Machines, Buildings, and Optimal Dynamic Taxes*

Ctirad Slavík[a] and Hakki Yazici[b]

[a]Goethe University Frankfurt, Frankfurt, Germany. Email: slavik@econ.uni-frankfurt.de

[b]Sabanci University, Istanbul, Turkey. E-mail: hakkiyazici@sabanciuniv.edu

March 21, 2014

Abstract

The effective taxes on capital returns differ depending on capital type in the U.S. tax code. This paper uncovers a novel reason for the optimality of differential capital taxation. We set up a model with two types of capital - equipments and structures - and equipment-skill complementarity. Under a plausible assumption, we show that it is optimal to tax equipments at a higher rate than structures. In a calibrated model, the optimal tax differential rises from 27 to 40 percentage points over the transition to the new steady state. The welfare gains of optimal differential capital taxation can be as high as 0.4% of lifetime consumption.

JEL classification: E62, H21.

Keywords: Differential capital asset taxation, equipment capital, structure capital, equipment-skill complementarity.

*Corresponding author: Ctirad Slavík, Grüneburgplatz 1 (Campus Westend) House of Finance, Room 3.49 60323 Frankfurt am Main, Germany. Telephone: +49 (0)69 798 33808.

1 Introduction

In the U.S. corporate tax code, the effective marginal tax rates on returns to capital assets show a considerable amount of variation depending on the capital type. For instance, according to Gravelle (2011), the effective marginal tax rate on the returns to communications equipment is 19%, whereas it is above 35% for non-residential buildings.[1] This feature of the tax code has been the subject of numerous reform proposals since the 1980s. Recently, President Obama called for a reform to abolish the tax rules that create differential taxation of capital assets in order to "level the playing field" across companies.[2] Many economists have argued in favor of the proposals to abolish tax differentials following an efficiency argument first raised by Diamond and Mirrlees (1971): taxing different types of capital at different rates distorts firms' production decisions, thereby creating production inefficiencies.

This paper takes a step back and reassesses whether differential taxation of capital income is a desirable feature of the tax code. Theoretically, the paper uncovers a novel economic mechanism that calls for optimality of differential capital asset taxation, but with an important caveat. In the current U.S. tax code, the effective tax rate on equipment capital (i.e., mostly machines) is on average 5% below the effective tax rate on structure capital (i.e., mostly non-residential buildings). In contrast, our theory suggests that capital equipments should be taxed at a higher rate than capital structures. We conduct a quantitative exercise to assess the quantitative importance of optimal differential capital taxation. In our baseline calibration, the tax rate on capital equipments should be at least 27 percentage points higher than the tax rate on capital structures in the transition and at the steady state. Furthermore, the welfare gains of optimal differential capital taxation are as high as 0.4% of lifetime consumption for reasonable parameter values.

We study dynamic optimal taxes in an economy in which people are heterogeneous in terms of their skills, and the government uses capital and labor income taxes to provide redistribution (insurance). The benchmark model considers an environment with permanent skills. The main theoretical results are then generalized to an environment with stochastic skills. Our approach to optimal dynamic taxation follows the recent New Dynamic Public Finance literature in the sense that taxes are allowed to be arbitrary functions of people's past and current incomes.

The key feature of our environment is equipment-skill complementarity in the production technology. Following Gravelle (2011), capital assets are grouped into two categories: structure capital and equipment capital. There are two types of labor: skilled and unskilled. Following the empirical evidence for the U.S. economy provided by Krusell, Ohanian, Ríos-Rull, and Violante (2000), we assume that the degree of complementarity between equipment capital and skilled labor is higher than the degree of complementarity between equipment capital and unskilled labor. Structure capital is neutral in terms of its complementarity with skilled and unskilled labor. More generally, Flug and Hercowitz (2000) provide evidence for equipment-skill complementarity for a large panel of countries.

[1] The capital tax differentials are created through tax depreciation allowances that differ from actual depreciation rates. Appendix A explains this in detail and provides further information on the historical evolution of capital tax differentials in the U.S. tax code.

[2] The 2011 U.S. President's State of the Union Address. Retrieved from http://www.whitehouse.gov/the-press-office/2011/01/25/remarks-president-state-union-address.

Equipment-skill complementarity implies that skilled and unskilled labor are not perfect substitutes and that the skill premium – defined as the ratio of the skilled wage to the unskilled wage – is endogenous. In particular, a decrease in the stock of equipment capital decreases the skill premium, thereby creating an indirect transfer from the skilled agents to the unskilled ones. Therefore, depressing the level of equipment capital creates an extra channel of redistribution and/or insurance. In order to depress equipment capital accumulation, the government taxes returns to equipment capital at a higher rate than it taxes returns to structure capital. This implies the optimality of differential capital taxation.

We assess the quantitative importance of differential capital taxation using the model with permanent skills calibrated to the U.S. economy. In our benchmark calibration, the optimal equipment capital income tax is 27.6 percentage points higher than the tax on structure capital in the first period. The tax differential rises along the transition path to 39.6 percentage points at the steady state.

The skill premium is about 40% in the first period after the optimal tax reform, and rises over the transition to 48% in the new steady state. Thus, the 'optimal' skill premium in any period is significantly lower than 80%, the empirical estimate for the current U.S. economy. This suggests that the optimal tax system relies much more on indirect redistribution than the current U.S. tax system. In addition, the optimal skill premium is rising over the transition because the economy is growing, and hence, the level of equipment capital increases. This result is interesting as it suggests that, even if the government cares about equality, an increasing skill premium is optimal in a growing economy.

Next, we evaluate the welfare gains of optimal differential capital taxation. This is achieved by comparing welfare in the optimal tax system with welfare in a tax system, in which the government is unrestricted in its choice of labor income taxes, but the tax rates on both types of capital are restricted to be equal to the values in the U.S. tax code. The additional welfare gains of allowing for differential capital taxation are 0.19% in terms of lifetime consumption in the benchmark and can be as high as 0.40% within the set of reasonable parameter values.

This paper focuses on the redistribution and insurance provision role of differential capital taxation. There could be other reasons for differential taxation of capital. For instance, some authors have argued that investment in equipment capital might create positive externalities. Other things being equal, positive externalities would be a reason to tax equipment capital at a lower rate than structure capital. Auerbach, Hassett, and Oliner (1994) point out, however, that it is hard to support the existence of such positive externalities on empirical grounds. This paper abstracts from all other possible reasons for differential capital taxation in order to isolate its redistributive and insurance provision role.

Related Literature. This paper relates to three distinct strands of literature. First, in their seminal paper Diamond and Mirrlees (1971) show that tax systems should maintain productive efficiency. In an environment with multiple capital types, this result implies that all capital should be taxed at the same rate. However, Auerbach (1979) and Feldstein (1990) show that it might be optimal to tax capital differentially if the government is exogenously restricted to a narrower set of fiscal instruments than in Diamond and Mirrlees (1971). Our paper is different in the sense that the optimality of differential capital taxation stems from redistribution and/or insurance motives.

Our paper follows the New Dynamic Public Finance (NDPF) tradition. This literature studies optimal capital and labor income taxation in dynamic settings in which agents' labor skills are allowed to change stochastically over time and the optimal tax system can be arbitrarily nonlinear in the history of capital and labor income.[3] No paper in this literature, however, has studied differential taxation of capital assets prior to the current paper. In addition, our paper contributes to the NDPF literature by adding to a set of recent papers that aim to provide practical policy recommendations by quantifying the theoretical implications of the NDPF literature, see e.g, Fukushima (2010), Huggett and Parra (2010), Farhi and Werning (2013), and Golosov, Troshkin, and Tsyvinski (2013).

