Redistributive Capital Taxation Revisited

Özlem Kina, Ctirad Slavík and Hakki Yazici*

September 15, 2022

This paper uses a rich quantitative model with endogenous skill acquisition to show that capital-skill complementarity provides a quantitatively significant rationale to tax capital for redistributive governments. The optimal capital income tax rate is 67%, while it is 61% in an identically calibrated model without capital-skill complementarity. The skill premium falls from 1.9 to 1.84 along the transition following the optimal reform in the capital-skill complementarity model, implying substantial indirect redistribution from skilled to unskilled workers. These results show that a redistributive government should take into account capital-skill complementarity when taxing capital.

JEL classification: E25, J31.

Keywords: Capital taxation, capital-skill complementarity, inequality, redistribution.

*Özlem Kina, University of Edinburgh, 30 Buccleuch Place, EH8 9JT, Edinburgh, UK. Email:okina@exseed.ed.ac.uk. Ctirad Slavík, CERGE-EI. A joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Czech Academy of Sciences. Politických vězňů 7, Praha 1, 111 21, Czech Republic. Email: ctirad.slavik@cerge-ei.cz. Hakki Yazici, University of Bristol. 12A Priory Rd, Bristol BS8 1TU, UK. Email: hakki.yazici@bristol.ac.uk.

We would like to thank Árpád Ábrahám, Ivo Bakota, Marek Kapička, Dirk Krueger, Remzi Kaygusuz, Baris Kaymak, Alex Ludwig, Ramon Marimon, Alex Monge-Naranjo, Gonul Sengul, Faisal Sohail, Dušan Stojanović, Hitoshi Tsujiyama, Gianluca Violante, and participants of various conferences, workshops and seminars for helpful suggestions and comments. We would also like to thank Andrea Downing for excellent English editing. Slavík acknowledges support from the Czech Science Foundation (grant project 17-27676S) and from the Ministry of Education, Youth and Sports of the Czech Republic through the project NPO "Systemic Risk Institute" (LX22NPO5101).

1 Introduction

The optimal tax rate on capital income has long been debated. Supporters of capital tax cuts stress the efficiency costs associated with capital taxation, mainly the slowing down of capital accumulation and, hence, reduced output growth. Proponents of higher capital taxes often cite their redistributive benefits: as wealth is often unequally distributed across people, increasing capital taxes in favor of lower labor taxes decreases after-tax inequality. Aiyagari (1995) and Domeij and Heathcote (2004), among others, show that the redistributive benefits of capital taxation can be large enough to justify significant optimal tax rates on capital income. In this paper, we contribute to the debate on optimal capital taxation by putting forward a mechanism through which capital taxes lead to additional redistributive benefits and by quantifying the implications of this mechanism on the optimal capital tax rate. We find that the proposed mechanism implies that the optimal tax rate on capital income should be considerably higher than the conventional economic models tell us.

At the heart of this mechanism is the assumption of capital-skill complementarity, which is the idea that capital is relatively more complementary with skilled labor than it is with unskilled labor.[1] A rise in the capital tax rate depresses capital accumulation, which then decreases the relative demand for skilled workers due to capital-skill complementarity. As a result, the skill premium - i.e., the wages of the skilled workers relative to those of the unskilled workers - declines. Since skilled workers normally earn higher wages and have more assets, this decline in the skill premium increases social welfare from the perspective of a government that values equality.

We measure the quantitative significance of this mechanism for the optimal capital tax rate using a model that embeds capital-skill complementarity into an incomplete markets model à la Bewley (1986), Imrohoroglu (1989), Huggett (1993) and Aiyagari (1994), where individuals face idiosyncratic wage risk and make a once-and-for-all skill

[1] Capital-skill complementarity was first empirically documented by Griliches (1969). It has received much attention from economists and has been successfully used in explaining the evolution of inequality in the returns to education. Among others, see Fallon and Layard (1975), Krusell, Ohanian, Ríos-Rull, and Violante (2000), Flug and Hercowitz (2000), and Duffy, Papageorgiou, and Perez-Sebastian (2004).

acquisition decision. We choose this model as it allows for a sufficiently rich modeling of earnings and wealth inequality, which is key to accurately assessing the redistributive benefits of capital taxation. We consider two versions of the model that differ from each other only in terms of the aggregate production functions. In the first economy, we model capital-skill complementarity (CSC) by assuming a production function that features a higher degree of complementarity between equipment capital and skilled labor than between equipment capital and unskilled labor, as documented empirically for the U.S. economy by Krusell, Ohanian, Ríos-Rull, and Violante (2000). As a benchmark for comparison, we also build a second economy with a standard Cobb-Douglas (CD) production function that does not feature capital-skill complementarity. We make the two model economies comparable by calibrating each one separately to the current U.S. economy along selected dimensions under the status-quo capital and labor tax system.

We consider the case of a government that chooses a linear tax rate on capital income to maximize a Utilitarian social welfare function with equal weights on all agents. The government takes into account the effect of tax changes on people's welfare over the transition to the new steady state. We find that the optimal capital tax rate for the capital-skill complementarity economy is significantly higher than that in the Cobb-Douglas economy, with respective optimal rates of 67% vs. 61%. Accordingly, the average labor income tax is lower in the economy with capital-skill complementarity. In response to the optimal tax reform, the skill premium falls from 1.90 to as low as 1.84 over the transition and to a final steady-state level of about 1.86 in the capital-skill complementarity economy. Meanwhile it remains virtually unchanged in the Cobb-Douglas economy. Since labor income taxes are distortionary, this indirect redistribution channel is valuable for the government and gives rise to a higher optimal capital tax rate in the economy with capital-skill complementarity. This finding shows that the debate over the optimal tax rate on capital income should take into account the presence of capital-skill complementarities in production.

Under the Utilitarian social welfare function, the welfare gains of the reform are equivalent to those of increasing consumption of all agents by 1.25% at every date and state in the economy with capital-skill complementarity. The corresponding welfare gains amount to 0.85% in the Cobb-Douglas economy. This difference in welfare gains implies that carrying out the optimal capital tax reform is considerably more important once capital-skill complementarity is taken into account. A welfare decomposition exercise reveals that the main gain of the reform is redistribution in both models, and as expected, this gain is higher in the model with capital-skill complementarity.

Through an extensive sensitivity analysis, we show that our results are quantitatively robust to an alternative degree of capital-skill complementarity estimated using more recent data, a lower level of elasticity of labor supply, and alternative specifications of the social welfare function.

While in the baseline reform, the government chooses a uniform tax rate on the two types of capital, Section 6 considers optimal tax reforms with more flexible instruments. These reforms are: a reform in which the government in the capital-skill complementarity economy can set different tax rates on different types of capital (equipment and structures), (ii) a comprehensive reform in which the government chooses the degree of labor tax progressivity in addition to the capital tax rate, and finally, (iii) a reform in which the tax rate on capital can vary over time. We find that the indirect redistribution channel of capital taxation is at work in all these reforms.

Related Literature. Taxation of capital income is a controversial topic in the macroeconomics literature. In the representative-agent paradigm, Chamley (1986) and Judd (1985) show that it is optimal not to tax capital at all in the long run.[2] Aiyagari (1995) shows that the optimal long-run capital income tax might be positive when there is heterogeneity across agents arising from uninsured labor income risk and incomplete markets. He points out that the optimal steady state capital income tax is between 25% and 45% depending on the values of various model parameters.[3] Domeij and Heathcote (2004)

[2]Straub and Werning (2020) provide a set of conditions under which the optimality of zero taxes on capital in the long run does not hold. Chari, Nicolini, and Teles (2020) show that with a richer set of tax instruments and under the assumption that initial confiscation of wealth is restricted, one recovers the long-run optimality of zero capital taxes.

[3]These numerical results are not included in the published version of the paper, and are only available in a working paper version. This version is available as Minneapolis Fed Working Paper Series #508. Moreover, Aiyagari (1995) only reports optimal taxes at the steady state. Recently, Acikgoz, Hagedorn, Holter, and Wang (2018) and Dyrda and Pedroni (2022) calculate time-varying paths of optimal capital and labor taxes in environments with uninsurable wage risk.

investigate the quantitative importance of heterogeneity and idiosyncratic labor income risk for capital taxation using an Aiyagari (1994) model. They consider the problem of a redistributive government that needs to choose constant (time-independent) tax rates on capital and labor income. They find that eliminating capital income taxes altogether brings large welfare gains if they assume a representative-agent economy. However, when there is heterogeneity and risk, the optimal capital tax rate can be quite high, namely 40%, according to their calculations. Imrohoroglu (1998) and Conesa, Kitao, and Krueger (2009) also analyze optimal capital taxation in quantitative models with rich heterogeneity and, in particular, a life cycle structure. Conesa, Kitao, and Krueger (2009) find that, due to the life-cycle structure, optimal capital taxes can be significantly positive at 36% even when the government maximizes steady-state welfare.[4] We add to this literature by assessing the quantitative impact of capital-skill complementarity on optimal capital taxation.

There is also a more recent and growing literature on taxation of capital in the presence of capital-skill complementarity. Jones, Manuelli, and Rossi (1997) provide an important backdrop to this literature. In an extension section, the authors analyze optimal linear taxation in a growth model with two types of labor, skilled and unskilled, and show that the optimal long-run capital tax rate may be positive if the labor income tax rate is not allowed to depend on skill type and there is capital-skill complementarity. The key difference between the current paper and that of Jones, Manuelli, and Rossi (1997) is that we quantitatively evaluate the effect of capital-skill complementarity for optimal capital tax rate in a model that allows for a rich modelling of earnings and wealth inequality, whereas they use a simple, representative agent framework to make a qualitative statement. Slavík and Yazici (2014) also build a model with capital-skill complementarity; however, they use it to study the optimality of differential capital taxation. They find that in their private information Mirrleesian model it is optimal to tax equipment at a higher rate

[4]See also the New Dynamic Public Finance literature, which has followed the seminal contribution of Golosov, Kocherlakota, and Tsyvinski (2003), for investigations of optimal capital taxation in dynamic Mirrlesian private information models with idiosyncratic labor income shocks.

than structures.[5] Bhattarai, Lee, Park, and Yang (2020) analyze the macroeconomic effects of specific capital tax reforms under capital-skill complementarity. Angelopoulos, Asimakopoulos, and Malley (2015) use a representative agent model to evaluate the optimality of labor tax smoothing under capital-skill complementarity, while Dolado, Motyovszki, and Pappa (2020) analyze monetary policy and its redistributive implications in a New Keynesian model with capital-skill complementarity.[6]

The rest of the paper is organized as follows. Section 2 lays out the model while Section 3 describes the optimal tax problem formally. Section 4 explains the calibration strategy and Section 5 discusses the main quantitative results. Section 6 explores tax reforms with more flexible instruments. Finally, Section 7 concludes.

2 Model

The economy consists of a unit measure of individuals, a firm, and a government. In the baseline model, the aggregate production function features capital-skill complementarity. Later on, for comparison, we also consider an economy that combines capital and labor using a standard Cobb-Douglas production function.

Demographics and Worker Choices. Each period a fraction $1 - \chi$ of workers are born with zero asset holdings. Life prior to labor market entry is not modeled. At the beginning of their lives, just before they enter the labor market, agents choose their skill level once-and-for-all. They become either skilled or unskilled, denoted by $i \in \{u, s\}$. After this, they enter labor market and work, consume and save every period. Workers survive from period to another at a constant rate of χ and the assets of deceased people are distributed among the survivors in proportion to the survivors' wealth. This assumption is equivalent to assuming that people can buy actuarially fair life insurance policy.

[5] There is a growing literature which analyzes the optimal taxation of robots, see e.g. Guerreiro, Rebelo, and Teles (2021), Costinot and Werning (2018) and Thuemmel (2020).

[6] Krueger and Ludwig (2016) and Heathcote, Storesletten, and Violante (2017) are also related to the current paper in the sense that they analyze optimal taxation in models with imperfect substitutability between skilled and unskilled labor in which there are general equilibrium effects of taxation on skill prices. Importantly, neither of these studies models capital-skill complementarity.

Skill Heterogeneity and Wage Risk. Skilled agents can only work in the skilled labor sector and unskilled agents only in the unskilled labor sector. Agents of skill type i receive a wage rate $w_{i,t}$ for each unit of effective labor they supply in period t. The total mass of type i workers in period t is denoted by $\pi_{i,t}$. In the quantitative analysis, skill types correspond to educational attainment at the time of entering the labor market. Workers who have at least a bachelor degree are classified as skilled agents and the rest of the agents are classified as unskilled.

There is also ex-post heterogeneity within each skill group arising from workers facing idiosyncratic labor productivity shocks over time. The productivity shock, denoted by z, follows a type-specific Markov chain with states $Z_i = \{z_{i,1}, ..., z_{i,I}\}$ and transitions $\Pi_i(z'|z)$. The productivity shock for labor market entrants is drawn from the stationary distribution associated with the Markov chain. When a skill type i worker draws productivity level z and works l units in a period, she produces $l \cdot z$ units of effective type i labor. Her wage per unit of time is $w_{i,t} \cdot z$.

Preferences. Preferences over consumption and labor, c and l, in a period are defined using a utility function which is separable between consumption and labor: $u(c) - v(l)$, where the utility and disutility functions satisfy standard assumptions: $u', -u'', v', v'' > 0$. Also, we assume people discount utility across periods by $\beta \in (0, 1)$.

Technology. The production process is summarized by a constant returns to scale production function: $Y = F(K_s,\ K_e,\ L_s,\ L_u)$, where K_s, K_e, L_s and L_u refer to the aggregate levels of structure capital, equipment capital, effective skilled labor supply and effective unskilled labor supply, respectively. The stocks of structure and equipment capital depreciate at rates δ_s and δ_e, respectively.

We assume that there is capital-skill complementarity in the production process. More specifically, technology features equipment-skill complementarity, which means that the degree of complementarity between equipment capital and skilled labor is higher than that between equipment capital and unskilled labor. This implies that an increase in the stock of equipment capital decreases the ratio of the marginal product of unskilled labor to that of skilled labor. Under the assumption of competitive factor markets, this implies that the skill premium, defined as the ratio of skilled to unskilled wages, is increasing in equipment capital. Structure capital, on the other hand, is assumed to be neutral in terms of its complementarity with skilled and unskilled labor. These assumptions on technology are consistent with the estimation results of Krusell, Ohanian, Ríos-Rull, and Violante (2000).

Production is carried out by a representative firm, which, in each period, rents the two types of capital and hires the two types of labor to maximize profits. The firm solves the following maximization problem in period t:

$$\max_{K_{s,t},K_{e,t},L_{s,t},L_{u,t}} \quad F(K_{s,t},K_{e,t},L_{s,t},L_{u,t}) - r_{s,t}K_{s,t} - r_{e,t}K_{e,t} - w_{s,t}L_{s,t} - w_{u,t}L_{u,t}, \quad (1)$$

where $r_{s,t}$ and $r_{e,t}$ are the rental rates of structure and equipment capital and $w_{u,t}$ and $w_{s,t}$ are the wages rates paid to unskilled and skilled effective labor in period t.

