cdot \theta}{\tilde{p}}\right) \ \right] d\theta. \quad (117)\end{aligned}$$

Combining this with the first order conditions with respect to $\tilde{v}(x)$ yields the Pareto efficiency test formula. Assuming no income effects, the formula for the entrepreneur's earnings is $\frac{\partial x}{\partial \theta} R(x) \leq 0$ for all x, where $R(x) = 1 - G(x) - \hat{\tilde{\mu}}(x)$. The tax formula for the entrepreneur's earnings is different from the standard tax formula with exogenous wage due to the incidence on the price of human capital associated with the perturbation in entrepreneur's effective effort—captured in $\tilde{\epsilon}_{\tilde{p},y(x)}$.