This paper is also related to a set of theoretical studies on optimal static Mirrleesian taxation with endogenous wages. Stiglitz (1982) assumes that the labor supplies of agents with different skills are imperfect substitutes and shows that the agent with the highest income should be subsidized. Naito (1999) shows that the uniform commodity taxation result of Atkinson and Stiglitz (1976) and productive efficiency result of Diamond and Mirrlees (1971) are no longer valid under imperfect labor substitutability. Ales, Kurnaz, and Sleet (2014) analyze a static optimal tax problem in which agents with different skills are assigned to tasks (occupations). They calculate optimal taxes for the U.S. economy for the 1970s and the 2000s and compare them to their empirical counterparts. In addition, they analyze the impact of technical change on optimal taxes. The current paper differs from this literature by focusing on a dynamic environment with different types of capital, which is used to analyze optimal differential taxation of capital assets both theoretically and quantitatively.

The rest of the paper is structured as follows. Section 2 lays out the model for the case of permanent skills. Section 3 shows that differential capital taxation is optimal in this environment. Section 4 generalizes the main results to an environment with stochastic skills. Section 5 discusses our quantitative results, and Section 6 concludes.[4]

2 Model

There is a continuum of measure one of agents who live for infinitely many periods. They differ in their skill levels: they are born either skilled or unskilled, $h \in H = \{u, s\}$. A fraction π_h of agents belong to skill group h. In the main body of the paper, we assume that agents' skills are permanent. Permanent skills is a natural assumption given that in our quantitative analysis skill levels are associated with educational attainment. Section 4 shows that the main theoretical results remain valid for a general stochastic skill process.

Production Technology. An agent of skill level h produces $l \cdot z_h$ units of effective h type labor when he works l units of labor. There are two different occupational sectors in this economy: a skilled occupation in which only skilled agents are allowed to work and an unskilled occupation in which only unskilled agents are allowed to work. The first assumption reflects the fact that unskilled people do not have the skills to work in the skilled occupation.

[3] For seminal contributions to NDPF, see Golosov, Kocherlakota, and Tsyvinski (2003), Kocherlakota (2005), and Albanesi and Sleet (2006). For an excellent review of this literature, see Kocherlakota (2010).

[4] A discussion of differential taxation of capital assets in the U.S. tax code, the proofs of the propositions, a formal implementation of the constrained efficient allocation in an incomplete markets environment, and the definitions of alternative social planning problems that are analyzed in Section 5 are presented in a separate online Appendix.

The second assumption can be rationalized as follows. In our model, agents get the same disutility from working in the two occupations. Therefore, a skilled agent will choose to work in the skilled occupation as long as he gets a higher wage in the skilled occupation. This reasoning holds in the presence of taxes under our assumption that taxes are functions of income histories only. The nature of the tax system is discussed in more detail below.

Output is produced according to a production function $Y = F(K_s, K_e, L_s, L_u)$, where L_s, L_u, K_s and K_e denote the aggregate amounts of effective skilled labor, effective unskilled labor, structure capital, and equipment capital. Output can be used for consumption or can be converted to structure or equipment capital one-for-one. The economy is initially endowed with $K^*_{s,1}$ and $K^*_{e,1}$ units of the capital goods. Define $\tilde{F}$ as the function that gives the total wealth of the economy: $\tilde{F} = F + (1-\delta_s)K_s + (1-\delta_e)K_e$, where δ_s and δ_e denote the depreciation rates of structure and equipment capital. Define $F_i(\cdot)$ and $\tilde{F}_i(\cdot)$ as partial derivatives of F and $\tilde{F}$ with respect to the i^{th} argument.

Wages. Agents of type $h \in H$ receive wage $w_{h,t}$ in period t for one unit of their labor:

$$w_{s,t} = F_3(K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t}) \cdot z_s, \qquad w_{u,t} = F_4(K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t}) \cdot z_u. \tag{1}$$

Equipment-Skill Complementarity. Following Krusell, Ohanian, Ríos-Rull, and Violante (2000), we assume that the production 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 assumption has two important implications that make our model different from the standard model in the NDPF literature. First, an increase in the stock of equipment capital decreases the ratio of the marginal product of unskilled labor to the marginal product of skilled labor. In other words, the ratio of skilled to unskilled wages (skill premium) is endogenous, and this ratio 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. Second, skilled and unskilled labor are no longer perfect substitutes which implies that the skill premium is decreasing in the total amount of skilled labor and increasing in the total amount of unskilled labor. These assumptions on technology are formalized as follows.

Assumption 1. $F_3(\cdot)/F_4(\cdot)$ *is independent of* K_s.

Assumption 2. $F_3(\cdot)/F_4(\cdot)$ *is strictly increasing in* K_e.

Assumption 3. $F_3(\cdot)/F_4(\cdot)$ *is strictly decreasing in* L_s *and strictly increasing in* L_u.

Assumptions (1) - (3) are maintained throughout the paper without further reference.

Preferences. An agent of type h evaluates a consumption-labor sequence, $(c_{h,t}, l_{h,t})_{t=1}^{\infty}$, with a utility function that is time-separable and separable between consumption and labor,

$$\sum_{t=1}^{\infty} \beta^{t-1} \left[u(c_{h,t}) - v(l_{h,t}) \right],$$

where $\beta \in (0,1)$ is the discount factor, $u, v : \mathbb{R}_+ \to \mathbb{R}$, and $u', -u'', v', v'' > 0$.

Allocation. An allocation is $x = ((c_{h,t}, l_{h,t})_{h \in H}, K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t})_{t=1}^{\infty}$.

Feasibility. An allocation is feasible if in any period $t \geq 1$,

$$\sum_{h=u,s} \pi_h c_{h,t} + K_{s,t+1} + K_{e,t+1} + G_t \leq \tilde{F}(K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t}), \tag{2}$$

$$L_{h,t} = \pi_h l_{h,t} z_h, \text{ for } h \in H \quad \text{ and } \quad K_{s,1} \leq K^*_{s,1}, K_{e,1} \leq K^*_{s,1}. \tag{3}$$

Here, $\{G_t\}_{t=0}^{\infty}$ is a sequence of exogenously given wasteful government consumption.

Optimal Tax Problem. As in the U.S. tax code, taxes are allowed to depend only on people's incomes, and not directly on their skills, occupations, wages, or labor supplies. We do not model why the government does not use this information in the tax code (there could be constitutional, administrative or other reasons), but rather focus on the best policy given the existing fiscal framework. Following Mirrlees (1971) and the recent New Dynamic Public Finance literature, no further restrictions are imposed on the tax code; specifically, taxes can be arbitrarily nonlinear functions of income histories.

Following Kocherlakota (2010), we make no explicit mention of private information in motivating why taxes are restricted to depend only on income. However, the fact that the government can condition taxes only on income implies that the optimal tax problem is isomorphic to a social planning problem, in which agents are privately informed about their skills, occupations, wages, and labor supplies. Income and consumption are public information. In the planning problem, each agent reports his skill type to the planner and receives an allocation as a function of his report.[5] The set of allocations available to the planner is constrained by incentive compatibility constraints, which ensure that agents do not misreport their types.[6]

Our strategy is to first characterize the solution to the planning problem and then use this characterization to back out properties of an optimal tax system.