Government. The government uses linear taxes on capital income net of depreciation. Let $\{\tau_t\}_{t=0}^{\infty}$ be the sequence of tax rates on capital income. It is irrelevant for our analysis whether capital income is taxed at the consumer or at the corporate level. We assume without loss of generality that all capital income taxes are paid at the consumer level. The government taxes labor income using a sequence of possibly non-linear functions $\{T_t(y)\}_{t=0}^{\infty}$, where y is labor income and $T_t(y)$ are the taxes paid by the consumer. We follow Heathcote, Storesletten, and Violante (2017) and assume that tax liability given labor income y is defined as:

$$T_t(y) = \bar{y}\left[\frac{y}{\bar{y}} - \lambda_t\left(\frac{y}{\bar{y}}\right)^{1-\tau_l}\right],$$

where $\bar{y}$ is the mean labor income in the economy, $1-\lambda_t$ is the average tax rate of a mean income individual and τ_l controls the progressivity of the tax code. When $\tau_l > 0$, labor taxes are progressive and the tax function implies transfers to people with sufficiently low income. The government uses taxes to finance a stream of expenditure $\{G_t\}_{t=0}^{\infty}$ and repay government debt $\{D_t\}_{t=0}^{\infty}$.

Asset Market Structure. Government debt is the only financial asset in the economy. It has a one period maturity and return R_t in period t. Consumers can also save through the two types of capital. In the absence of aggregate shocks, the returns to savings in the form of the two capital types are certain, as is the return on government bonds. Therefore, all three assets must yield the same after-tax return in equilibrium, $R_t = 1 + (r_{s,t} - \delta_s)(1 - \tau_t) = 1 + (r_{e,t} - \delta_e)(1 - \tau_t)$. As a result, one does not need to distinguish between savings via different types of assets in the consumer's problem. Consumers' (total) asset holdings will be denoted by a and $\mathcal{A} = [0, \infty)$ denotes the set of possible asset levels that agents can hold. Our assumptions imply that, in every period, the total savings of consumers must be equal to the total borrowing of the government plus the total capital stock in the economy.

Worker's Problem. In period t, agent of skill type i with productivity shock and asset level $(z_{i,t}, a_{i,t})$ solves:

$$v_{i,t}(z_{i,t}, a_{i,t}) = \max_{(c_{i,t}, l_{i,t}, a_{i,t+1}) \geq 0} u(c_{i,t}, l_{i,t}) + \beta\chi \sum_{z_i \in \mathcal{Z}_i} \Pi_i(z_{i,t+1}|z_{i,t}) v_{i,t+1}(z_{i,t+1}, a_{i,t+1}) \qquad \text{s.t.}$$

$$c_{i,t} + \chi a_{i,t+1} \leq w_{i,t} z_{i,t} l_{i,t} - T_t(w_{i,t} z_{i,t} l_{i,t}) + R_t a_{i,t},$$

$$c_{i,t}, a_{i,t+1} \geq 0 \text{ and } l_{i,t} \in (0,1), \tag{2}$$

where expectation is taken over the realizations of the productivity shock. The fact that assets of the deceased are distributed among the survivors in proportion to their wealth is captured by the survival probability χ multiplying $a_{i,t+1}$ in the budget constraint above.

Skill Acquisition. A tax reform that raises capital income taxes have an additional redistributive benefit in the presence of capital-skill complementarity because it reduces the skill premium. However, such a reduction in the skill premium may have adverse incentive effects on the skill acquisition decision of cohorts that make this decision after the reform. Taking this behavioral response into account is important as the implied decline in the relative number of skilled people may partially offset the decline in the skill premium, curtailing the indirect redistribution benefit of capital taxation under capital-skill complementarity. Moreover, by reducing the number of skilled workers, capital taxation may decrease average labor productivity in the economy. We model endogenous skill acquisition to account for these effects of capital taxation under capital-skill complementarity.

Newborns make a skill choice just before entering the labor market. As in Heathcote, Storesletten, and Violante (2010), there is a utility cost of attaining a college degree, $\psi \geq 0$, which is idiosyncratic and drawn from a distribution $H(\psi)$. This distribution is a reduced form way of capturing the cross-sectional variation in the psychological and pecuniary costs of acquiring a college degree such as variation in scholastic talent, tuition fees, parental resources, access to credit, and government aid programs. Upon drawing the cost of education, the agent compares this cost to the benefit of attaining a college degree, which is simply the net present utility gain of receiving the skilled wage rather than the unskilled wage in each date and state after entering the labor market. An agent born in period t chooses to become skilled if and only if

$$\psi \leq E_{s,t}[v_{s,t}(z_s, 0)] - E_{u,t}[v_{u,t}(z_u, 0)], \tag{3}$$

where expectation is taken over labor market entrants' initial productivity draw. Let $\overline{\psi}_t$ be the level of utility cost at which (3) holds with equality in period t. All agents with ψ at or below this threshold level attend college and all above do not.

Competitive Equilibrium. Before we provide a formal definition of equilibrium, it is useful to introduce some concepts and notation. The state of a worker of type i in a period t is fully described by the worker's productivity and asset holdings. Let $(z_i, a_i) \in \mathcal{Z}_i \times \mathcal{A}$ denote this state. Let $\Lambda_{i,t}(a_i, z_i)$ denote the distribution of workers of type i across productivities and assets. The initial, $t = 0$, distributions are given exogeneously.

Definition: Given initial conditions, a recursive competitive equilibrium is a government policy $(T_t(.), \tau_t, D_t, G_t)_{t=0}^{\infty}$, allocation for the firm, $(K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t})_{t=0}^{\infty}$, value and policy functions for agents, $(v_{i,t}(z_i, a_i), c_{i,t}(z_i, a_i), l_{i,t}(z_i, a_i), a_{i,t+1}(z_i, a_i))_{t=0,i=u,s}^{\infty}$, skill choices, shares of population who are skilled, $(\pi_{s,t})_{t=0}^{\infty}$, a price system $(r_{s,t}, r_{e,t}, w_{s,t}, w_{u,t}, R_t)_{t=0}^{\infty}$ and distributions over individual states, $(\Lambda_{i,t}(z_i, a_i))_{t=0,i=u,s}^{\infty}$, such that:

1. In each period $t \geq 0$, taking factor prices as given, $(K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t})$ solves the firm's problem given by (1).

2. Given government policy and the price system, the policy functions solve the consumer's problem given by (2) the solution of which defines the value functions.

3. Skill choice is consistent with (3), that is in any period t, all those with $\psi \leq \overline{\psi}_t$ attend college and all other do not. Moreover, the evolution of the fraction of skilled in each period is consistent with skill choice: $\pi_{s,t} = \chi\pi_{s,t-1} + (1-\chi)\pi_{s,t}^n$, where $\pi_{s,t}^n = \int_{\mathbb{R}_+} I_{\psi \leq \overline{\psi}_t}(\psi) dH(\psi)$ is the fraction of newborns who choose to become skilled in period t and $I_{\psi \leq \overline{\psi}_t}(\psi)$ is the indicator function, $\pi_{u,t}^n = 1 - \pi_{s,t}^n$ for all t, and $\pi_{s,0}$ is given.

4. The evolution of distributions of agents across productivities and assets over time is consistent with agent choices. That is, for all $t \geq 0$, $i = u, s$ and $(z_i', a_i') \in \mathcal{Z}_i \times \mathcal{A}$:

$$\Lambda_{i,t+1}(z_i', a_i') = \frac{\chi \sum_{z_i \in \mathcal{Z}_i} \Pi_i(z_i'|z_i) \int_{\{a_i : a_{i,t+1}(z_i,a_i) \leq a_i'\}} d\Lambda_{i,t}(z_i, a_i) + (1-\chi)\pi_{i,t+1}^n \Lambda_i^z(z_i')}{\chi + (1-\chi)\pi_{i,t+1}^n},$$

where $(\Lambda_{i,0}(z_i, a_i))_{i=u,s}$ is given and Λ_i^z is the stationary distribution associated with the Markov chain that describes the evolution of the productivity shock for type i.

5. Markets for assets, labor and goods clear: for all $t \geq 0$,

$$K_{s,t} + K_{e,t} + D_t = \sum_{i=u,s} \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} a_{i,t}(z_i, a_i) d\Lambda_{i,t-1}(z_i, a_i),$$

$$L_{i,t} = \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} l_{i,t}(z_i, a_i) z_i d\Lambda_{i,t}(z_i, a_i), \text{ for } i = u, s,$$

$$G_t + C_t + K_{s,t+1} + K_{e,t+1} = F(K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t}) + (1-\delta_s)K_{s,t} + (1-\delta_e)K_{e,t},$$

where

$$C_t = \sum_{i=u,s} \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} c_{i,t}(z_i, a_i) d\Lambda_{i,t}(z_i, a_i)$$

is aggregate consumption in period t.

6. The government's budget constraint is satisfied every period: for all $t \geq 0$,

$$G_t + R_t D_t = D_{t+1} + \sum_{j=s,e} \tau_t (r_{j,t} - \delta_j) K_{j,t} + \sum_{i=u,s} \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} T_t(l_{i,t}(z_i, a_i) w_{i,t} z_i) d\Lambda_{i,t}(z_i, a_i).$$

2.1 Cobb-Douglas Economy

To assess the quantitative significance of capital-skill complementarity for optimal capital taxes, we consider a second, benchmark, economy in which the production function does not feature capital-skill complementarity. In this economy, we do not distinguish between equipment capital and structure capital; there is only one type of capital, which depreciates every period at rate δ. First, the skilled and unskilled labor inputs are combined to give aggregate labor L. The details of how the two types of labor are aggregated will be discussed in Section 4. Next, capital and labor are combined to produce aggregate output using a standard Cobb-Douglas production function $Y = AK^{\theta} L^{1-\theta}$. We preserve all the other properties of the first model.

Importantly, under this production function, the ratio of the marginal product of skilled labor to that of unskilled labor, hence the skill premium, is independent of the amount of capital in the economy. The changes in the aggregate capital level do not affect the skill premium, therefore, capital income taxation has no direct impact on wage inequality. The definition of competitive equilibrium for this economy is very similar to that given for the capital-skill complementarity economy, and hence is relegated to Appendix A.

3 The Optimal Tax Problem

We consider the following optimal fiscal policy reform. The economy is initially at a steady state under a status-quo fiscal policy. Given the initial distribution of workers across the productivity-asset space implied by this steady state, the government introduces a once and for all unannounced change in the tax rate that applies to capital income. We assume that the levels of government spending and debt in the reform period and all the periods that follow are constant at the levels given by the initial steady state. At the same time, to ensure that its budget holds, the government adjusts the parameter that controls the average labor income tax, $\{\lambda_t\}_{t=0}^{\infty}$, along the transition to the new steady state.

In the baseline analysis, we assume that the government does not change the progressivity of the labor tax function. We do so because, perhaps due to political constraints, it is difficult for governments to carry out comprehensive reforms in which capital and labor tax codes are changed substantially at the same time. In Section 6.2, we analyze the effect of capital-skill complementarity on optimal capital taxes in the presence of such a comprehensive reform. Another assumption maintained in the baseline reform is that government is restricted to choose a capital tax rate that applies to all future dates, that is time invariant. This is a plausible assumption given that it may be harder for governments to commit to time-varying taxes. Yet, it is interesting to see the impact of capital-skill complementarity on optimal capital taxation in the presence of time-varying capital taxes. This extension is analyzed in Section 6.3.

The government evaluates the consequences of the reform by aggregating agents' welfare using a Utilitarian social welfare function that puts an equal weight on all agents who are alive at the time of the reform. Importantly, the government takes into account the effect of the tax reform on people's welfare over the transition. The optimal tax problem then is to find the tax rate τ on capital income that leads to the competitive equilibrium that achieves the highest social welfare. Formally, the government solves the following problem:

$$\max_{\tau} \sum_{i=u,s} \pi_{i,0} \int_{\mathcal{Z}_i \times \mathcal{A}} v_{i,0}(z_i, a_i; \tau) d\Lambda_{i,0}(z_i, a_i) \tag{4}$$

such that, for every τ, $v_{i,0}(z_i, a_i; \tau)$ is the value in the corresponding competitive equilibrium.

This baseline social welfare function: (i) puts a uniform weight on all agents and (ii) does not take into account the welfare of future generations. In Section, 5.2 we conduct optimal tax exercises and analyze the impact of capital-skill complementarity on optimal capital taxes under different social welfare functions that: (i) put all weight on the most unfortunate member of society, (ii) ignore redistribution, and (iii) take into account future generations' welfare.

4 Calibration

This section first explains how we calibrate the baseline model with capital-skill complementarity to the U.S. economy. We first fix a number of parameters to values from the data or from the literature. These parameters are summarized in Table 1. We then calibrate the remaining parameters so that the stationary recursive competitive equilibrium of the model economy matches the U.S. economy around 2017 along selected dimensions that are key for our investigation.[7] Our calibration procedure is summarized in Table 3. Whenever data is not available until 2017 for some variable, we use the most recent data. The details and definitions of the data are included in Appendix B.

Preferences and Demographics. One period in the model corresponds to one year. We assume that the period utility function takes the form

$$u(c) - v(l) = \frac{c^{1-\sigma}}{1-\sigma} - \phi \frac{l^{1+\gamma}}{1+\gamma},$$

[7]The existence of stationary equilibrium requires the assumption that policies (government expenditure, debt and taxes) do not change over time. Given this, a stationary recursive competitive equilibrium is a recursive competitive equilibrium defined exactly as in Section 2 in which the firm allocation, consumer value and policy functions, skill choices, prices and distributions over individual states are all independent of time. We choose 2017 U.S. economy as the calibration target because we want to focus on the economy before the capital tax reform of President Trump's administration entered into effect on January 1, 2018.

where σ equals the coefficient of relative risk aversion while γ controls the Frisch elasticity of labor supply. In the benchmark case, we use $\sigma = 1$ and $\gamma = 1$. These are within the range of values that have been considered in the literature. We calibrate ϕ to match the average labor supply. Agents in the model are born at the real life age of 25 and enter the labor market immediately. Following Castaneda, Díaz-Giménez, and Ríos-Rull (2003), the survival probability χ is set to 0.978 to match the average working life-span of 45 years. The discount rate, β, is calibrated internally as explained below.

The fraction of skilled agents is calculated to be 0.3544 using the Current Population Survey (CPS) data for 2017. We focus on males who are 25 years old or older and who have earnings. To be consistent with Krusell, Ohanian, Ríos-Rull, and Violante (2000), skilled people are defined as those who have at least a bachelor's degree. We set the fraction of skilled workers in the model to 0.3544 exogeneously and not specify a cost distribution yet since this is not needed to compute the status-quo stationary economy. The cost distribution, which is needed for the optimal tax analysis, is calibrated in Section 4.3 to be consistent with this number in equilibrium with endogenous skill choice.

Technology. In the baseline economy with capital-skill complementarity, the production function takes the same form as in Krusell, Ohanian, Ríos-Rull, and Violante (2000):

$$Y = F(K_s, K_e, L_s, L_u) = K_s^{\alpha} \left(\nu \left[\omega K_e^{\rho} + (1-\omega) L_s^{\rho} \right]^{\frac{\eta}{\rho}} + (1-\nu) L_u^{\eta} \right)^{\frac{1-\alpha}{\eta}}. \tag{5}$$

In this formula, ρ controls the degree of complementarity between equipment capital and skilled labor while η controls the degree of complementarity between equipment capital and unskilled labor. Krusell, Ohanian, Ríos-Rull, and Violante (2000) estimate ρ and η, and we use their estimates. Their estimates of these two parameters imply that equipment capital is more complementary with skilled than unskilled labor. The parameter α gives the income share of structure capital. The other two parameters in this production function, ω and ν jointly control the income shares of equipment capital, skilled labor and unskilled labor. These three parameters are calibrated internally, as explained in detail later.

Table 1: Benchmark Parameters

Parameter	**Symbol**	**Value**	**Source**
Technology (Capital-skill complementarity)			
Structure capital depreciation rate	δ_s	0.056	GHK
Equipment capital depreciation rate	δ_e	0.124	GHK
Elasticity of substitution between K_e and L_u	η	0.401	KORV
Elasticity of substitution between K_e and L_s	ρ	-0.495	KORV
Technology (Cobb-Douglas)			
Capital's share of output	θ	1/3	
Elasticity of substitution between L_s and L_u	ε	0.2908	KM
Depreciation rate of capital	δ	0.0787	
Common parameters			
Relative risk aversion parameter	σ	1	
Inverse Frisch elasticity	γ	1	
Survival probability	χ	0.978	CDR
Relative supply of skilled workers	π_s	0.3544	CPS
Labor tax progressivity	τ_l	0.1	FN
Linear tax rate on capital income	τ	0.36	TU
Government consumption	G/Y	0.16	NIPA
Government debt	D/Y	0.60	FRED

This table reports the benchmark parameters that we take directly from the literature or the data. The acronyms CDR, FN, GHK, KORV, KL, KM and TU stand for Castaneda, Díaz-Giménez, and Ríos-Rull (2003), Ferrière and Navarro (2018), Greenwood, Hercowitz, and Krusell (1997), Krusell, Ohanian, Ríos-Rull, and Violante (2000), Heathcote, Storesletten, and Violante (2017), Krueger and Ludwig (2016), Katz and Murphy (1992) and Trabandt and Uhlig (2011), respectively. NIPA stands for the National Income and Product Accounts, CPS for Current Population Survey and FRED for the FRED database of the Federal Reserve Bank of St. Louis.

Government. As reported in the National Income and Product Accounts (NIPA), the government consumption-to-output ratio has been fairly stable with an average of about 16% since the 1980's. This is the value we use. We assume a government debt of 60% of GDP, which equals the federal debt held by private investors over GDP in 2015 according to the Federal Reserve Bank of Saint Louis FRED database.

We follow Trabandt and Uhlig (2011) and assume that the current tax rate on capital income is $\tau = 36\%$. As for labor income taxes, modeling the progressivity of the U.S. tax system may be important for our exercise since progressive tax systems can already provide substantial redistribution from skilled to unskilled workers, dwarfing the importance of taxing capital for indirect redistribution. Using longitudinal IRS (Internal Revenue Service) data, Ferrière and Navarro (2018) provide annual estimates of τ_l until 2012.[8] They find that τ_l is about 0.1 during 2010-2012. This is consistent with Dyrda and Pugsley (2019) who estimate a progressivity parameter of slightly below 0.1 for the same period. We use this estimate and calibrate λ, which controls the average labor tax in the economy, to clear the government budget.

Wage Risk. It is well known that the class of models used in this paper together with Gaussian individual labor productivity shocks falls short of matching earnings and wealth inequality simultaneously, especially at the top end of the corresponding distributions. One way to resolve this issue, proposed by Castaneda, Díaz-Giménez, and Ríos-Rull (2003), which we follow, is to assume the existence of a superstar individual productivity state. Specifically, for each worker type i, productivity, z can be either in a normal or a superstar state. In the normal state, z follows a skill-type specific AR(1) process: $\log z_{t+1} = \rho_i \log z_t + \varepsilon_{i,t}$, which we approximate by finite number Markov chains using the Rouwenhorst method described in Kopecky and Suen (2010). At any given time, from any normal state, productivity transits to superstar state with probability p_i. When at the superstar state, productivity is e_i times larger than the average productivity across normal

[8] We do not use the estimate provided in Heathcote, Storesletten, and Violante (2017) because the income base that the tax function applies to is labor plus capital income in their paper, whereas in our paper the tax function applies to labor income only. This is also the approach taken by Ferrière and Navarro (2018).

states. The probability of remaining at the superstar state is q_i. When agents return to the normal state, they draw a new labor market ability from the ergodic distribution associated with the AR(1) process. Together with the persistence parameter, ρ_i, and variance of the shocks, $var(\varepsilon_i)$, the productivity process introduces ten parameters to be calibrated.[9]

Internal Calibration. In addition to the ten parameters related to the productivity processes, there are six parameters to be determined: the three production function parameters, α, ω and ν, the labor disutility parameter ϕ, the discount factor β, and the parameter governing the overall level of taxes in the tax function, λ. These 16 parameter values are jointly chosen to ensure that the model matches the data along a number of selected moments. Although the calibration is carried out jointly, it is instructive to think about the calibration of the two sets of parameters separately. First, the ten parameters that describe the individual wage risk processes are calibrated to match ten distributional targets. These targets are the six Gini coefficients of the overall, skilled and unskilled earnings and wealth distributions, top 1%'s share in the earnings and wealth distribution, the ratio of average wealth of skilled workers to that of unskilled workers, and the autocorrelation of earnings. Table 2 reports the model's ability to match calibration targets in Panel A and the calibrated parameter values in Panel B.

Second, the remaining six parameters are chosen to match six aggregate moments. The income shares of equipment capital, skilled labor and unskilled labor are governed by ω and ν, and α governs the income share of structure capital. We calibrate α, ω and ν so that (i) the share of equipment capital in total capital is 1/3 as it is approximately in the Fixed Asset Tables (FAT) in 2017, (ii) the labor share equals 2/3, and (iii) the skill premium equals 1.9 as reported by Heathcote, Perri, and Violante (2010).[10] We choose ϕ

[9]We cannot identify the mean levels of the idiosyncratic labor productivity shock z for the two types of agents separately from the remaining parameters of the production function and, therefore, normalize average productivity of each skill type to 1. This assumption implies that the marginal product of labor for type i, w_i, equals the average wage rate of workers of that skill type. As a result, the skill premium in the model economy is given by w_s/w_u. This is in line with the benchmark estimation of Krusell, Ohanian, Ríos-Rull, and Violante (2000) who abstract from time variation in average productivity differentials across skill types.

[10]Heathcote, Perri, and Violante (2010) use CPS data and compute the skill premium for the period 1967-2005 for males between ages of 25 and 60, working at least 260 hours a year. In subsequent work,

Table 2: Calibration: Distributional Moments

Panel A: Moments	Data	Model
Earnings Gini	0.68	0.66
Earnings Gini - skilled	0.66	0.66
Earnings Gini - unskilled	0.61	0.62
Earnings Top 1%'s share	0.23	0.24
Earnings autocorrelation	0.94	0.95
Wealth Gini	0.86	0.85
Wealth Gini - skilled	0.81	0.81
Wealth Gini - unskilled	0.82	0.82
Wealth Top 1%'s share	0.39	0.38
Relative skilled wealth	5.6	5.6
Panel B: Parameters	**Symbol**	**Value**
Normal state persistence (skilled)	ρ_s	0.8219
Normal state volatility of shocks (skilled)	$var(\varepsilon_s)$	0.1338
Transit into superstar state (skilled)	p_s	1×10^{-3}
Remain in superstar state (skilled)	q_s	0.9473
Productivity superstar state (skilled)	e_s	35.57
Normal state persistence (unskilled)	ρ_u	0.9915
Normal state volatility of shocks (unskilled)	$var(\varepsilon_u)$	0.0333
Transit into superstar state (unskilled)	p_u	8×10^{-5}
Remain in superstar state (unskilled)	q_u	0.0216
Productivity superstar state (unskilled)	e_u	43.43

This table reports calibration results regarding the wage risk parameters. The model's ability to match calibration targets are reported in Panel A and the calibrated parameter values are reported in Panel B. All data moments correspond to 2016 U.S. economy and are taken from Kuhn and Ríos-Rull (2020), with the exception of the autocorrelation of earnings, which is reported in Boar and Midrigan (2022). Relative skilled wealth refers to the ratio of the average skilled asset holdings to the average unskilled asset holdings.

so that the aggregate labor supply in steady state equals 1/3 as commonly assumed in the macro literature. We calibrate β so that the capital-to-output ratio in the model equals 2.07. This number is calculated using the NIPA and Fixed Asset Tables for year 2017. Krusell, Ohanian, Ríos-Rull, and Violante (2000) exclude housing from both capital stock and output time series when they estimate the parameters of the production function. Since we use their estimates, we also exclude housing from both capital stock and output when we calculate the capital-to-output ratio. Following Heathcote, Storesletten, and Violante (2017), we choose λ to clear the government budget constraint in equilibrium. Table 3 summarizes the internal calibration procedure. Data targets are not reported in the table as the model is able to match the targets precisely.

they update skill premium data series until 2016. They find that the skill premium has been stable around 1.9 during 2005-2016 period.

Table 3: Calibration: Aggregate Moments

Parameter	**Symbol**	**Value**	**Target**	**Source**
Technology (CSC)				
Production parameter	ω	0.2824	Labor share = 2/3	NIPA
Production parameter	ν	0.6581	Skill premium = 1.9	CPS
Production parameter	α	0.1909	Share of equipments, $\frac{K_e}{K} = 1/3$	FAT
Technology (CD)				
Total factor productivity	A	0.7870	Output level of CSC economy	
Production parameter	κ	0.5581	Skill premium = 1.9	CPS
Common parameters				
Discount factor	β	0.9365	Capital to output ratio = 2.07	NIPA, FAT
Tax function parameter	λ	0.8844	Government budget balance	
Disutility of labor	ϕ	6.90	Labor supply = 1/3	

This table reports the calibration procedure for parameters that target aggregate moments. Model generated target moments are not reported as the match is perfect. The production function parameters α, ν and ω control the income shares of structure capital, equipment capital, skilled and unskilled labor in the capital-skill complementarity model (CSC). The production function parameter κ controls the income shares of the skilled and unskilled labor in the Cobb-Douglas model (CD). The tax function parameter λ controls the labor income tax rate of the mean income agent. The acronyms CPS, FAT, and NIPA stand Current Population Survey, Fixed Asset Tables, and National Income and Product Accounts, respectively.

4.1 Calibration of the Cobb-Douglas Economy

In the second economy, we eliminate capital-skill complementarity, and use the following production function:

$$Y = AK^{\theta}(\kappa L_s^{\varepsilon} + (1-\kappa)L_u^{\varepsilon})^{\frac{1-\theta}{\varepsilon}}$$

where A is total factor productivity, θ is the usual Cobb-Douglas parameter that governs the income share of capital, κ is a share parameter that allows for skilled labor to be more effective than unskilled labor, and ε controls the degree of substitutability between skilled and unskilled labor. We set $\theta = 1/3$ as is common in the literature. This is also in line with the labor share target of the capital-skill complementarity economy. Following Katz and Murphy (1992), we set the elasticity of substitution between skilled and unskilled labor to 1.41, which implies $\varepsilon = 0.2908$. The depreciation rate of capital, δ, is assumed to equal the weighted average of depreciation rates of structure capital and that of equipment capital in the capital-skill complementarity economy. These exogenously calibrated technology parameters for the Cobb-Douglas economy are summarized in Table 1. The rest of the externally calibrated parameters in the Cobb-Douglas economy are chosen identically to the complementarity economy and are also summarized in the same table. Similarly, the same income process is used in the Cobb-Douglas economy as in the complementarity economy.

The rest of internal calibration procedure in the Cobb-Douglas economy is identical to that in the capital-skill complementarity economy except that there are only five parameter values left to be determined. The first parameter is the total factor productivity parameter, A, which is calibrated so that the Cobb-Douglas economy has the same total output as the capital-skill complementarity economy in the status-quo steady state. The calibrated value for A is reported in Table 3. The second parameter is κ, which is chosen to ensure that the skill premium equals 1.9. The remaining three parameters are the labor disutility parameter ϕ, the discount factor β, and the parameter governing the overall level of taxes in the tax function, λ. We calibrate these parameters to match the exact same targets as in the complementarity economy. As a result, the calibrated parameter values for these three are identical to those in the complementarity economy, and are given in the last four rows of Table 3.

It is worth emphasizing that the calibration procedures render the two economies completely identical. That is, the real interest rate, the skilled and unskilled wages, aggregate output, aggregate capital stock, aggregate labor and consumption, as well as the distributions of consumption, labor supply, assets, earnings and welfare across workers are identical in the initial steady states of the two economies. This synchronization of the capital-skill complementarity and Cobb-Douglas economies is important as it assures us that the difference in the optimal tax rates across the two economies cannot be coming from differences in initial conditions. The difference in optimal tax rates emerges from the fact that the two economies respond differently to identical tax reforms.

4.2 Model Fit

In this section, we provide a further validation of our calibration by comparing the calibrated model to the data along a number of non-targeted moments.

Table 4: Non-Targeted Moments

	A: Earnings Quintiles				
	1st	**2nd**	**3rd**	**4th**	**5th**
Data	0	2.9%	9.8%	19.1%	68.3%
Model	1.7%	3.3%	10.6%	14.9%	69.5%
	B: Wealth Quintiles				
	1st	**2nd**	**3rd**	**4th**	**5th**
Data	-0.5%	0.6%	2.9%	8.6%	88.3%
Model	0%	0.1%	2.7%	9.1%	88.0%

This table reports the fit of the model with respect to some non-targetted moments of the earnings and wealth distributions. All data moments correspond to 2016 U.S. economy and are taken from Kuhn and Ríos-Rull (2020).

Cross Sectional Moments. Table 4 summarizes the performance of the model vis-a-vis the data in terms of cross-sectional earnings and wealth moments that are not targeted in our calibration. Panel A reports the earnings share of each earnings quintile both in the model and in the data whereas Panel B does the same for wealth. We find that our model reproduces well the degree of inequality in earnings and wealth that is present in the U.S. economy. We also investigate how our model performs regarding the distribution of hours worked by comparing a number of key moments to their empirical counterparts calculated by Heathcote, Perri, and Violante (2010) and their subsequent work (updated to year 2016). The variance of log hours delivered by the model is 0.13 which compares well to its empirical counterpart of 0.12. The Gini coefficient of hours worked in the model is 0.1, which falls moderately short of 0.14 in the data. Hong, Seok, and You (2019) estimates hours volatility by skill type and find that the coefficient of variation of skilled hours is 0.2 and 0.22 among skilled and unskilled (their data refers to year 2000). The model delivers 0.11 and 0.33.