Incentive Compatibility. With permanent types, people report their type only once in the first period. Moreover, since agents cannot switch occupations in our model, an agent can only mimic the other type's income level by adjusting his labor hours. As a result, the planner faces only two incentive constraints.

We say that an allocation is incentive compatible if and only if for all $h \in H$

$$\sum_{t=1}^{\infty} \beta^{t-1} \left[u(c_{h,t}) - v(l_{h,t})\right] \quad \geq \quad \sum_{t=1}^{\infty} \beta^{t-1} \left[u(c_{j,t}) - v(\frac{l_{j,t} w_{j,t}}{w_{h,t}})\right], \tag{4}$$

where j denotes the complement of h in the set H.

Social Planning Problem. We analyze the problem of a planner who maximizes a

[5] Agents only report their skill types, because given that income is observable and skilled (unskilled) agents can only work in the skilled (unskilled) occupation, knowing an agent's skill type reveals all his private information.

[6] The restriction to direct truth-telling mechanisms is without loss of generality because of the following argument. Any market arrangement with taxes is a particular mechanism. By revelation principle, no such mechanism can do better than the optimal direct truth-telling mechanism. Conversely, Proposition C.1 in Appendix C shows that there is a tax system that implements the allocation that arises from the optimal direct truth-telling mechanism. Therefore, finding the optimal tax system reduces to finding the optimal direct truth-telling mechanism, which is the problem of a social planner who assigns allocations as functions of agents' types subject to incentive compatibility constraints.

Utilitarian objective with equal weights on all agents. The social planning problem is

$$\max_x \sum_{h\in H} \pi_h \sum_{t=1}^{\infty} \beta^{t-1} \left[u(c_{h,t}) - v(l_{h,t})\right] \qquad \text{s.t.} \qquad (1), (2), (3), \text{ and } (4).$$

The allocation that solves the social planning problem is called the *constrained efficient* allocation and is denoted with an asterisk throughout the paper.

3 Optimality of Differential Taxation of Capital

This section uncovers the economic mechanism that calls for differential capital taxation. We show that, with equipment-skill complementarity, as long as only the incentive constraint that prevents skilled agents from pretending to be unskilled binds, the optimal tax on equipment capital is strictly higher than the optimal tax on structure capital. Assumption 4 formalizes the assumption on the pattern of binding incentive constraints.

Assumption 4. *The incentive constraint (4) binds for $h = s$ and is slack for $h = u$ at the solution to the social planning problem.*

In all quantitative exercises in Section 5, in which the model is parameterized to match the U.S. data, the skilled wage is higher than the unskilled wage in every period. However, in our environment with endogeneous wages, it is not possible to guarantee that skilled wages are always higher than unskilled wages without making very restrictive assumptions on F. Without monotonic wages, it is not possible to determine the pattern of binding incentive constraints. Therefore, this section proceeds directly with Assumption 4, see Stiglitz (1982) for the same approach. Assumption 4 is satisfied in all our quantitative exercises.

3.1 Capital Return Wedge

In the standard growth model with two types of capital, aggregate savings are allocated between the two types of capital in a way that equates their marginal returns. Proposition 1 below shows that this is not true in the constrained efficient allocation, meaning it is optimal to create a wedge between the marginal returns to structure and equipment capital. This result forms the basis for the optimality of differential taxation of capital: to create the optimal wedge in the market equilibrium, the two types of capital should be taxed differently.

Proposition 1. *Suppose Assumption 4 holds. Then, at the constrained efficient allocation, in any period $t \geq 2$, $\tilde{F}_1(K^*_{s,t}, K^*_{e,t}, L^*_{s,t}, L^*_{u,t}) < \tilde{F}_2(K^*_{s,t}, K^*_{e,t}, L^*_{s,t}, L^*_{u,t})$.*

Proof. Let $\lambda_t \beta^{t-1}$ be the multiplier on period t feasibility constraint and μ be the multiplier on skilled agents' incentive constraint. The first order optimality conditions with respect to the two types of capital are:

$$
\begin{aligned}
(K_{e,t}) : \lambda^*_{t-1} &= \beta \left[\lambda^*_t \tilde{F}_2(K^*_{s,t}, K^*_{e,t}, L^*_{s,t}, L^*_{u,t}) + X^*_t \right], \\
(K_{s,t}) : \lambda^*_{t-1} &= \beta \lambda^*_t \tilde{F}_1(K^*_{s,t}, K^*_{e,t}, L^*_{s,t}, L^*_{u,t}), \qquad \textit{where} \\
X^*_t &= \mu^* v' \left(\frac{l^*_{u,t} w^*_{u,t}}{w^*_{s,t}} \right) l^*_{u,t} \frac{\partial \left(\frac{w^*_{u,t}}{w^*_{s,t}} \right)}{\partial K^*_{e,t}}.
\end{aligned}
$$

By equipment-skill complementarity, $\partial \left(\frac{w^*_{u,t}}{w^*_{s,t}} \right) / \partial K^*_{e,t} < 0$. Since $\mu^* > 0$, $X^*_t < 0$. Using $X^*_t < 0$ together with the first-order conditions gives the result. □

Because of equipment-skill complementarity, increasing the level of equipment capital in period t decreases the wage ratio $w^*_{u,t}/w^*_{s,t}$. This makes it more profitable for the skilled agents to pretend to be unskilled and, hence, tightens the skilled incentive constraint. From a planning perspective, this means that increasing equipment capital has an extra negative return, $X^*_t < 0$, in addition to the physical return, $\tilde{F}^*_{2,t}$, where $F^*_{i,t}$ denotes $\tilde{F}_i(K^*_{s,t}, K^*_{e,t}, L^*_{s,t}, L^*_{u,t})$. Since structure capital is neutral, changing its level does not affect the incentive constraint, and hence its only return is its physical return, $\tilde{F}^*_{1,t}$. In order for the overall return on the two types of capital to be equal, the physical return on equipment capital must higher than the physical return on structure capital at the constrained efficient allocation.

This result is intuitive: decreasing the level of equipment capital has an additional marginal benefit for the planner, because it decreases the skill premium and thus indirectly redistributes from the skilled to the unskilled. Decreasing the level of equipment capital increases its return above the return on structure capital due to diminishing marginal returns. This intuition shows that there is an extra reason to depress equipment capital accumulation relative to structure capital. This implies that equipment capital should be taxed at a higher rate than structure capital, as shown in Section 3.2.

Two features of the model are key for the optimality of the capital return wedge. First, if equipment capital was also neutral in terms of its complementarity with the two types labor, then, $X^*_t = 0$, and hence, it would be efficient to equate the physical marginal returns to the two types of capital. Second, if the government could condition taxes on skill types, it could redistribute via type-specific lump-sum taxes at zero efficiency cost and would not need the indirect (and distortionary) channel of redistribution, which works through the capital return wedge. In terms of the planning problem, this would mean that skills were not private information but publicly known. As a result, there would be no incentive constraints, and hence, $X^*_t = 0$, and the optimal capital return wedge would again be zero.

3.2 Optimal Differential Capital Taxes

This section provides a link between the optimality of the capital return wedge and the optimality of differential capital taxation. Proposition 2 characterizes the properties of optimal wedges (distortions) that a planner has to create in the intertemporal allocation of resources in order to implement the constrained efficient allocation in a competitive market environment, in which people are allowed to save through both types of capital. Formally, the optimal intertemporal wedge that the planner has to create for an agent of type h for capital of type $i \in \{s, e\}$ from period t to $t+1$ is defined as $\tau^*_{i,t+1}(h) = 1 - u'(c^*_{h,t}) / \left[\beta \tilde{F}^*_{i,t+1} u'(c^*_{h,t+1})\right]$.