Long-Term Changes in Macroeconomic Variables. Krusell, Ohanian, Ríos-Rull, and Violante (2000) argue that, under the assumption of capital-skill complementarity, the declining price of equipment and the changes in the relative supply of skilled workers explain most of the changes in the skill premium between 1960's and 1990's. In this section, we investigate how well our calibrated model performs in terms of matching long-run changes in the skill premium and some other macroeconomic variables. Specifically,

Table 5: Non-Targeted Moments: Macroeconomic Variables

	Data			Model		
	1967	**2017**	**Change**	**1967**	**2017**	**Change**
Skill premium	1.50	1.90	27%	1.48	1.90	28%
Share of equipment	0.36	0.32	-12%	0.39	0.33	-15%
Labor share	0.66	0.61	-9%	0.63	0.66	5%
Real output			137%			147%
Investment-to-output ratio	0.21	0.20	-6%	0.17	0.16	-4%

This table reports the performance of the calibrated model in terms of matching long-run changes in the skill premium, share of equipment in total capital stock, labor share, real output, and investment-to-capital ratio. For the details on data construction, see Appendix C.

we take the model economy that is calibrated to the 2017 U.S. economy and feed in the price of equipment, the relative supply of skilled workers and government policies in the 1967 U.S. economy, and solve for the stationary equilibrium of the model that corresponds to 1967. Appendix C provides a further description of the steady state that corresponds to the 1967 economy.

We then compare the model generated changes in the skill premium and other macroeconomic moments between 1967 and 2017 with their empirical counterparts. Table 5 summarizes our results. We find that our model matches quite well the long-run change in the skill premium in response to changes in equipment price and relative supply of skilled workers.[11] During the same time period, the share of equipment in total capital stock decreased from 0.36 to 0.32, or by 12%. The corresponding decline in the share of equipment implied by the model is similar at 15%. The model fails to capture the decline of the labor share in the last few decades which has been argued by a recent literature; see, for instance, Karabarbounis and Neiman (2014).[12] The model matches well the growth in real output per capita as well as the decline in the share of output that is used for investment.

[11] Slavík and Yazici (2022) report a similar finding using an open economy incomplete markets model.

[12] The fact that the production function we use does not capture the recent decline in the labor share is known from Krusell, Ohanian, Ríos-Rull, and Violante (2000) and the literature that follows. For example, Ohanian, Orak, and Shen (2021), who find strong evidence for continued capital-skill complementarity and that the Krusell, Ohanian, Ríos-Rull, and Violante (2000) model continues to closely account for the skill premium in the most recent data, also report that the model overpredicts the level of the labor share by about four percentage points throughout most of the 2010s.

4.3 Calibration of the Cost of Skill Acquisition

Following the calibration strategy in Heathcote, Storesletten, and Violante (2010), we assume a log-normal distribution of the cost of skill acquisition, H, and pin down the two unknown parameters of this distribution, its mean and the variance, to ensure that the 1967 and the 2017 steady states of the model economy that were described in the previous section match the fraction of skilled workers in the U.S. economy. Further details of this calibration are provided in Appendix D.

The calibration of the cost distribution H is important as it controls the elasticity of the fraction of skilled workers with respect to the skill premium, which itself is important for the quantitative strength of our mechanism. A number of papers estimate the elasticity of college enrolment with respect to the skill premium for the U.S. economy by estimating the following relationship using time series data: $log(enr_t) = a + b \cdot log(sp_t)$. The coefficient b gives the percentage change in enrolment that is associated with a one-percent change in skill premium and is interpreted as the elasticity of enrolment with respect to the skill premium. The estimates of this elasticity vary considerably across studies depending on the exact definition of variables used and the time periods taken into account, but fall mostly within the range of 1 to 2 as reported in the meta analysis by Freeman (1982). Estimating the same relationship using model simulated data over the transition following various capital tax reforms (including the optimal one), we find that the elasticity of college enrolment implied by our model is quite stable and within this range of empirical estimates provided by the literature.

5 Optimal Capital Taxation

This section describes the key features of optimal capital tax reforms for economies with and without capital-skill complementarity. After providing baseline results - for the economies calibrated in Section 4 and under Utilitarian social welfare function, we check how our results are affected by alternative calibrations and social welfare functions.

Table 6: Optimal Taxes: Baseline Results

	τ	λ
Status Quo	0.36	0.89
Cobb-Douglas	0.61	0.95
Capital-Skill Complementarity	0.67	0.96

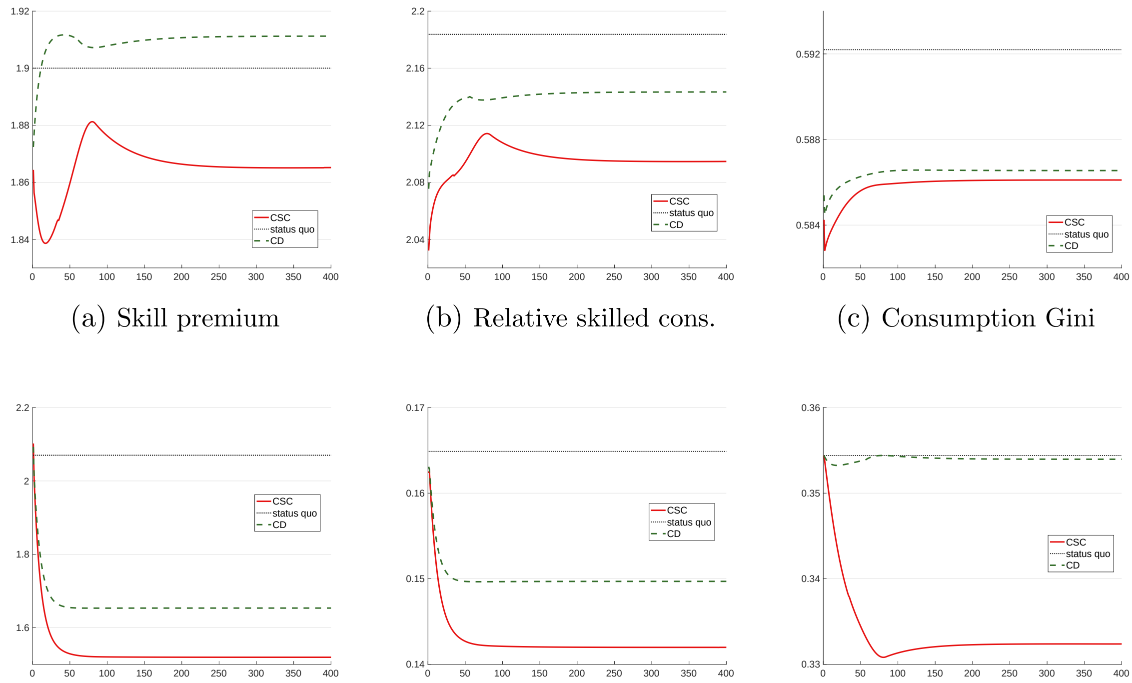
The first row of the table reports status-quo capital tax rate used in our calibration and the corresponding average labor income tax parameter in the corresponding steady state. The second and third rows report the optimal capital tax rate and the labor income tax parameter in the resulting final steady state for both the Cobb-Douglas and capital-skill complementarity models.

5.1 Baseline Results

The first row of Table 6 reports the status-quo capital tax rate used in our calibration and the corresponding steady-state average labor income taxes, controlled by $1-\lambda$. The second and third rows report the values of corresponding variables under the optimal reforms for the Cobb-Douglas and capital-skill complementarity economies. The main finding is that the optimal capital tax rate in the capital-skill complementarity economy is significantly larger than that in the Cobb-Douglas economy, 67% vs. 61%. Accordingly, optimal average labor taxes in the steady state are relatively lower in the capital-skill complementarity economy.

In the Cobb-Douglas economy, increasing the tax rate on capital income has the benefit of decreasing consumption inequality since capital income is more unevenly distributed across the population than labor income. However, taxing capital also entails the usual cost of discouraging its accumulation, and hence, depressing output. That the optimal capital tax rate is positive and large, 61% in our calculation, arises mainly from this trade off. Similar large capital tax rates have been found to be optimal previously in the literature, for instance, by Dyrda and Pedroni (2022).

What is more interesting is the finding that under capital-skill complementarity, the capital tax rate should be set significantly, namely 6 percentage points, higher. The reason for this difference is that, in the capital-skill complementarity economy, besides the trade off explained above, increasing capital taxes has an additional redistributive benefit. Higher capital taxes slow down aggregate capital accumulation, and in particular the accumulation of equipment capital. When there is capital-skill complementarity, this



(a) Skill premium (b) Relative skilled cons. (c) Consumption Gini

(d) Capital-output ratio (e) Aggregate output (f) Relative skill supply

Figure 1: Dynamics of key macroeconomic variables following the optimal reform

The six graphs report how the skill premium, ratio of average consumption of the skilled workers to that of unskilled workers, Gini coefficient of the consumption distribution, capital-output ratio, aggregate output and fraction of skilled workers change over the transition following the optimal tax reform. CSC and CD refer to capital-skill complementarity and Cobb-Douglas economies, respectively.

decreases the relative demand for skilled labor, which then diminishes the skill premium. As a result, increasing capital taxes provides indirect redistribution from skilled to unskilled agents. To the extent that unskilled agents are poorer, they have higher marginal utility from consumption, and hence, this redistribution increases social welfare from the perspective of a Utilitarian planner. Observe that this indirect redistribution channel is partly mitigated by the fact that the decline in the skill premium coming from higher capital taxes discourages skill acquisition, preventing the skill premium from declining further.

The indirect redistribution channel at work under capital-skill complementarity can be observed from Figure 1a which shows that the reform reduces the skill premium considerably, from 1.90 to as low as 1.84 over the transition to a final steady-state level of just above 1.86. Rising capital taxes have virtually no effect on the skill premium in the Cobb-Douglas case. Table 7 provides further details of how changes in capital

Table 7: Skill Premium Decomposition

	Pre-reform	**+Capital**	**+Extensive**	**+Intensive**
Cobb-Douglas	1.90	1.90	1.90	1.91
Capital-skill complementarity	1.90	1.73	1.86	1.86

This table decomposes the change in skill premium in response to optimal capital tax reforms across steady states for Cobb-Douglas (CD) and capital-skill complementarity (CSC) economies into components coming from changes in the capital stock and the extensive and intensive margins of labor supplies of both skill types.

and labor allocations affect the skill premium. The first and fourth columns in the table are the pre-reform and the post-reform steady state values of the skill premium. The second column computes the skill premium for an artificial allocation which fixes the values of aggregate skilled and unskilled effective hours, L_s and L_u, to pre-reform steady state levels while setting the stock of capital (of both types in the CSC economy) to the post-reform level, thereby isolating the capital channel. Relative to the second column, the third column changes only the fraction of skilled workers to the post-reform level, holding the average skilled and unskilled labor hours fixed, and computes the implied skill premium. This way, the third column isolates the extensive margin effect. Finally, a comparison of the third and fourth columns gives the intensive margin effect on the skill premium (by adjusting the skilled and unskilled hours to the post reform steady-state values). As expected, in the CSC economy, the decline in capital stock has the strongest effect on the skill premium which is partially offset by the resulting decline in the fraction of skilled workers. The capital channel is not operational and the extensive margin channel is negligible in the CD economy. The intensive margin effects are small in both cases.

The redistributive benefit of the decline of the skill premium in the CSC economy can be seen from Figure 1b: average consumption inequality between the two groups falls over the transition. The decline in consumption inequality in the Cobb-Douglas economy in response to increasing capital taxes is significantly less pronounced. A similar pattern can be observed looking at Figure 1c: consumption Gini decreases more in the CSC economy.

Higher capital taxes have similar aggregate implications in the two economies: they reduce capital intensity and output. This happens to a larger extent in the CSC economy as displayed by Figure 1d and Figure 1e because the capital tax increase is larger in the

CSC economy. As Figure 1f shows, an important difference between the two economies is the decline in the fraction of skilled workers observed only in the CSC economy, which is caused by the decline in the skill premium.

Welfare Gains. The welfare gains of the reform are equivalent to increasing the consumption of all agents (who were alive at the time of the reform) at all dates and states by 1.25% in the economy with capital-skill complementarity, while the corresponding welfare gains number is 0.85% in the standard Cobb-Douglas economy. This implies that carrying out the optimal capital tax reform is more important in terms of its welfare effects when capital-skill complementarity in production is taken into account.

Components of Welfare Gains. Following Benabou (2002) and Floden (2001), we decompose the total welfare gains, Δ, into three components: level, Δ_L, redistribution, Δ_R, and insurance, Δ_I, where $1+\Delta = (1+\Delta_I)(1+\Delta_R)(1+\Delta_L)$.[13] The level component measures the welfare gains that arise from improvements in aggregate quantities between the pre-reform and post-reform allocations. It aims to capture efficiency gains that result from better allocation of productive resources and reduction in distortionary taxes. The redistribution component measures the gains that arise from a reduction in inequality between the two allocations. Finally, the insurance component measures the welfare gains that arise from a reduction in risk associated with pre-reform vs. post-reform allocations, and aims to capture the magnitude of insurance that the tax reform provides.

The first row of Table 8 shows that the reform brings substantial redistributive gains at the cost of large level losses and a modest increase in the risk faced by individuals in the CD economy. Recall that the reform increases the capital tax rate and lowers labor taxes. The rise in capital tax rate generates redistribution as wealth is distributed very unevenly. The decline in average labor taxes goes in the opposite direction but is clearly trumped by the capital tax rate. This is in line with the findings of Dyrda and

[13]Our decomposition is more closely related to, but distinct from, Dyrda and Pedroni (2022), who extend the methods developed by prior literature to measure welfare gains over transitions. Our and their methods produce identical level effects and similar but non-negligibly different redistribution and insurance effects. With our method, one does not need to define certainty equivalent allocations over transition. The precise definitions used in our decomposition are provided in Appendix E. There, we also formally prove the claim $1 + \Delta = (1 + \Delta_I)(1 + \Delta_R)(1 + \Delta_L)$.

Table 8: Decomposition of Welfare Gains in Baseline Reform

	$\boldsymbol{\Delta}$	$\boldsymbol{\Delta_L}$	$\boldsymbol{\Delta_R}$	$\boldsymbol{\Delta_I}$
Cobb-Douglas	0.85	-1.45	2.44	-0.11
Skilled	-1.09	-3.68	2.81	-0.12
Unskilled	1.93	1.12	0.90	-0.10
Capital-Skill Complementarity	1.25	-2.12	3.19	0.24
Skilled	-1.95	-5.46	3.15	0.55
Unskilled	3.04	1.86	1.09	0.07

The first panel of the table reports the total welfare gains of the reform (Δ) and its decomposition to level (Δ_L), redistribution (Δ_R) and insurance (Δ_I) effects in consumption equivalence units for the Cobb-Douglas economy while the second panel reports the same for the capital-skill complementarity economy.