Proposition 2. *Suppose Assumption 4 holds. Then,*

1. *In all periods $t \geq 2$, the optimal wedge on equipment capital is strictly positive and independent of agent type, whereas the optimal wedge on structure capital is zero, i.e., for all $h \in H$, $\tau^*_{e,t} \equiv \tau^*_{e,t}(h) > \tau^*_{s,t} \equiv \tau^*_{s,t}(h) = 0$.*

2. *If a steady state of the constrained efficient allocation exists, then the optimal wedge on equipment capital is strictly positive at the steady state.*

Proof. Relegated to Appendix B. □

Part 1 of Proposition 2 calls for zero taxation of structure capital and positive taxation of equipment capital in every period. Recall that, by Assumption 1, a change in the level of structure capital does not affect the skill premium. Therefore, there is no indirect redistribution motive to distort structure capital accumulation. In addition, it follows from the uniform commodity taxation result of Atkinson and Stiglitz (1976) that in the absence of skill risk, it is optimal not to tax structure capital.[7] In contrast, taxing equipment capital has the extra benefit of decreasing the skill premium, thus providing indirect redistribution. Therefore, the planner finds it optimal to tax equipment capital.[8] Finally, part 1 of the proposition also shows that the capital tax rates are type independent.

Part 2 of Proposition 2 says that the optimal wedge on equipment capital is positive in steady state. This result is interesting because it shows that the indirect redistribution channel calls for taxing equipment capital not only in the short run but also in the long run. This result is in contrast with the long run optimality of zero capital taxation in the Ramsey literature shown by Chamley (1986) and Judd (1985).

3.3 Intratemporal Wedges

Our model has interesting implications for intratemporal wedges (i.e., marginal labor income taxes) as well. The optimal intratemporal wedge in period t for an agent of skill type h, defined as $\tau^*_{y,t}(h) = 1 - v'(l^*_{h,t}) / \left[w^*_{h,t} u'(c^*_{h,t})\right]$, measures the efficient distortion that the planner needs to create in this agent's intratemporal allocation of consumption and labor in period t. The famous no distortion at the top result, proven originally by Sadka (1976) and Seade (1977), states that in a static Mirrleesian economy, if the distribution of skills has a finite support, then the consumption-labor decision of the agent with the highest skill level should not be distorted. Huggett and Parra (2010) prove this result for a dynamic Mirrleesian economy in which skill types are permanent and a version of our Assumption 4 holds. Proposition 3 shows that the no distortion at the top result does not hold in the presence of equipment-skill complementarity. In particular, the proposition shows that the skilled agents' labor income should be subsidized.

[7] The optimality of not taxing structure capital is closely related to Werning (2007), who shows that with permanent types zero capital taxation is optimal in a dynamic Mirrleesian model with standard Cobb-Douglas production function.

[8] If Assumption 4 is not satisfied, it will still be generically optimal to tax the two types of capital differentially, as shown explicitly in a more general environment in Section 4. However, in that case, it is not possible to determine which capital good will be taxed at a higher rate.

Proposition 3. *Suppose Assumption 4 holds. Then, in any period $t \geq 1$, the optimal intratemporal wedge of the skilled agent is negative, i.e., $\tau^*_{y,t}(s) < 0$.*

Proof. Relegated to Appendix B. □

The intuition for this result is as follows. Under the equipment-skill complementarity assumption, skilled and unskilled labor are imperfect substitutes. This implies that increasing the labor supply of the skilled agents decreases the skill premium which means that increasing skilled labor supply creates indirect redistribution. In order to encourage the supply of skilled labor, the government finds it optimal to subsidize skilled labor at the margin. This result is in line with Stiglitz (1982), who shows that when two types of labor are imperfect substitutes, the more productive agents' labor supply should be subsidized.

4 Generalization to Stochastic Skills

In the model laid out in Section 2, agents' skill types are permanent. The current section allows for agents' skills to evolve stochastically over time. This level of generality is useful because it allows us to establish that the main theoretical results of Section 3 remain valid if people's skills change after they enter the labor market, or if one takes a dynastic interpretation of our model in which skills change from one generation to another. Notice that in this environment with stochastic skills the government uses taxes to provide insurance in addition to providing redistribution and financing public spending.

We first show that differential taxation of capital is optimal for any stochastic skill process. Moreover, under an assumption regarding the pattern of binding incentive compatibility conditions, it is optimal to tax equipment capital at a higher rate than structure capital.

The environment is the same as in Section 2 except that people are born identical, but their skills evolve stochastically over time. A skill realization in period t is denoted by $h_t \in H$. A partial skill history in period t is denoted by $h^t = (h_1, h_2, \ldots, h_t) \in H^t$, where H^t denotes the set of all period t histories. Let $\pi_t(h^t)$ be the unconditional probability of h^t.

Wages. An agent of type h in period t receives a wage $w_{h,t}$, defined in equation (1), independent of his skill history before period t. For expositional convenience, in this section, wages are denoted by $w_t(h_t)$ instead of $w_{h,t}$.

Preferences. Preferences are now defined over stochastic processes of consumption and labor, $(c_t, l_t)_{t=0}^{\infty}$, where $c_t, l_t : H^t \to \mathbb{R}_+$, using an ex ante expected discounted utility function,

$$\sum_{t=1}^{\infty} \sum_{h^t \in H^t} \pi_t(h^t) \beta^{t-1} \left[u(c_t(h^t)) - v(l_t(h^t)) \right]. \tag{5}$$

Allocation. An allocation is $x = (c_t, l_t, K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t})_{t=1}^{\infty}$.

Feasibility. An allocation is feasible if in any period $t \geq 1$,

$$\sum_{h^t \in H^t} \pi_t(h^t) c_t(h^t) + K_{s,t+1} + K_{e,t+1} + G_t \leq \tilde{F}(K_{s,t}, K_{e,t}, L_{s,t}, L_{u,t}), \tag{6}$$

$$L_{h,t} = \sum_{\{h^t \in H^t | h_t = h\}} \pi_t(h^t) l_t(h^t) z_h \text{ for } h \in H, \qquad \text{and} \qquad K_{s,1} \leq K^*_{s,1}, K_{e,1} \leq K^*_{s,1}. \tag{7}$$

Incentive Compatibility. Define $\sigma_t : H^t \to H$. A reporting strategy is $\sigma = (\sigma_t)_{t=1}^{\infty}$. Let Σ denote the set of all reporting strategies. The truth-telling strategy, which is denoted by σ^*, prescribes reporting the true type at each and every node: for all h^t, $\sigma_t^*(h^t) = h_t$. Let $\sigma^t(h^t) = (\sigma_1(h^1), ..., \sigma_t(h^t))$ denote the history of reports along history h^t. Define

$$W(\sigma|x) = \sum_{t=1}^{\infty} \sum_{h^t \in H^t} \pi_t(h^t)\beta^{t-1} \left[u(c_t(\sigma^t\left(h^t\right))) - v\left(\frac{l_t(\sigma^t\left(h^t\right))w_t(\sigma_t\left(h^t\right))}{w_t(h_t)} \right) \right],$$

as the expected discounted value of using reporting strategy σ given an allocation x. An allocation x is called incentive compatible if and only if for all $\sigma \in \Sigma$, $W(\sigma^*|x) \geq W(\sigma|x)$.