Pedroni (2022) who argue that, in an economy that is similar to ours, changes to capital income taxes are more important for redistributive gains than changes to labor taxes. Large distortions created by higher capital taxes lower productivity (through mainly lower capital intensity in production), which manifests itself in negative level effects.[14] Lower labor taxes increase the share of total household after-tax income that is risky. This is a force for negative insurance effect of the reform. However, since it increases aggregate labor and reduces the capital stock, the reform also decreases wages, which has a positive insurance effect. This counteracting general equilibrium force partially offsets the former, resulting in a small and negative insurance effect.

A comparison of the first and the fourth rows reveals that, by pushing the capital tax rate even higher in the capital-skill complementarity economy, the government achieves more redistribution but this comes at the cost of larger level losses. The insurance effect changes sign relative to the Cobb-Douglas economy, albeit it is still quite small relative to other components.

Figure 2 depicts a more complete picture by comparing how identical tax reforms affect welfare via the three components in the two economies. While rising capital taxes have comparable level effects in CSC and CD economies, they generate larger redistributive

[14] From Dyrda and Pedroni (2022), we know that increasing the capital tax rate can work in the direction of increasing labor productivity if there are wealth effects on labor supply. Lowering labor taxes can also increase labor productivity, especially since our labor tax schedule is progressive. These positive effects on productivity are, however, trumped by the aforementioned distortionary effects of the reform, which, therefore, generates overall negative level effects. This is, in part, due to the fact that, as shown by Dyrda and Pedroni (2022), time variation of fiscal instruments is important for the productivity enhancing effect of rising capital taxes.

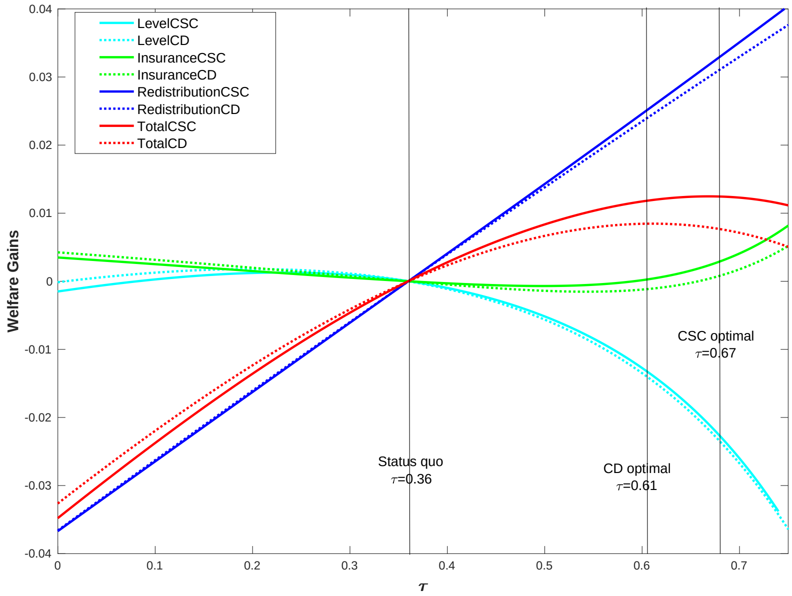


Figure 2: Components of Welfare Gains

This figure displays the three components of welfare gains, level, insurance, and redistribution, as well as total welfare gains from capital tax reforms for Cobb-Douglas and capital-skill complementarity economies. From left to right, the solid black lines represent the status-quo, optimal Cobb-Douglas and optimal capital-skill complementarity capital tax rates.

gains in the former, which is expected given that the indirect redistribution channel present only in that economy. The insurance effects of increasing capital taxes are smaller in magnitude relative to other components. They are non-monotonic in general, but positive and increasing at higher tax rates, and are larger in the CSC economy. This is partly due to the fact that, for a given capital tax rate increase, averages wages for both skilled and unskilled workers decline more in the CSC economy.

Distribution of Welfare Gains. We also find that the distribution of welfare gains is more tilted toward the unskilled in the CSC model relative to the CD model. As the second and third rows of Table 8 display, the welfare of unskilled agents as a group increases by 1.93% in consumption equivalence units in the CD economy while the skilled agents' welfare decreases by 1.09%. The corresponding numbers are more extreme in the CSC economy: a 3.04% increase for unskilled and a 1.95% decrease for skilled. Looking at the decomposition of these gains, perhaps most notable is that the skilled agents face much larger level losses relative to the aggregate level losses while unskilled agents experience level gains in both economies. This is because, in addition to the aggregate level loss coming from increased distortions in the economy, the reform also taxes away their wealth to be redistributed to unskilled workers. Importantly, the level losses of skilled and gains of the unskilled are more pronounced in the CSC economy because of the indirect redistribution channel that is also at work.

Not reported in Table 8 is a closer look at the winners and the losers of the reform within each skill group. It is the asset-poor agents who gain and the asset-rich agents who lose in both groups in both economies. In the CSC economy, while 88% of the unskilled gain, only 49% of the skilled do so. Since the indirect redistribution channel is missing in the CD economy, the welfare implications are more symmetric within the two groups than in the capital-skill complementarity case: 87% of the unskilled and 54% of the skilled gain.

5.2 Alternative Social Welfare Criteria

The baseline analysis assumes that the government evaluates the outcome of the reform by aggregating citizens' welfare using a Utilitarian social welfare function that puts an equal weight on all agents who were alive at the time of the reform. In this section, we consider alternative assumptions regarding how society adds up individual utilities.

Rawlsian Social Welfare. A substantially more redistributive alternative is the Rawlsian social welfare criterion. This social welfare function maximizes the welfare of the least fortunate member of society. The optimal tax problem then is to find the tax rate τ on capital income that leads to the competitive equilibrium that achieves the highest welfare for the agent with the lowest welfare among all the agents who were alive at the time of the reform. Formally, the government solves the following problem:

$$\max_{\tau} \min_{i \in u,s; (z_i,a_i) \in \mathcal{Z}_i \times \mathcal{A}} v_{i,0}(z_i, a_i; \tau) \tag{6}$$

such that, for every τ, $v_{i,0}(z_i, a_i; \tau)$ is the value in the corresponding competitive equilibrium.

Table 9: Optimal Taxes under Alternative Social Welfare Functions

	τ	λ
Status Quo	0.36	0.89
Baseline		
Cobb-Douglas	0.61	0.95
Capital-Skill Complementarity	0.67	0.96
Rawlsian Social Welfare		
Cobb-Douglas	0.70	0.97
Capital-Skill Complementarity	0.74	0.97
Ignoring Redistribution		
Cobb-Douglas	0.08	0.821
Capital-Skill Complementarity	0.14	0.837
Weight on Unborn Generations		
Cobb-Douglas	0.53	0.93
Capital-Skill Complementarity	0.56	0.93

The first row of the table reports status-quo capital tax rate used in our calibration and the corresponding average labor income tax parameter in the corresponding steady state. The second and third rows report the optimal capital tax rate and the labor income tax parameter in the resulting final steady state for both the Cobb-Douglas and capital-skill complementarity models for the baseline exercise. The fourth and fifth rows report the values of the same variables for the exercise in which the social welfare function is Rawlsian while the sixth and the seventh rows report these for the case in which social welfare function consists of the multiplication of the level and insurance components and ignores redistribution component. Finally, last two rows report optimal capital tax rate and the corresponding final steady-state labor income tax parameter for a social welfare function that weighs utility of generations who were yet to be born at the reform date.

The results of this exercise are reported in the second panel of Table 9. We find that the optimal capital tax rate is 74% in the economy with capital-skill complementarity while it is 70% in the economy without. Since redistributive considerations are more important under the Rawlsian social welfare criterion, the government uses capital taxes more heavily in both economies in order to tax asset-rich agents. The difference between the optimal rates in the CSC and CD economies is four percentage points, somewhat lower than the differential in the benchmark case.[15]

[15] One explanation as to why the differential falls is as follows. Changing the social welfare function from Utilitarian, which puts some weight on skilled workers' welfare, to Rawlsian, which puts all weight on unskilled workers (since lowest welfare agent is unskilled), one may expect that our channel - which redistributes from skilled to unskilled as a group - becomes more pronounced, and hence, the optimal capital tax differential between CD and CSC economies should increase. However, moving from Utilitarian social welfare function to Rawlsian also implies that although we were putting weight on people with all asset levels before, now we put weight only on workers with the lowest asset level (namely workers with no assets). This means that the standard channel for redistributive capital taxation also becomes more pronounced. Whether the optimal tax differential increases depends on which of the two mechanisms' strength increases more, which is a quantitative matter, and in part, depends on the relative degrees of wealth inequality vs. skill premium.

Ignoring Redistribution. Next, we consider an optimal tax problem of a planner that does not value redistribution across initial types. We do so by making use of the decomposition of welfare gains introduced in Section 5.1. Specifically, we look for the capital tax rate that maximizes the combination of the efficiency and insurance gains of reform, which corresponds to $(1 + \Delta_L)(1 + \Delta_I)$. The results are given in the third panel of Table 9. First, compared to the baseline exercise, the optimal capital tax rates are much lower in both economies. This is expected as the main benefit of capital taxation is redistribution. Second, the optimal capital tax rate is still significantly higher in the CSC economy. A glance at Figure 2 shows that this is because the insurance and level gains associated with cutting capital taxes are lower in the CSC economy.

Weight on Unborn Generations. In the baseline exercise, we assume that the planner weighs only the welfare of generations who are alive at the time of the reform in the social welfare calculus. In this section, we investigate the impact of capital-skill complementarity on optimal capital taxation under the assumption that society also takes into account the utility of the generations who are yet to be born as of the time of the reform. For comparability to the baseline exercise, we assume that the planner cares about all citizens who were alive at the date of the reform equally. All citizens who are born on a certain date t after the reform enter the social calculus with a welfare weight of $\hat{\beta}^t = \beta^t$.

The social welfare is then given by:

$$\sum_{i=u,s} \pi_{i,0} \int_{\mathcal{Z}_i \times \mathcal{A}} v_{i,0}(z_i, a_i; \tau) d\Lambda_{i,0}(z_i, a_i) + (1 - \chi) \sum_{t=1}^{\infty} \hat{\beta}^t \sum_{i=u,s} \pi_{i,t} \int_{\mathcal{Z}_i} [v_{i,t}(z_i, 0) - \psi_i] d\Lambda_i^z(z_i),$$

where Λ_i^z is the stationary distribution associated with the Markov chain that describes the evolution of the productivity shock for type i. We believe that $\hat{\beta} = \beta$ is a natural case to consider since this amounts to assuming that the social discount factor that applies to utility from consumption in a given period only depends on the period and is independent of the cohort that enjoys that consumption.

The results of this exercise are reported in the bottom two rows of Table 9. First, in both CD and CSC economies, the optimal capital tax rates are lower relative to the benchmark exercise. This is expected since while the redistributive benefit of capital taxation is enjoyed more in the short run, the cost of capital taxation - the slowing down of capital accumulation - is a forward looking cost, and the social welfare function employed in this section puts more weight on the future relative to the baseline case. We also find that the impact of capital-skill complementarity on optimal taxes, as measured by the difference in the optimal capital tax rates between CSC and CD economies, is lower relative to the baseline exercise. The same intuition is at play here. The indirect redistribution channel that capital-skill complementarity unleashes brings more benefits in the short run while the additional distortion that this creates, namely the decline in the fraction of skilled workers, is a forward looking cost. Therefore, the additional capital tax that capital-skill complementarity implies is smaller when future generations are taken into account.

5.3 Sensitivity Analysis

Cost of Skill Acquisition. In the baseline economy, following Heathcote, Storesletten, and Violante (2010), we assume that the cost of skill acquisition is distributed according to a log-normal distribution. In this section, we recalibrate the distribution of cost of skill acquisition assuming it is distributed according to another commonly used two-parameter distribution, the logistic distribution (see, among others, Guerreiro, Rebelo, and Teles (2021)). The calibration of the model is identical up to the cost distribution, which is calibrated in a way identical to the calibration of the log-normal cost distribution in the baseline case. The results, which are summarized in Table 10 in the second panel from the top, show that the optimal capital taxes are five percentage points higher in the CSC economy (66%) relative to the CD economy (61%). The difference is 1 percentage point smaller than in the baseline.

We also consider an exercise with inelastic skill supply in which we keep the fraction of skilled workers exogenously at the pre-reform level. This is obviously an extreme assumption but this exercise is useful as it informs us about the significance of skill choice as a behavioral response to capital tax reforms. As the third panel of Table 10 displays,

Table 10: Optimal Taxes: Sensitivity Results

	τ	λ
Status quo	0.36	0.89
Baseline		
Cobb-Douglas	0.61	0.95
Capital-Skill Complementarity	0.67	0.96
Logistic Cost of Skill Acquisition		
Cobb-Douglas	0.61	0.95
Capital-Skill Complementarity	0.66	0.96
Exogenous Skills		
Cobb-Douglas	0.62	0.95
Capital-Skill Complementarity	0.73	0.99
Capital-Skill Complementarity		
Cobb-Douglas (baseline)	0.61	0.95
Capital-Skill Complementarity	0.67	0.96
Lower Labor Supply Elasticity		
Cobb-Douglas	0.60	0.94
Capital-Skill Complementarity	0.65	0.95

The first column of the table reports status-quo capital and average labor income taxes under the status-quo tax system. The second and third columns report optimal capital taxes and the steady-state value of average labor income taxes under the optimal tax system for both the Cobb-Douglas (CD) and capital-skill complementarity (CSC) models.

in this case optimal taxes are higher in both CD and CSC economies: 62% and 73%, respectively. This is intuitive: in both economies, higher capital taxes reduce people's incentives to acquire skills as skilled workers earn more and acquire more wealth. In fact, this is the only channel operating in the CD economy. In the CSC economy, in addition to this channel, higher capital taxes also reduce the skill premium, thereby disincentivizing skill acquisition further. The fact that ignoring skill choice increases optimal capital taxes much more in the CSC economy implies that this latter channel is quantitatively important.

Capital-Skill Complementarity. The mechanism that calls for higher optimal taxes on capital income works through the presence of capital-skill complementarity in production. In this regard, our results may be sensitive to the degree of relative substitutability of capital and skilled labor. Krusell, Ohanian, Ríos-Rull, and Violante (2000), whose elasticity estimates we employ in our quantitative work, use data from the period 1963-1992. However, the world in general and the U.S. economy in particular has been going through an unprecetended technological change in the last three decades. A recent working paper by Maliar, Maliar, and Tsener (2020) estimates the same production function, given by (5), using more recent data, namely data from the period 1963-2017. They find that in the recent data, the elasticity of substitution between equipment and unskilled labor is about 1.71, and the one between equipment and skilled labor of about 0.76. These values are higher than the corresponding numbers in Krusell, Ohanian, Ríos-Rull, and Violante (2000), which are 1.67 and 0.67, respectively. This implies that equipment capital has become more substitutable with both skilled and unskilled labor. Nonetheless, Maliar, Maliar, and Tsener (2020) conclude that the production function estimated by Krusell, Ohanian, Ríos-Rull, and Violante (2000) and the capital-skill complementarity mechanism remain remarkably successfull in explaining the skill premium dynamics.