Following Fernandes and Phelan (2000), without loss of generality, we restrict attention to the set of reporting strategies that has lying only at a single node. This allows us to replace the incentive compatibility constraints defined above with a sequence of *temporary incentive constraints*, one for each node. An allocation x is called temporary incentive compatible if and only if, in any period t and at any node h^{t-1} and for all $h_t \in H$,

$$\begin{aligned} & u(c_t(h^{t-1}, h_t)) - v(l_t(h^{t-1}, h_t)) + \sum_{m=t+1}^{\infty} \sum_{h^m \in \bar{H}^m} \pi_m(h^m)\beta^{m-t} \left[u(c_m(h^m)) - v(l_m(h^m)) \right] \qquad (8) \\ & \geq u(c_t(h^{t-1}, h_t^o)) - v\left(\frac{l_t(h^{t-1}, h_t^o)w_t(h_t^o)}{w_t(h_t)} \right) + \sum_{m=t+1}^{\infty} \sum_{h^m \in \bar{H}^m} \pi_m(h^m)\beta^{m-t} \left[u(c_m(\tilde{h}^m)) - v(l_m(\tilde{h}^m)) \right], \end{aligned}$$

where h_t^o is the complement of h_t in the set H, $\bar{H}^m$ denotes the set of period m histories that follow from h^t, i.e., $\bar{H}^m \equiv \{h^m \in H^m : h^m \succ h^t\}$, and $\tilde{h}^m = (h^{t-1}, h_t^o, h_{t+1}, ..., h_m)$ is identical to h^m except in period t. From now on, (8) is used to represent incentive compatibility.[9]

Social Planning Problem. The social planning problem that defines the constrained efficient allocation is: $\max_x$ (5) s.t. (1), (6), (7), and (8).

Optimality of Differential Capital Taxation. Now, we prove the optimality of differential taxation of capital for the general environment with skill shocks. First, define the intertemporal wedge for an agent with skill history h^t and for capital of type $i \in \{s, e\}$ from period t to period $t+1$, as

$$\tau_{i,t+1}(h^t) = 1 - \frac{u'(c_t(h^t))}{\beta \tilde{F}_{i,t+1} E_t \left\{ u'(c_{t+1}(h^{t+1})) | h^t \right\}}. \qquad (9)$$

The first part of Proposition 4 generalizes Proposition 1 by showing that it is optimal to create a wedge between the marginal returns to structure and equipment capital when skills evolve stochastically over time. The second part of Proposition 4 shows that the optimal intertemporal wedges for structure and equipment capital are different. Thus, optimality of differential taxation of capital does not depend on the permanent skill type assumption.

[9] Temporary incentive constraints were first shown to be necessary and sufficient for incentive compatibility by Green (1987) for an environment with i.i.d. shocks. Fernandes and Phelan (2000) generalized this result to environments with persistent shocks. To be precise, two more assumptions are needed to guarantee the necessity and sufficiency of temporary incentive constraints. First, each skill history should be reached with strictly positive probability. Second, a transversality condition, which is automatically satisfied if one assumes that instantaneous utility is bounded, should hold.

Proposition 4. *1. At the constrained efficient allocation, in any period $t \geq 2$,*

$$\tilde{F}_1(K^*_{s,t}, K^*_{e,t}, L^*_{s,t}, L^*_{u,t}) = \tilde{F}_2(K^*_{s,t}, K^*_{e,t}, L^*_{s,t}, L^*_{u,t}) + X^*_t/\lambda^*_t, \qquad \textit{where}$$

$$X^*_t = \sum_{\{h^t \in H^t\}} \mu^*_t(h^t) v'\left(\frac{l^*_t(h^{t-1}, h^o_t) w^*_t(h^o_t)}{w^*_t(h_t)}\right) l^*_t(h^{t-1}, h^o_t) \frac{\partial \frac{w^*_t(h^o_t)}{w^*_t(h_t)}}{\partial K^*_{e,t}}$$

and $\lambda_t \beta^{t-1}$ and $\mu_t(h^t)$ are Lagrange multipliers on period t feasibility constraint and the incentive constraint at history h^t.

2. (a) The optimal wedge on structure capital in any period $t \geq 2$ and history h^{t-1} satisfies $\tau^*_{s,t}(h^{t-1}) \geq 0$. The inequality is strict if and only if there is no $h \in H$ such that $\pi_t(h^{t-1}, h|h^{t-1}) = 1$.

(b) The optimal wedge on equipment capital in any period $t \geq 2$ and history h^{t-1} is

$$\left[1 - \tau^*_{e,t}(h^{t-1})\right] = \left[1 - \tau^*_{s,t}(h^{t-1})\right] \cdot \left[1 + X^*_t / \left(\lambda^*_t \tilde{F}^*_{2,t}\right)\right]. \tag{10}$$

Proof. Relegated to Appendix B. □

The idea behind the first part of Proposition 4 is very similar to the one for Proposition 1: under equipment-skill complementarity, increasing the amount of equipment capital has an effect on incentives, summarized by the term X^*_t. In contrast, changing the amount of structure capital does not affect incentives. As a result, it is optimal to create a wedge between the physical returns to the two types of capital. The main distinction from the permanent type model is that, in the case with stochastic skills, a change in period t equipment capital affects all the binding incentive constraints in that period. Thus, X^*_t measures the cumulative effect of a change in equipment capital on all the binding incentive constraints. Since at this level of generality it is not possible to determine the pattern of binding incentive constraints, the sign of X^*_t is ambiguous.

Part 2(a) of Proposition 4 states that the intertemporal wedge on structure capital is positive if there is skill risk. Intuitively, if an agent is allowed to save at the marginal rate of return to structure capital, he will save more than the efficient level. In the next period, he will work less than socially optimal if he turns out to be of the skilled type. To prevent this double deviation, it is optimal to discourage savings. The government achieves that with a positive wedge on structure capital.[10] Naturally, with permanent types there is no skill risk and, hence, no reason to tax structure capital, as already shown in Proposition 2.

Equation (10) in part 2(b) of the proposition is a version of the no-arbitrage condition for this economy. The equation shows that the intertemporal wedge on equipment capital can be decomposed into two parts. First, the government wants to discourage savings in equipment capital for the same reason that it wants to discourage savings in structure capital, which is captured by the first term on the right-hand side of equation (10). The second term on the

[10]The positive wedge on structure capital follows from the inverse Euler equation, see equation (B.6) in Appendix B. This condition was first derived by Rogerson (1985) and then generalized by Golosov, Kocherlakota, and Tsyvinski (2003). The inverse Euler equation does not hold for equipment capital because of the effect that equipment capital has on incentives. We derive a modified version of the inverse Euler equation for equipment capital in Appendix B, see equation (B.7).

right-hand side of equation (10) is present in order to create the optimal wedge between the returns to the two types of capital. The presence of this term implies that generically the optimal wedges on the two types of capital are *different* in any period and history, which establishes the optimality of differential taxation of capital.

A Special Case. Assumption 5 below assumes that the only incentive constraints that bind are those that prevent the skilled from pretending to be unskilled. These incentive constraints are called *downward incentive constraints*. There is no theoretical result that establishes the pattern of binding incentive constraints for general skill processes in dynamic Mirrleesian environments, even when wages are exogeneous.[11] Indeed, there are examples in which some upward incentive constraints bind. In this regard, Assumption 5 is stronger than Assumption 4, which is used in Section 3.