To assess the sensitivity of our result to these elasticities, we now compute the 2017 steady state using the parameter values from the baseline calibration except for the values of ρ and η, which we take from Maliar, Maliar, and Tsener (2020). We find that the resulting stationary equilibrium still matches the U.S. economy very well in terms of our calibration targets. We then conduct the optimal capital tax reform for the CSC economy with the parameters ρ and η from Maliar, Maliar, and Tsener (2020), and as reported in the fourth panel of Table 10, we find that the optimal capital tax rate is 67%, the same as in the baseline case.[16]

Elasticity of Labor Supply. In our benchmark exercise, we take the parameter that controls the Frisch elasticity of labor supply to be $\gamma = 1$, which implies an elasticity of 1. As a sensitivity check, we conduct optimal tax exercise for an economy that is recalibrated assuming $\gamma = 2$ (Frisch elasticity equals 0.5).[17] The results of this exercise are reported in

[16]A key assumption in the analysis of Krusell, Ohanian, Ríos-Rull, and Violante (2000) is their choice of the time series of the price of equipment capital. This choice determines the time series of real stock of equipment capital in the data, which affects the estimation of elasticities. Polgreen and Silos (2008) conduct two sensitivity checks to Krusell, Ohanian, Ríos-Rull, and Violante (2000) by using two alternative series for the price of equipment capital. They estimate the production function given by (5) using these two alternative series. In a previous version of this paper, we conducted the optimal tax analysis for these two additional capital-skill complementarity economies and find that the optimal capital tax rates for these economies are also very similar to the one in the baseline capital-skill complementarity economy, providing further robustness to our baseline findings.

[17]The values of parameters that are taken from the literature are identical to those in the baseline calibration, and hence are reported in Table 1. The values of internally calibrated parameters are reported in Table 13 and Table 14, which are relegated to Appendix F for brevity.

the last panel of Table 10. The optimal capital tax rate equals 65% in the economy with capital-skill complementarity while it is 60% in the Cobb-Douglas economy. We conclude that the main finding - that the presence of capital-skill complementarity in production calls for higher optimal capital taxes - is not affected by Frisch elasticity, at least in the region of empirically plausible elasticities.

6 Tax Reforms with Richer Instruments

6.1 Differential Taxation of Equipment and Structures

In the baseline optimal tax exercise, we assume that the government is not allowed to tax equipment and structures differently in the CSC economy. This is mainly motivated by the fact that statutory tax rates on capital income derived from different types of capital is the same. However, effective tax rates can differ by capital type (mainly due to tax depreciation allowances that differ from actual depreciation rates). In this section, we consider a tax reform in which the government is allowed to tax equipment and structures at different rates. We find that optimal tax rate of structures is 65% while that on equipment is 69%, see the second panel of Table 11. The fact that equipment capital is optimally taxed at a higher rate than capital in the CD economy follows the same logic as in the baseline exercise. Capital structures are also taxed at a higher rate because the government does not want to set the two tax rates too far apart from each other in order to keep the productive efficiency distortions associated with taxing two capital types differently. Similar to the baseline exercise, average optimal tax rate on capital income in the CSC economy is about six percentage points higher than the optimal capital tax rate in CD economy.

The optimality of differential taxation of capital is in line with Slavík and Yazici (2014) who, relative to our four percentage point differential between taxes on equipment and structures, find that a much larger differential is optimal. This is mainly due to the fact that they do not take into account endogenous skill supply nor do they model heterogeneity beyond differences in skills.

Table 11: Optimal Taxes: Richer Instruments

	τ_s	τ_e	τ_l	λ
Status quo	0.36	0.36	0.10	0.89
Baseline				
Cobb-Douglas	0.61	0.61	0.10	0.95
Capital-Skill Complementarity	0.67	0.67	0.10	0.96
Differential Capital Taxation				
Cobb-Douglas (baseline)	0.61	0.61	0.10	0.95
Capital-Skill Complementarity	0.65	0.69	0.10	0.96
Comprehensive Reform				
Cobb-Douglas	0.63	0.63	0.37	0.75
Capital-Skill Complementarity	0.71	0.71	0.37	0.69

The first column of the table reports status-quo capital and average labor income taxes under the status-quo tax system. The second and third columns report optimal capital taxes and the steady-state value of average labor income taxes under the optimal tax system for both the Cobb-Douglas (CD) and capital-skill complementarity (CSC) models.

6.2 Comprehensive Reform

So far, we have focused on the effect of capital-skill complementarity on the optimal capital tax rate in the context of a tax reform in which the government is only able to adjust the capital tax rate along with the parameter that controls the average labor income tax, λ. In particular, this reform does not involve setting the labor income tax progressivity parameter, τ_l, optimally. We pursue this route mainly because, perhaps due to political constraints, it is often quite difficult for the government to implement comprehensive reforms in which the capital and labor tax codes are reformed substantially at the same time. This section aims to gauge the effect of capital-skill complementarity on the optimal capital tax rate in the context of such a comprehensive tax reform. To be precise, we consider the problem of a government which introduces a once and for all unannounced change in the capital tax rate, τ, and in labor tax progressivity, τ_l. As in the baseline, to ensure that its budget holds, the government adjusts the parameter that controls the average labor income tax, $\{\lambda_t\}_{t=0}^{\infty}$, along the transition to the new steady state. The welfare criterion puts equal weight on all the agents who are alive at the time of the reform and takes transition into account as in the baseline case.

The third panel of Table 11 summarizes our findings. Looking at the last two rows of the table, we see that the optimal capital tax rate differential between the two economies is even higher, namely 8 percentage points, in the comprehensive reform. Moreover, in

Table 12: Decomposition of Welfare Gains in Comprehensive Reform

	$\boldsymbol{\Delta}$	$\boldsymbol{\Delta_L}$	$\boldsymbol{\Delta_R}$	$\boldsymbol{\Delta_I}$
Cobb-Douglas	16.40	-13.46	22.25	10.02
Capital-Skill Complementarity	17.27	-14.89	23.60	11.47

The first panel of the table reports the total welfare gains of the reform (Δ) and its decomposition to level (Δ_L), redistribution (Δ_R) and insurance (Δ_I) effects in consumption equivalence units for the Cobb-Douglas economy while the second panel reports the same for the capital-skill complementarity economy.

both economies, the government finds it optimal to increase the capital tax rate beyond the level that is optimal in the baseline exercise, in tandem with much higher labor tax progressivity and average labor taxes. A glance at welfare gains decomposition given by Table 12 helps us make sense of this finding. The ability to increase labor tax progressivity in the comprehensive reform implies much larger insurance and redistribution gains relative to the baseline reform. This is due to the fact that, with higher progressivity, labor taxes are much more potent in making targeted transfers. This can be seen by looking at Figure 3 which compares average labor tax rates across the earnings distribution between the baseline and the comprehensive reforms. The high degree of progressivity in the comprehensive reform allows for substantial subsidies especially at the lower end of the earnings distribution.

The large redistribution and insurance gains come at the expense of immense level losses relative to the partial reform. This contraction of the economy implies that to finance the same level of spending, the government needs to raise substantially larger revenue (relative to income) through both capital and labor income taxes. This is why we observe a rise in both the capital tax rate and average labor taxes. Overall welfare gains from the comprehensive reform are very large which is a finding in line with Ferrière, Gruebener, Navarro, and Vardishvili (2022) who evaluate welfare gains of a similar reform. Unlike us, these authors find that such an optimal reform involves a reduction in tax progressivity. The divergence in findings follows mainly from the fact that in their model there is a transfer function, distinct from the progressive tax function we use, through which the government is able to transfer resources to the poor while the main way to achieve this in the context of the current paper is via progressive labor taxes.

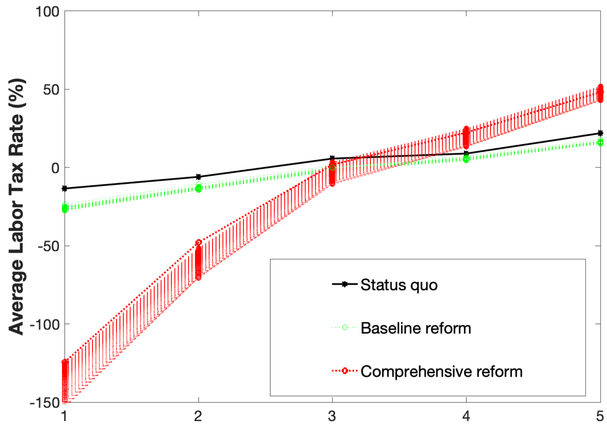


Figure 3: Average Labor Tax Rates in Baseline vs. Comprehensive Reform

This figure displays the average labor tax rate that applies to the mean earner in each earnings quintile under the status quo tax system (solid black line), baseline reform (green circles) and comprehensive reform (red circles). The vertical variation that corresponds to each quintile in baseline and comprehensive reforms report average tax rates in every period during transition. The figure is drawn for capital-skill complementarity economy.

6.3 Time-Varying Optimal Capital Taxes

In the baseline environment, we assume that the government chooses a capital tax rate that is constant over time. Although this is not an unreasonable assumption with regards to how actual tax rates are set in place, it is also interesting to consider the effect of capital-skill complementarity on optimal capital taxes in a world in which the government can commit to a time-varying sequence of capital tax rates.[18] We allow for time variation in the capital tax rate in the following parsimonious way:

$$\tau_t = \underline{\tau} exp(-\xi \cdot t) + (1 - exp(-\xi \cdot t)\overline{\tau}, \tag{7}$$

where t is time and $(\underline{\tau}, \overline{\tau}, \xi)$ denote the initial tax rate, the final tax rate, and a parameter that controls the speed of transition from the initial to the final tax rate, respectively. At the beginning of the reform, the government announces and commits to a capital tax

[18]Following Aiyagari (1995), only recently researchers have begun to analyze optimal time-varying Ramsey tax problems in economies with heterogenous agents. Dyrda and Pedroni (2022) provides one such optimal tax analysis in an incomplete markets framework with realistic degrees of heterogeneity similar to ours but using a much more flexible capital tax function. See also Acikgoz, Hagedorn, Holter, and Wang (2018) for a similar analysis.

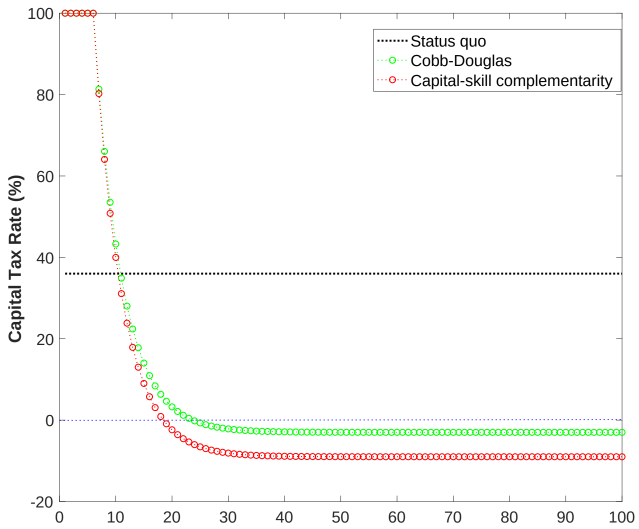


Figure 4: Optimal Time-Varying Capital Taxes

The six graphs report how the skill premium, relative skilled wealth, relative skilled consumption, aggregate capital stock, aggregate output, and the capital-output ratio change over the transition following the optimal tax reform. CSC and CD refer to capital-skill complementarity and Cobb-Douglas economies, respectively.

policy along with a sequence, $\{\lambda_t\}_{t=0}^{\infty}$, which ensures that the government budget balances every period.

In addition to equation (7), we also assume an upper bound on the tax rate $\tau_t \leq 1$, following the literature. At the optimum, this constraint binds for a certain number of periods after which the planner finds it optimal to decrease capital taxes below 100%. Therefore, rather than choosing the initial tax rate, the planner chooses the number of periods of $\tau_t = 100\%$. After these periods, the capital tax policy follows equation (7) with an additional simplification, namely that the decay rate $\xi = 0.2$.[19]

The optimal tax rates on capital in the economies with and without capital-skill complementarity are depicted in Figure 4. In both economies the planner finds it optimal to set $\tau_t = 100\%$ for six 6 periods. The result that if the planner can choose capital taxes that vary over time, she will indeed choose very high capital taxes early on to combat inequality, is well known from the literature. After these 6 initial periods, the

[19]We fix ξ mainly for computational tractability. This assumption is justified based on the observation that the planner has one too many instruments locally. With this tax function, the planner can approximate a given path of tax rates, at least in the short run, with another combination of decay rate and terminal tax rate with virtually no impact on welfare.

optimal capital tax rates are higher in the CD economy. The reason for this seemingly contradictory finding is as follows.

The initial 6 periods of high capital taxation are very effective in reducing inequality in both economies, but especially in the CSC economy, as the massive decline in capital stock reduces skill premium substantially in this economy (but not in the CD economy). As a result, the two economies are not identical any more once the planner finds it optimal to leave the upper bound of 100% in period 7: the skill premium is substantially lower, namely 1.85, and hence, the need for redistribution is smaller in the CSC economy, which calls for lower capital taxes. On the other hand, capital taxation is still a more effective redistribution tool in the CSC economy due to the indirect redistribution channel it provides. In period 7, these two forces offset each other to a large extent, and optimal capital tax in CD economy is only about 1% higher than that in CSC economy.

Capital taxation is an especially effective redistribution tool at the beginning of the reform since most of the population - except for 2.2% that are newly born - are fixed into their skill types. Over time, the fraction of the population that makes a skill choice increases, increasing the elasticity of the fraction of skilled with respect to capital taxes and reducing the effectiveness of capital taxation as a redistributive tool in the CSC economy. This implies a rise in the difference between the optimal capital tax rate across the two economies over time. Eventually, the differential converges to 6% as the elasticity of the fraction of skilled workers converge.

In contrast to the theoretical characterization of Aiyagari (1995), we find that the optimal long-run capital tax rate is negative in both economies. This is possible since our analysis differs from his in three substantive ways. First, we have a different demographic structure where people die and are replaced by newborns, and the future generations do not enter the social welfare function. Second, we model endogenous skill choice, which is an additional margin through which capital taxation distorts the economy. Finally, we assume a parametric tax function (7) with an additional restriction that the decay rate is fixed.

7 Conclusion

This paper shows that capital-skill complementarity provides a quantitatively significant rationale for taxing capital for redistributive governments. Importantly, it does so using a rich quantitative model with endogenous skill acquisition that allows us to replicate the degree of earnings and wealth inequality observed in the U.S. economy. The paper finds that it is optimal to rely more on capital income taxes and less on labor income taxes when capital-skill complementarity is taken into account. The welfare gains of an optimal tax reform are also significantly larger in the presence of capital-skill complementarity. Given the overwhelming empirical evidence on the presence of capital-skill complementarities in production, our analysis suggests that governments should take into account the presence of such complementarities when setting capital tax rates.

References

Acikgoz, O., M. Hagedorn, H. Holter, and Y. Wang (2018): "The Optimum Quantity of Capital and Debt," Cepr discussion paper 12952.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 109(3), 659–84.