Assumption 5. *In any period $t \geq 1$, history h^t, only downward incentive constraints bind.*

Assumption 5 allows us to show that $X_t^* > 0$ in all periods. It is then possible to sign the capital return wedge, and show that the optimal equipment capital wedge is higher than the optimal structure capital wedge. These results are summarized by the following proposition.

Proposition 5. *Suppose Assumption 5 holds. Then, in any period $t \geq 2$ and history h^{t-1}, $\tilde{F}_1(K_{s,t}^*, K_{e,t}^*, L_{s,t}^*, L_{u,t}^*) < \tilde{F}_2(K_{s,t}^*, K_{e,t}^*, L_{s,t}^*, L_{u,t}^*)$ and $\tau_{e,t}^*(h^{t-1}) > \tau_{s,t}^*(h^{t-1})$.*

Proof. Relegated to Appendix B. □

Intratemporal Wedges. Under Assumption 5, Proposition 6 generalizes the optimality of subsidizing skilled labor supply, shown for the permanent type case in Section 3.3, for the environment in which skills evolve stochastically over time. First, define the optimal intratemporal wedge at history h^t as $\tau_{y,t}^*(h^t) = 1 - v'(l_t^*(h^t))/(w_t^*(h_t)u'(c_t^*(h^t)))$.

Proposition 6. *Suppose Assumption 5 holds. In any period $t \geq 1$ and history h^{t-1}, $\tau_{y,t}^*(h^{t-1}, s) < 0$.*

Proof. Relegated to Appendix B. □

Implementation. Appendix C provides an implementation of the constrained efficient allocation through a tax system in a competitive market environment in which agents trade a risk free bond and capital. The implementation result holds for any stochastic process, including the permanent type model. An interesting feature of this tax system is that the optimal tax differentials across equipment and structure capital can be implemented at the firm level, as is the case in the current U.S. tax system. This is possible because, as the second term on the right-hand side of equation (10) shows, the differential between optimal intertemporal wedges of structure and equipment capital is history independent in any period. Another notable feature of the implementation is that the optimal tax system mimics the actual U.S. tax code in the sense that capital tax differentials are created through depreciation allowances that differ from actual economic depreciation. Therefore, creating the optimal capital tax differentials would not require complicating the U.S. tax code further.

[11] Downward incentive constraints are the only binding incentive constraints when skills are i.i.d. and wages are exogeneous.

5 Quantitative Analysis

The main goal of this section is to analyze the quantitative importance of differential taxation of capital in a calibrated version of our model. As in the main part of the paper, agents' skill types are assumed to be permanent. Since there is no labor income risk in this environment, the only role of taxation is redistribution (along with financing government consumption). Permanent skills is a natural assumption given that in the data we associate skill levels with educational attainment. In addition, there is empirical evidence that initial conditions account for a large part of the cross-sectional variation in lifetime earnings.[12]

First, model parameters are calibrated to the U.S. economy using a competitive equilibrium framework with the actual U.S. tax code and government consumption level. Then, we solve a social planning problem with endogeneous factor prices in which the planner "inherits" the initial capital stocks from the steady state of the competitive equilibrium.[13] We solve for the whole time series of the constrained efficient allocation, thus taking into account the transition to a new steady state, and recover the optimal wedges (taxes) from the constrained efficient allocation. In line with Proposition 2, the optimal taxes on equipment capital are higher than those on structure capital. Specifically, in our benchmark calibration, the optimal tax differential increases from 27.6% in the first period to 39.5% in the steady state. Moreover, the welfare gains of optimal differential capital taxation can be as high as 0.4% in terms of lifetime consumption.

5.1 Calibration

To calibrate the parameters in the social planning problem, we assume that the steady state of the competitive equilibrium (abbreviated as SCE in what follows) defined in Appendix C represents the current U.S. economy. We first fix a number of parameters to values from the data or from the literature and then calibrate the remaining parameters so that the SCE matches the U.S. data along selected dimensions.

One period in our model corresponds to one year. The period utility function takes the form $u(c) - v(l) = c^{1-\sigma}/(1-\sigma) - \phi l^{1+\gamma}/(1+\gamma)$. In the benchmark case, $\sigma = 2$ and $\gamma = 1$. These are within the range of values that have been considered in the literature. The production function takes the same form as in Krusell, Ohanian, Ríos-Rull, and Violante (2000):

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

The values of α, ρ, η are taken from Krusell, Ohanian, Ríos-Rull, and Violante (2000) who

[12] Keane and Wolpin (1997) estimate that initial conditions account for 90% of the cross-sectional variation in life-time earnings. Huggett, Ventura, and Yaron (2011) estimate this number to be over 60%, and Storesletten, Telmer, and Yaron (2004) estimate it to be almost 50%.

[13] It would not be possible to assess the role of differential capital taxation in a partial equilibrium environment, because the skill premium would not be affected by changes in the level of equipment capital. To the contrary, most quantitative papers in the NDPF literature consider partial equilibrium environments. As Farhi and Werning (2012) show, considering general equilibrium effects might be important *even* with a standard production function without complementarities.

estimate these parameters using U.S. data. ω and ρ (which Krusell, Ohanian, Ríos-Rull, and Violante (2000) do not estimate) are calibrated to U.S. data, as explained in detail below. This production function satisfies Assumptions 1 – 3 if $\eta > \rho$, which is what Krusell, Ohanian, Ríos-Rull, and Violante (2000) find.

The government consumption-to-output ratio is assumed to be 16%, which is close to the average ratio in the United States during the period 1980 – 2012, as reported in the National Income and Product Accounts (NIPA) data. Following Heathcote, Storesletten, and Violante (2010), we assume a flat labor income tax rate of $\tau_y = 27\%$ (for a discussion of the construction of this number, see Domeij and Heathcote (2004)). Gravelle (2011) documents that because of differences in tax depreciation rates, the effective tax rates on structure capital and equipment capital differ at the firm level. Specifically, she estimates the effective corporate tax rate on structure capital to be 32%, and that on equipment capital to be 26%. The capital income tax rate at the consumer level is 15% in the U.S. tax code. This implies an overall tax on structure capital $\tau_s = 1 - 0.85 \cdot (1 - 0.32) = 42.2\%$ and an overall tax on equipment capital $\tau_e = 1 - 0.85 \cdot (1 - 0.26) = 37.1\%$. These numbers are in line with a 40% tax on aggregate capital that is reported by Domeij and Heathcote (2004). Unspent government tax revenue is distributed back to the agents in a lump-sum manner, which implies that in the SCE average taxes are in general not equal to marginal taxes. The ratio of skilled to unskilled agents, π_s/π_u, is set so as to be consistent with the 2011 US Census data. As in Section 2, π_s refers to the fraction of skilled agents and π_u refers to the fraction of unskilled agents.

For a given tax system, steady-state equilibrium is not unique in our environment with permanent types. In particular, in the absence of idiosyncratic uncertainty, depending on the initial asset distribution across skill groups, there are many steady-state equilibrium asset distributions. To calibrate the model, we select the steady-state equilibrium which matches the distribution of assets between skilled and unskilled agents observed in the U.S. data. Formally, denote the steady-state asset holdings of a skilled agent by a_s and of an unskilled agent by a_u. Given aggregate capital levels K_s, K_e consistent with the SCE, any asset distribution of the form $\pi_s a_s = \zeta(K_s + K_e)$ and $\pi_u a_u = (1-\zeta)(K_s + K_e)$ with $\zeta \in (0, 1)$ can arise in the SCE. This means that skilled agents hold fraction ζ of aggregate wealth and unskilled agents hold fraction $(1 - \zeta)$ of aggregate wealth. ζ is chosen so that the SCE asset distribution matches the observed asset distribution between skilled and unskilled agents in the 2010 U.S. Census data. Table 1 summarizes the benchmark parameters that are taken directly from the data or the literature.

[Table 1 about here.]

This leaves us with several parameters to be determined. z_u and z_s cannot be identified separately from the remaining parameters of the production function, and therefore, are set to $z_u = z_s = 1$. The parameter that controls the income share of equipment capital ω, the parameter that controls the income share of unskilled labor ν, the labor disutility parameter ϕ, and the discount factor β are calibrated. These parameters are calibrated so that (i) the labor share equals 2/3 (approximately the average labor share in 1980 – 2010 as reported in the NIPA data), (ii) the capital-to-output ratio equals 2.9 (approximately the average of 1980 – 2011 as reported in the NIPA and Fixed Asset Tables), (iii) the skill premium equals 1.8 (as reported by Heathcote, Perri, and Violante (2010) for the 2000s), and (iv) the aggregate labor supply in steady state equals 1/3 (as is commonly used in the macro literature). Table 2 summarizes the calibration procedure.

[Table 2 about here.]

5.2 Quantitative Results

This section analyzes the quantitative properties of the optimal tax system. This is achieved by solving the social planning problem (SPP) defined in Section 2 with parameters calibrated in Section 5.1 to the U.S. economy.[14] In the SPP, the planner inherits the initial capital stocks from the SCE and needs to finance the same level of government consumption as in the SCE.

Steady-State Comparison. We first discuss the properties of the optimal tax system in steady state and compare it to the current U.S. tax system. The first column of Table 3 summarizes the current U.S. tax system. The second column reports its counterpart in the optimal tax system at the steady state. The first two rows of Table 3 report capital income taxes net of depreciation.[15] The equipment capital tax τ_e is substantial at the steady state of the solution to the SPP. It is 39.54% – that is, 39.54 percentage points higher than the tax on structure capital τ_s, which is zero. This is in contrast with the current effective tax rates in the United States where structure capital is taxed by 5.1 percentage points more than equipment capital overall. As for the labor wedges, they are 27% for both types of labor in the SCE because we approximate the U.S. labor income tax code by a 27% linear tax. At the steady state of the solution to the SPP, the labor wedge for unskilled labor $\tau_y(u)$ is 26.6%, which is almost the same as in the SCE. The skilled labor wedge $\tau_y(s)$, on the other hand, is -11.14%. Both higher taxes on equipment capital and marginal subsidies on skilled labor are in line with our theoretical results from Section 3.

[Table 3 about here.]

The higher taxes on equipment capital relative to structure capital, together with marginal subsidies on skilled labor, are used to indirectly redistribute from the skilled to the unskilled. Table 4 shows how the optimal tax system achieves indirect redistribution by comparing the allocations at the SCE and the SPP. The higher tax on equipment capital discourages the accumulation of equipment capital relative to structure capital at the SPP in comparison to the SCE. At the same time, the marginal subsidy on skilled labor income increases the ratio of skilled to unskilled labor. Both capital and labor taxes decrease the skill premium at the SPP. This way, the planner provides indirect redistribution from the skilled to the unskilled.

[14] The SPP is solved assuming that the economy converges to a steady state in 200 periods. Changing the number of periods does not affect the results. In other words, the economy gets very close to steady state long before period 200.

[15] Table 3 reports capital income taxes net of depreciation rather than the capital wedges defined in Section 3.2 because the former correspond to the taxes used in the U.S. tax code. With a slight abuse of notation, τ_i, which refers to capital wedge for capital of type i in the rest of the paper, refers to capital income tax net of depreciation in this section. In the column denoted "SPP," the capital income taxes are recovered from the constrained efficient allocation by using the following definition for each skill type $h \in H$, capital type i, and period t: $\tau_{i,t+1}(h) \equiv 1 - \left(\frac{u'(c_{h,t})}{\beta u'(c_{h,t+1})} - 1 \right) / (F_{i,t+1} - \delta_i)$. Part 1 of Proposition 2 implies that these taxes only depend on time and not on agent type; therefore, only one number (time series) is reported.

[Table 4 about here.]

The marginal subsidy on skilled labor income *seems* to imply that there is direct redistribution from the unskilled to the skilled at the SPP. However, recall that optimal taxes are nonlinear in labor income. In this case, at a given income level, the average income tax can be quite different from the marginal income tax.[16] As a consequence, a tax system can be progressive overall even though the marginal taxes are regressive. This is precisely what happens at the optimal tax system. To assess the overall progressivity of the optimal tax system, we compute a measure of average labor taxes that an agent has to pay at the steady state of the SPP. This measure is defined as $1 - c_h/(w_h l_h)$ for agents of type h, following Farhi and Werning (2013). The optimal average labor taxes computed using this measure are progressive: 6% for the unskilled and 18% for the skilled. Therefore, the optimal labor taxes do provide direct redistribution from the skilled to the unskilled.[17]

Transition. This section summarizes the evolution of the optimal taxes (wedges) along the transition to the new steady state. The left panel of Figure 1 shows that the optimal structure capital income tax (net of depreciation) is 0 and the equipment capital tax is positive in all periods. These properties are in line with Proposition 2. The equipment capital tax is growing over time. To understand this finding, one needs to look at the evolution of the stocks of the two types of capital, which is shown in the left panel of Figure 2. It shows that both capital stocks are growing along the transition path. The overall capital stock is growing in the constrained efficient allocation because the planner inherits an inefficiently low level of capital from the SCE, which is due to the inefficiently high overall level of capital taxes at the SCE. As the quantity of equipment capital grows, so does the skill premium (see Figure 3). The planner wants to prevent an unfettered growth of the skill premium because of its adverse redistributive effects. To keep the growth of the skill premium under control, the planner finds it optimal to increase the tax on equipment capital.[18]

Optimal labor wedges are almost constant along the transition, as shown in the right panel of Figure 1. In fact, Werning (2007) shows that with our utility function labor wedges are exactly constant over time in a permanent-type model without equipment-skill complementarity. Figure 1 suggests that the extra distortions in labor wedges arising from equipment-skill complementarity are also approximately constant over time. Since skilled labor is subsidized, skilled agents work more than unskilled agents in each period, as shown in the right panel of Figure 2. As the economy grows, both types of agents become richer, and because of the income effect, they decrease their labor supply even though labor wedges

[16]Suppose, e.g., that the tax formula for an agent with income \$200,000 is $T(y) = \$100,000 - 0.1 \cdot y$. This agent pays \$80,000 in taxes, implying an average tax of 40%, even though he gets a marginal subsidy of 10%.

[17]The non-linear nature of the optimal labor income tax code also explains how government budget is balanced under the optimal tax system. Table 3 seems to suggest that - except for a small increase in equipment capital taxes - government revenue from all other sources declines significantly when the economy moves from the current system to the optimal one. However, with a non-linear tax system the total amount of labor income taxes collected can increase even if the marginal taxes decline.

[18]We check the validity of this intuition by conducting exercises, in which the planner inherits inefficiently high amounts of capital from the SCE. In those cases, as our intuition suggests, the planner decreases both types of capital over the transition to the new steady state, and optimal equipment taxes decline over the transition period.

do not change much.