——— (1995): "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting," *Journal of Political Economy*, 103(6), 1158–75.

Angelopoulos, K., S. Asimakopoulos, and J. Malley (2015): "Tax Smoothing in a Business Cycle Model with Capital-Skill Complementarity," *Journal of Economic Dynamics and Control*, 51, 420–444.

Benabou, R. (2002): "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 70(2), 481–517.

Bewley, T. F. (1986): "Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers," in *Contributions to Mathematical Economics in Honor of Gerard Debreu*, ed. by W. Hildenbrand, and A. Mas-Colell, pp. 79–102. North Holland.

Bhattarai, S., J. W. Lee, W. Y. Park, and C. Yang (2020): "Macroeconomic Effects of Capital Tax Rate Changes," Working paper.

Boar, C., and V. Midrigan (2022): "Efficient Redistribution," *Journal of Monetary Economics*, forthcoming.

Castaneda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003): "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111(4), 818–857.

Chamley, C. (1986): "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 54(3), 607–622.

Chari, V., J. P. Nicolini, and P. Teles (2020): "Optimal capital taxation revisited," *Journal of Monetary Economics*, 116, 147–165.

Conesa, J. C., S. Kitao, and D. Krueger (2009): "Taxing Capital? Not a Bad Idea after All!," *American Economic Review*, 99(1), 25–48.

Costinot, A., and I. Werning (2018): "Robots, Trade, and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation," NBER Working Papers 25103, National Bureau of Economic Research, Inc.

Cummins, J. G., and G. L. Violante (2002): "Investment-Specific Technical Change in the US (1947-2000): Measurement and Macroeconomic Consequences," *Review of Economic Dynamics*, 5(2), 243–284.

DiCecio, R. (2009): "Sticky wages and sectoral labor comovement," *Journal of Economic Dynamics and Control*, 33(3), 538–553.

Dolado, J. J., G. Motyovszki, and E. Pappa (2020): "Monetary Policy and Inequality under Labor Market Frictions and Capital-Skill Complementarity," *American Economic Journal: Macroeconomics*, forthcoming.

Domeij, D., and J. Heathcote (2004): "On the Distributional Effects of Reducing Capital Taxes," *International Economic Review*, 45(2), 523–554.

Duffy, J., C. Papageorgiou, and F. Perez-Sebastian (2004): "Capital-Skill Complementarity? Evidence from a Panel of Countries," *Review of Economics and Statistics*, 86(1), 327–344.

Dyrda, S., and M. Pedroni (2022): "Optimal Fiscal Policy in a Model with Uninsurable Idiosyncratic Shocks," *Review of Economic Studies*, forthcoming.

Dyrda, S., and B. Pugsley (2019): "Taxes, Private Equity, and Evolution of Income Inequality in the US," Working paper.

Fallon, P. R., and P. R. G. Layard (1975): "Capital-Skill Complementarity, Income Distribution, and Output Accounting," *Journal of Political Economy*, 83(2), 279–302.

Ferrière, A., P. Gruebener, G. Navarro, and O. Vardishvili (2022): "On the Optimal Design of Transfers and Income-Tax Progressivity," Working paper.

Ferrière, A., and G. Navarro (2018): "The Heterogeneous Effects of Government Spending: It's All About Taxes," International Finance Discussion Papers 1237, Board of Governors of the Federal Reserve System (U.S.).

Floden, M. (2001): "The effectiveness of government debt and transfers as insurance," *Journal of Monetary Economics*, 48(1), 81–108.

Flug, K., and Z. Hercowitz (2000): "Equipment Investment and the Relative Demand for Skilled Labor: International Evidence," *Review of Economic Dynamics*, 3, 461–485.

Freeman, R. B. (1982): "Elasticities of Demand for Educated Labor and Elasiticities of Supply of Educated Labor," Discussion paper.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

Greenwood, J., Z. Hercowitz, and P. Krusell (1997): "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, 87(3), 342–362.

Griliches, Z. (1969): "Capital-Skill Complementarity," *Review of Economics and Statistics*, 51(4), 465–468.

Guerreiro, J., S. Rebelo, and P. Teles (2021): "Should Robots Be Taxed?," *The Review of Economic Studies*, 89(1), 279–311.

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., K. Storesletten, and G. L. Violante (2010): "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, 118(4), 681–722.

——— (2017): "Optimal Tax Progressivity: An Analytical Framework," *The Quarterly Journal of Economics*, 132(4), 1693–1754.

Hong, J. H., B. H. Seok, and H. M. You (2019): "Wage Volatility and Changing Patterns of Labor Supply," *International Economic Review*, 60(2), 595–630.

Huggett, M. (1993): "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 17(5-6), 953–969.

Imrohoroglu, A. (1989): "Cost of Business Cycles with Indivisibilities and Liquidity Constraints," *Journal of Political Economy*, 97(6), 1364–1383.

Imrohoroglu, S. (1998): "A Quantitative Analysis of Capital Income Taxation," *International Economic Review*, 39(2), 307–28.

Jones, L., R. Manuelli, and P. Rossi (1997): "On the Optimal Taxation of Capital Income," *Journal of Economic Theory*, 73(1), 93–117.

Judd, K. L. (1985): "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28(1), 59–83.

Karabarbounis, L., and B. Neiman (2014): "The Global Decline of the Labor Share," *The Quarterly Journal of Economics*, 129(1), 61–103.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

Kopecky, K., and R. Suen (2010): "Finite State Markov-chain Approximations to Highly Persistent Processes," *Review of Economic Dynamics*, 13(3), 701–714.

Krueger, D., and A. Ludwig (2016): "On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium," *Journal of Monetary Economics*, 77, 72 – 98, Inequality, Institutions, and Redistribution held at the Stern School of Business, New York University, April 24-25, 2015.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000): "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 68(5), 1029–1054.

Kuhn, M., and J.-V. Ríos-Rull (2020): "2016 Update on the U.S. Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics," Working paper.

Maliar, L., S. Maliar, and I. Tsener (2020): "Capital-Skill Complementarity and Inequality: Twenty Years After," Cepr discussion paper no: 15228.

Mendoza, E. G., A. Razin, and L. L. Tesar (1994): "Effective Tax rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption," *Journal of Monetary Economics*, 34(3), 297–323.

Ohanian, L., M. Orak, and S. Shen (2021): "Revisiting Capital-Skill Complementarity, Inequality, and Labor Share," Working paper.

Polgreen, L., and P. Silos (2008): "Capital-Skill Complementarity and Inequality: A Sensitivity Analysis," *Review of Economic Dynamics*, 11(2), 302–313.

Ríos-Rull, J.-V., and R. Santaeulàlia-Llopis (2010): "Redistributive shocks and productivity shocks," *Journal of Monetary Economics*, 57(8), 931–948.

Slavík, C., and H. Yazici (2014): "Machines, Buildings, and Optimal Dynamic Taxes," *Journal of Monetary Economics*, 66, 47 – 61.

Slavík, C., and H. Yazici (2022): "Wage Risk and the Skill Premium," *The Economic Journal*, 132(646), 2207–2230.

Straub, L., and I. Werning (2020): "Positive Long-Run Capital Taxation: Chamley-Judd Revisited," *American Economic Review*, 110(1), 86–119.

Thuemmel, U. (2020): "Optimal Taxation of Robots," Cesifo working papers.

Trabandt, M., and H. Uhlig (2011): "The Laffer curve revisited," *Journal of Monetary Economics*, 58(4), 305–327.

Appendix

A Definition of Competitive Equilibrium for the Cobb-Douglas Economy

The state of a worker of type i in a period t is fully described by the worker's productivity and asset holdings. Let $(z_i, a_i) \in \mathcal{Z}_i \times \mathcal{A}$ denote this state. Let $\Lambda_{i,t}(z_i, a_i)$ denote the distribution of workers of type i across productivities and assets. The initial, $t = 0$, distributions are given exogeneously.

Definition: Given initial conditions, a recursive competitive equilibrium is a government policy $(T_t(.), \tau_t, D_t, G_t)_{t=0}^{\infty}$, allocation for the firm, $(K_t, L_{s,t}, L_{u,t})_{t=0}^{\infty}$, value and policy functions for agents, $(v_{i,t}(z_i, a_i), c_{i,t}(z_i, a_i), l_{i,t}(z_i, a_i), a_{i,t+1}(z_i, a_i))_{t=0,i=u,s}^{\infty}$, skill choices, shares of workers who are skilled, $(\pi_{s,t})_{t=0}^{\infty}$, a price system $(r_t, w_{s,t}, w_{u,t}, R_t)_{t=0}^{\infty}$ and distributions over individual states, $(\Lambda_{i,t}(z_i, a_i))_{t=0,i=u,s}^{\infty}$, such that:

1. In each period $t \geq 0$, taking factor prices as given, $(K_t, L_{s,t}, L_{u,t})$ solves the firm's problem given by:

$$\max_{K_t, L_{s,t}, L_{u,t}} \quad F(K_t, L_{s,t}, L_{u,t}) - r_t K_t - w_{s,t} L_{s,t} - w_{u,t} L_{u,t},$$

2. Given government policy and the price system, the policy functions solve the consumer's problem given by (2).

3. Skill choice is consistent with (3), that is in any period t, all those with $\psi \leq \overline{\psi}_t$ attend college and all other do not. Moreover, the evolution of the fraction of skilled in each period is consistent with skill choice: $\pi_{s,t} = \chi \pi_{s,t-1} + (1 - \chi)\pi_{s,t}^n$, where $\pi_{s,t}^n = \int_{\mathbb{R}_+} I_{\psi \leq \overline{\psi}_t}(\psi) dH(\psi)$ is the fraction of newborns who choose to become skilled in period t and $I_{\psi \leq \overline{\psi}_t}(\psi)$ is the indicator function, $\pi_{u,t}^n = 1 - \pi_{s,t}^n$ for all t, and $\pi_{s,0}$ is given.

4. The evolution of distributions of agents across productivities and assets over time is consistent with agent choices. That is, for all $t \geq 0$, $i = u, s$ and $(z_i', a_i') \in \mathcal{Z}_i \times \mathcal{A}$:

$$\Lambda_{i,t+1}(z_i', a_i') = \frac{\chi \sum_{z_i \in \mathcal{Z}_i} \Pi_i(z_i'|z_i) \int_{\{a_i : a_{i,t+1}(z_i, a_i) \leq a_i'\}} d\Lambda_{i,t}(z_i, a_i) + (1 - \chi)\pi_{i,t+1}^n \Lambda_i^z(z_i')}{\chi + (1 - \chi)\pi_{i,t+1}^n},$$

   where $(\Lambda_{i,0}(z_i, a_i))_{i=u,s}$ is given and Λ_i^z is the stationary distribution associated with the Markov chain that describes the evolution of the productivity shock for type i.

5. Markets for assets, labor and goods clear: for all $t \geq 0$,

$$K_t + D_t = \sum_{i=u,s} \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} a_{i,t}(z_i, a_i) d\Lambda_{i,t-1}(z_i, a_i),$$

$$L_{i,t} = \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} l_{i,t}(z_i, a_i) z_i d\Lambda_{i,t}(z_i, a_i), \text{ for } i = u, s,$$

$$G_t + C_t + K_{t+1} = F(K_t, L_{s,t}, L_{u,t}) + (1 - \delta)K_t,$$

where

$$C_t = \sum_{i=u,s} \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} c_{i,t}(z_i, a_i) d\Lambda_{i,t}(z_i, a_i)$$

is aggregate consumption in period t.

6. The government's budget constraint is satisfied every period: for all $t \geq 0$,

$$G_t + R_t D_t = D_{t+1} + \tau_t(r_t - \delta)K_t + \sum_{i=u,s} \pi_{i,t} \int_{\mathcal{Z}_i \times \mathcal{A}} T_t(l_{i,t}(z_i, a_i) w_{i,t} z_i) d\Lambda_{i,t}(z_i, a_i).$$

B Data Construction

Fraction of skilled agents. The fraction of skilled agents is calculated using Current Population Survey ASEC (March) data administered by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics. We use data from the 2018 survey which includes information about 2017. We focus on males aged 25 and older with earnings and follow Krusell, Ohanian, Ríos-Rull, and Violante (2000) by defining the fraction of skilled agents as the ratio agents with a bachelor's degree or more divided by the total number of agents in Table P-16.

Government consumption-to-GDP ratio. The government consumption-to-output ratio is recovered from the National Income and Product Accounts (NIPA) data. It is defined as the ratio of nominal government consumption expenditure (line 15 in NIPA Table 3.1) to nominal GDP (line 1 in NIPA Table 1.1.5).

Government debt-to-GDP ratio. The government debt to GDP ratio is taken from the St. Louis FED database FRED for year 2015. The data series is called "Federal Debt Held by Private Investors as Percent of Gross Domestic Product" (series ID: HBPIGDQ188S). The precise number for 2015 is 59.2% which we round to 60% (government debt-to-GDP ratio keeps increasing after 2015).

Share of equipments in total capital stock. The share of equipment capital in total capital stock is calculated using Fixed Asset Tables (FAT) data. It is defined as the ratio of private equipment capital (line 5 in FAT Table 1.1) to the sum of private equipment and structure capital (line 5 + line 6 in FAT Table 1.1). This calculation gives a value of 0.32 in 2017, which we round to 1/3.

Capital-to-output ratio. Housing is excluded from both output and capital when calculating the capital-to-output ratio. For this calculation, output is defined using Table

1.5.5 in NIPA as GDP (line 1) net of Housing and utilities (line 16) and Residential investment (line 41). Capital stock is calculated using the Fixed Asset Tables (FAT), Table 1.1 as the sum of the stocks of private and government structure and equipment capital (line 5 + line 6 + line 11 + line 12). The ratio is relatively stable after 2015. We use the value of 2.07, which is the value for 2017.

Cross-sectional inequality statistics. All cross-sectional income and wealth moments (Gini for earnings and wealth, top 1% shares, quintile shares and relative skilled wealth) reported in Table 2 and in Table 4 are taken from Kuhn and Ríos-Rull (2020) and correspond to year 2016. The data source used in Kuhn and Ríos-Rull (2020) is the Survey of Consumer Finances. The definition of skilled and unskilled agents is consistent with the rest of the paper: Skilled agents are those of 16 years of education or more. In SCF, this corresponds to bachelors degree or higher (as reported for the head of the household, typically a male).

C The 1967 Economy

This section provides a detailed description of the steady state that corresponds to the 1967 U.S. economy. The 1967 economy is computed by taking the capital-skill complementarity model with all of its parameters that are calibrated to match the 2017 U.S. economy, as reported in Table 1, Table 2, and Table 3, and changing only the price of equipment, the relative supply of skilled workers and tax policy to their values from 1967. Below we explain how we constructed the changes in these three key factors.

Price of equipment in 1967. Following the methodology of Cummins and Violante (2002), DiCecio (2009) calculates the historical price of equipment capital in consumption good units. To quantify the decline in the price of equipment across the two steady states, we calculate the ratio of the price of equipment in 1967 to that in 2017. The price of equipment decreased by a factor of 16 over this period. (Averaging the price of equipment over five year periods centered around 1967 and 2017 does not change the resulting ratio.) Since we normalize the price of equipment to 1 in 2017 steady state, the price of equipment is set to 16 in 1967 steady state. Since different types of labor have different elasticity of substitution with equipment capital, the decline in the relative price of equipment capital endogeneously implies a change in the skill premium, i.e., skill-biased technical change. In the calculations provided by both Cummins and Violante (2002) and DiCecio (2009), the price of structure capital relative to consumption remains virtually constant during this period. For this reason, we keep the price of structures in 1967 at its normalized price of 1.

Supply of Skilled Workers in 1967. We compute the fraction of skilled workers for 1967 following the same procedure we use to compute it for 2017. We consider only males who are 25 years and older and who have earnings and use data from CPS 1967. We find that the fraction of skilled workers was 0.1356 in 1967.

Government policies in 1967. Trabandt and Uhlig (2011), from whom we take the capital tax rate for the 2017 steady state, use the methodology of Mendoza, Razin, and Tesar (1994) in calculating this tax rate. Since Trabandt and Uhlig (2011) only go back in time as far as 1995, we take the tax rate on capital income for 1967 directly from Mendoza, Razin, and Tesar (1994). Since effective capital tax rate estimates are sensitive to short-term fluctuations in the inflation rate, we take an average over the five year window centered around 1967, which gives a capital tax rate of 41%. As for labor income taxes, Ferrière and Navarro (2018) estimate a value of 0.12 for the five year period centered around 1967 for the tax parameter τ_l, which represents the progressivity of the U.S. tax system.

As noted earlier, government consumption to GDP ratio is relatively stable over time at 16%, we use this number for the 1967 steady state as well. In contrast, the government debt to GDP ratio, defined as before as "Federal Debt Held by Private Investors as Percent of Gross Domestic Product" (series ID: HBPIGDQ188S) is 21% in 1970, the earliest date for this time series. We use this number to represent the late 1960s (the government-debt-to-GDP ratio was relatively stable throughout the 1970s).

Comparison of the 1967 and 2017 economies. Table 5 compares the 1967 and 2017 model economies to the data along several dimension. (i) Skill premium comes from Heathcote, Perri, and Violante (2010), (ii) the share of equipment in 1967 is computed analogously to 2017, as described above. (iii) The labor share is computed from NIPA using the methodology described in Ríos-Rull and Santaeulàlia-Llopis (2010) and for details, we refer the reader to that paper. It offers several alternative ways of calculating the labor share. We use the following: we first calculate what Ríos-Rull and Santaeulàlia-Llopis (2010) call "unambiguously capital income" and "unambiguously labor income." Income which cannot be unambiguously classified as labor or capital income is then divided between capital and labor using the ratio between capital and labor income in unambiguously assigned income. To get the labor share, labor income is then divided by GNP. (iv) Real GDP is the series A939RX0Q048SBEA from St. Louis FED FRED database. (v) Investment-to-output ratio. Housing is excluded from both output and investment when calculating the capital-to-output ratio. For this calculation, output is defined using Table 1.5.5 in NIPA as GDP (line 1) net of Housing and utilities (line 16) and Residential investment (line 41). Investment is calculated using same table as the sum of the stocks of private and government non-residential investment (line 28 + line 56 + line 59 + line 62).

D Calibration of Cost of Skill Acquisition

This section provides a full description of the details of the calibration of the cost of skill acquisition. Consider the 2017 steady state. The model implies a utility premium for skilled workers at this steady state that is given by:

$$E_{s,2017}[v_{s,2017}(z_s, 0)] - E_{u,2017}[v_{u,2017}(z_u, 0)].$$

For the marginal individual who is indifferent between acquiring a college degree or not, this utility premium equals the cost of skill acquisition. That is, letting $\bar{\psi}_{s,2017}$ be the cost of skill acquisition for marginal worker, we have

$$\bar{\psi}_{s,2017} = E_{s,2017}[v_{s,2017}(z_s, 0)] - E_{u,2017}[v_{u,2017}(z_u, 0)].$$

Therefore, it has to be that $H(\bar{\psi}_{s,2017}) = \pi_{s,2017}$. An identical argument applied to 1967 steady state implies $H(\bar{\psi}_{s,1967}) = \pi_{s,1967}$. Assuming H is log-normally distributed with a mean and variance, we have two unknowns and two equations, which pins down the mean and variance of the distribution. The mean and the standard deviation of the normal distribution that corresponds to the calibrated H are 0.045 and 0.434.

E Decomposition of Welfare Gains

In welfare gains decompositions, it is more convenient to work with sequential definitions of allocations rather than the recursive definitions given until now. For that reason, we first give equivalent sequential definitions of allocations. Let $v_0 = (i, z_{i,0}, a_{i,0}) \in V_0$ denote a person's type in the initial steady-state distribution. This initial type is distributed according to some distribution $\Lambda_0(v_0)$. Although Λ_0 can be constructed from $\Lambda_{i,0}(z_{i,0}, a_{i,0})$ for $i = u, s$ which is given in the definition of equilibrium in Section A, we will not do so here as this is not needed for the welfare gains decomposition.

Denote the uncertain consumption-labor allocation a for type v_0 by $\Big\{\{c^a_{v_0,t}\}, \{l^a_{v_0,t}\}\Big\}$. Utility from this allocation is given by

$$U\Big(\{c^a_{v_0,t}\}, \{l^a_{v_0,t}\}\Big) = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c^a_{v_0,t}) - v(l^a_{v_0,t})\right],$$

where the expectation is taken over productivity shocks conditional on initial type.

Define welfare gains of moving from allocation b to allocation a as:

$$\int_{v_0 \in V_0} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c^a_{v_0,t}) - v(l^a_{v_0,t})\right] d\Lambda_0(v_0) = \int_{v_0 \in V_0} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u((1+\Delta)c^b_{v_0,t}) - v(l^b_{v_0,t})\right] d\Lambda_0(v_0). \tag{8}$$

Insurance effect. Let average levels of consumption and labor in period t for a given initial type in allocation a be

$$C^a_{v_0,t} = \mathbb{E}_t c^a_{v_0,t} \quad \text{and} \quad L^a_{v_0,t} = \mathbb{E}_t l^a_{v_0,t}.$$

Define the cost of risk for initial type v_0 in allocation a as

$$\sum_{t=0}^{\infty} \beta^t \left[u((1-p^a_{v_0,risk})C^a_{v_0,t}) - v(L^a_{v_0,t})\right] = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c^a_{v_0,t}) - v(l^a_{v_0,t})\right]. \tag{9}$$

The insurance effect Δ_I is then defined as

$$\log(1+\Delta_I) = \int_{v_0 \in V_0} \log\left(1+\Delta_{I,v_0}\right) d\Lambda_0(v_0). \tag{10}$$

where $1 + \Delta_{I,v_0} = \frac{1-p^a_{v_0,risk}}{1-p^b_{v_0,risk}}$ is the insurance effect for initial type v_0. Notice that the (aggregate) insurance effect is a weighted average of individual insurance effects in logs.

Redistribution effect. Let aggregate levels of consumption and labor in period t in allocation a be

$$C_t^a = \int_{v_0 \in V_0} C_{v_0,t}^a d\Lambda_0(v_0) \quad \text{and} \quad L_t^a = \int_{v_0 \in V_0} L_{v_0,t}^a d\Lambda_0(v_0).$$

Define cost of inequality in allocation a as

$$\sum_{t=0}^{\infty} \beta^t \left[u((1-p_{ineq}^a)C_t^a) - v(L_t^a) \right] = \int_{v_0 \in V_0} \sum_{t=0}^{\infty} \beta^t \left[u(C_{v_0,t}^a) - v(L_{v_0,t}^a) \right] d\Lambda_0(v_0). \tag{11}$$

The redistribution effect Δ_R is then defined by

$$1 + \Delta_R = \frac{1-p_{ineq}^a}{1-p_{ineq}^b}. \tag{12}$$

Level effect. Define level effect as

$$\sum_{t=0}^{\infty} \beta^t \left[u(C_t^a) - v(L_t^a) \right] = \sum_{t=0}^{\infty} \beta^t \left[u((1+\Delta_L)C_t^b) - v(L_t^b) \right]. \tag{13}$$

Proposition 1. *If* $u(c) = \log(c)$, *then*

$$1 + \Delta = (1+\Delta_I)(1+\Delta_R)(1+\Delta_L).$$

Proof.

$$\begin{aligned}
& \int_{v_0 \in V_0} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c_{v_0,t}^a) - v(l_{v_0,t}^a) \right] d\Lambda_0(v_0) \\
= & \int_{v_0 \in V_0} \log(1-p_{v_0,risk}^a) d\Lambda_0(v_0) + \int_{v_0 \in V_0} \sum_{t=0}^{\infty} \beta^t \left[u(C_{v_0,t}^a) - v(L_{v_0,t}^a) \right] d\Lambda_0(v_0) \\
= & \int_{v_0 \in V_0} \log(1-p_{v_0,risk}^a) d\Lambda_0(v_0) + \log(1-p_{ineq}^a) + \sum_{t=0}^{\infty} \beta^t \left[u(C_t^a) - v(L_t^a) \right] \\
= & \int_{v_0 \in V_0} \log(1-p_{v_0,risk}^a) d\Lambda_0(v_0) + \log(1-p_{ineq}^a) + \log(1+\Delta_L) + \sum_{t=0}^{\infty} \beta^t \left[u(C_t^b) - v(L_t^b) \right] \\
= & \int_{v_0 \in V_0} \log(1-p_{v_0,risk}^a) d\Lambda_0(v_0) + \log(1-p_{ineq}^a) \\
& + \log(1+\Delta_L) - \log(1-p_{ineq}^b) + \int_{v_0 \in V_0} \sum_{t=0}^{\infty} \beta^t \left[u(C_{v_0,t}^b) - v(L_{v_0,t}^b) \right] d\Lambda_0(v_0) \\
= & \int_{v_0 \in V_0} \log(1-p_{v_0,risk}^a) d\Lambda_0(v_0) + \log(1-p_{ineq}^a) + \log(1+\Delta_L) - \log(1-p_{ineq}^b) \\
& - \int_{v_0 \in V_0} \log(1-p_{v_0,risk}^b) d\Lambda_0(v_0) + \int_{v_0 \in V_0} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u(c_{v_0,t}^b) - v(l_{v_0,t}^b) \right] d\Lambda_0(v_0) \\
= & \int_{v_0 \in V_0} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[u((1+\Delta_I)(1+\Delta_R)(1+\Delta_L)c_{v_0,t}^b) - v(l_{v_0,t}^b) \right] d\Lambda_0(v_0),
\end{aligned}$$

where the first equality follows from (9), the second one follows from (11), the third one from (13), the fourth one follows from (11) and the fifth equality follows from (9). A comparison of the ultimate equality with the definition of welfare gains given by (8) finishes the proof. $\square$

Remark. *An alternative way of defining aggregate insurance component would be as follows:*

$$\sum_{t=0}^{\infty}\beta^t\int_{v_0\in V_0}\left[u((1-p^a_{risk})C^a_{v_0,t})-v(L^a_{v_0,t})\right]d\Lambda_0(v_0)=\int_{v_0\in V_0}\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\left[u(c^a_{v_0,t})-v(l^a_{v_0,t})\right]d\Lambda_0(v_0)$$

and

$$1+\Delta_I=\frac{1-p^a_{risk}}{1-p^b_{risk}}. \tag{14}$$

In case of logarithmic utility, the two definitions, given by (10) *and* (14)*, deliver an identical aggregate insurance effect.*

F $\gamma = 2$ Calibration

Table 13 and Table 14 below report the values of internally calibrated parameters for the version of the model in which $\gamma = 2$. As in the baseline, we solve another version of the model which represents 1967 and recalibrate the distribution of the cost of skill acquisition, H, to match skill acquisition in the data.The mean and the standard deviation of the normal distribution that corresponds to the calibrated H are now 0.015 and 0.451 (not reported in the tables below).

Table 13: $\gamma = 2$ Calibration: Aggregate Moments

Parameter	**Symbol**	**Value**	**Target**	**Source**
Technology (CSC)				
Production parameter	ω	0.2833	Labor share = 2/3	NIPA
Production parameter	ν	0.6573	Skill premium = 1.9	CPS
Production parameter	α	0.1909	Share of equipments, $\frac{K_e}{K}=1/3$	FAT
Technology (CD)				
Total factor productivity	A	0.7869	Output level of CSC economy	
Production parameter	κ	0.5570	Skill premium = 1.9	CPS
Common parameters				
Discount factor	β	0.9378	Capital to output ratio = 2.07	NIPA, FAT
Tax function parameter	λ	0.8839	Government budget balance	
Disutility of labor	ϕ	29.40	Labor supply = 1/3	

This table reports the calibration procedure for parameters that target aggregate moments for the case when $\gamma = 2$. Model generated target moments are not reported as the match is perfect. The production function parameters α, ν and ω control the income shares of structure capital, equipment capital, skilled and unskilled labor in the capital-skill complementarity model (CSC). The production function parameter κ controls the income shares of the skilled and unskilled labor in the Cobb-Douglas model (CD). The tax function parameter λ controls the labor income tax rate of the mean income agent. The acronyms CPS, FAT, and NIPA stand Current Population Survey, Fixed Asset Tables, and National Income and Product Accounts, respectively.

Table 14: $\gamma = 2$ Calibration: Distributional Moments

Panel A: Moments	**Data**	**Model**
Earnings Gini	0.68	0.66
Earnings Gini - skilled	0.66	0.66
Earnings Gini - unskilled	0.61	0.62
Earnings Top 1%'s share	0.23	0.24
Earnings autocorrelation	0.94	0.95
Wealth Gini	0.86	0.85
Wealth Gini - skilled	0.81	0.81
Wealth Gini - unskilled	0.82	0.81
Wealth Top 1%'s share	0.39	0.38
Relative skilled wealth	5.6	5.6
Panel B: Parameters	**Symbol**	**Value**
Normal state persistence (skilled)	ρ_s	0.7853
Normal state volatility of shocks (skilled)	$var(\varepsilon_s)$	0.1848
Transit into superstar state (skilled)	p_s	1×10^{-3}
Remain in superstar state (skilled)	q_s	0.9496
Productivity superstar state (skilled)	e_s	46.00
Normal state persistence (unskilled)	ρ_u	0.9947
Normal state volatility of shocks (unskilled)	$var(\varepsilon_u)$	0.0342
Transit into superstar state (unskilled)	p_u	8×10^{-5}
Remain in superstar state (unskilled)	q_u	0.0216
Productivity superstar state (unskilled)	e_u	43.45

This table reports calibration results regarding the wage risk parameters for the case when $\gamma = 2$. The model's ability to match calibration targets are reported in Panel A and the calibrated parameter values are reported in Panel B. All data moments correspond to 2016 U.S. economy and are taken from Kuhn and Ríos-Rull (2020), with the exception of the autocorrelation of earnings, which is reported in Boar and Midrigan (2022). Relative skilled wealth refers to the ratio of the average skilled asset holdings to the average unskilled asset holdings.