Figure 3 depicts the evolution of the optimal skill premium over time. First, the optimal skill premium is much lower in each period than it is in the U.S. data. This result suggests that the current U.S. tax system does not generate enough indirect redistribution. Second, the skill premium is increasing over time because the equipment capital level increases. This result implies that an increasing skill premium is optimal in a growing economy, even if the government cares about equality.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

Welfare Gains of Optimal Differential Taxation of Capital. The importance of optimal differential taxation of capital is evaluated by answering the following question: how much of the welfare gains of the full reform (which is called optimal DTC in this section) is lost if the government is restricted to use the current capital taxes and is allowed to choose only the labor income taxes optimally? To answer this question, we solve an additional version of the planning problem. In this problem, the planner is unrestricted in his choice of labor taxes, but he must use the capital income taxes as in the U.S. tax code. This tax system is called *current differential taxation of capital* (current DTC). The planning problem that gives rise to the current DTC is stated in Appendix D. For the benchmark parameters, reforming the current tax system to the optimal DTC implies 0.19% more welfare gains than reforming labor taxes alone (i.e., moving to the current DTC).[19] The additional gains of optimal DTC can be as high as 0.40% for reasonable parameter values, as discussed in more detail in the sensitivity analysis below.

In addition, we solve a version of the social planning problem, in which the planner is unrestricted in his choice of labor taxes, but is not allowed to tax the two types of capital differentially. This tax system is called the optimal *nondifferential taxation of capital* (optimal NDTC). The planning problem that gives rise to the optimal NDTC is stated in Appendix D. The welfare gains of the current DTC fall 0.14% short of the welfare gains of the optimal NDTC for the benchmark parameters. This difference in welfare gains can be as high as 0.27% for reasonable parameter values.[20]

One can also assess how people rank the different capital tax reforms. Relative to the current DTC, the optimal DTC helps both types. The reason is that the overall level of capital taxes at the current DTC is inefficiently high. Under the optimal DTC, structure capital taxes are zero while equipment capital taxes are virtually unchanged. As a result, there is more capital of both types at the optimal DTC. This increases the productivity of

[19] The welfare gains of allocation x relative to allocation y are measured as a fraction by which consumption in allocation y has to be increased in each date and state to make its welfare equal to allocation x welfare.

[20] These results suggest that setting capital tax rates to a uniform rate, as proposed recently by President Obama's administration, might imply substantial welfare gains. However, our results here are only suggestive, since that proposal only involves reforming capital taxes, but would leave labor taxes intact. Slavik and Yazici (2014) evaluate the consequences of such a proposal in a world with multiple layers of heterogeneity across agents.

both types of agents, and they both benefit from this reform. In addition, relative to the current DTC, the optimal NDTC helps the skilled and hurts the unskilled.

Sensitivity Analysis. Each sensitivity exercise changes the parameter of interest and redoes the calibration procedure. Table 5 summarizes the sensitivity results. In this table, optimal taxes are only reported for the optimal DTC reform. With a higher σ, the curvature of utility from consumption, the planner wants to provide more redistribution. Therefore, the indirect redistribution channel becomes more important. Hence, as σ increases, the tax on equipment capital as well as the marginal subsidy to skilled labor increase. Table 5 also reports the sensitivity of our results to changes in γ, the curvature of disutility from labor. As γ decreases, the tax on equipment capital and the skilled labor subsidy increase.

[Table 5 about here.]

As the penultimate row of Table 5 reports, the welfare gains of the optimal DTC reform are around 0.20% higher than the gains of the current DTC reform for all values of σ and γ considered.[21] The welfare gains of optimal NDTC relative to current DTC are decreasing in σ and increasing in γ, as shown in the last row of Table 5. The reason is that with a larger σ or lower γ, the optimal capital tax differential is larger, as one can see in the rows denoted by τ_e and τ_s in Table 5. Therefore, optimal NDTC, which forces capital taxes to be uniform, is more restrictive and implies smaller welfare gains for higher σ or lower γ.

The welfare gains of optimal DTC relative to current DTC are as high as 0.28% for $\sigma = 4$ and $\gamma = 0.5$. He and Liu (2008) use a higher elasticity of substitution between equipment capital and unskilled labor, namely, $\eta = 0.79$, which is based on an empirical estimate by Duffy, Papageorgiou, and Perez-Sebastian (2004). For this value of η and $\sigma = 4$ and $\gamma = 0.5$, the welfare gains of optimal DTC relative to current DTC are 0.40%.

6 Conclusion

The effective marginal tax rates on returns to capital assets differ substantially depending on the capital asset type in the U.S. tax code. In particular, the marginal tax rate on capital structures is about 5% higher than the marginal tax rate on capital equipments. This paper assesses the optimality of differential capital asset taxation both theoretically and quantitatively from the perspective of a government whose aim is to provide redistribution and insurance. Contrary to the actual practice in the U.S. tax code, the paper shows that, under a plausible assumption, it is optimal to tax equipment capital at a higher rate than structure capital. Intuitively, in an environment with equipment-skill complementarity, taxing equipment capital and hence depressing its accumulation decreases the skill premium, providing indirect redistribution from the skilled to the unskilled agents. In a quantitative version of the model, the optimal tax rate on equipment capital is at least 27 percentage points higher than the optimal tax rate on structure capital during transition and at the

[21] We also compute the welfare gains of optimal DTC under alternative social welfare weights. If the planner cares more about the unskilled, the welfare gains of optimal DTC are larger. This is intuitive: not being able to use one of the channels of indirect redistribution optimally has more severe welfare consequences when society care more about redistribution.

steady state. Furthermore, the welfare gains of optimal differential capital taxation can be as high as 0.4% of lifetime consumption.

7 Acknowledgements

We thank the associate editor, Christopher Sleet, an anonymous referee, as well as Laurence Ales, Alex Bick, Felix Bierbrauer, V. V. Chari, Nicola Fuchs-Schündeln, Kenichi Fukushima, Mike Golosov, Christian Hellwig, Larry Jones, Juan Pablo Nicolini, Fabrizio Perri, Dominik Sachs, Florian Scheuer, Yuichiro Waki, and participants at various conferences, workshops and seminars for their comments. Yazici greatly appreciates IIES (Stockholm University) for their hospitality during his visit.

References

Albanesi, S., and C. Sleet (2006): "Dynamic Optimal Taxation with Private Information," *Review of Economic Studies*, 73(1), 1–30.

Ales, L., M. Kurnaz, and C. Sleet (2014): "Tasks, Talents, and Taxes," Working paper.

Atkinson, A. B., and J. E. Stiglitz (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6(1–2), 55–75.

Auerbach, A. J. (1979): "The Optimal Taxation of Heterogeneous Capital," *Quarterly Journal of Economics*, 93(4), 589–612.

Auerbach, A. J., K. A. Hassett, and S. D. Oliner (1994): "Reassessing the Social Returns to Equipment Investment," *Quarterly Journal of Economics*, 109(3), 789–802.

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

Diamond, P. A., and J. A. Mirrlees (1971): "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review*, 61(1), 8–27.

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.

Farhi, E., and I. Werning (2012): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," *Journal of Political Economy*, 120(3), 398–445.

Farhi, E., and I. Werning (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596–635.

Feldstein, M. (1990): "The Second Best Theory of Differential Capital Taxation," *Oxford Economic Papers*, 42(1), 256–267.

Fernandes, A., and C. Phelan (2000): "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, 91(2), 223–247.

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

Fukushima, K. (2010): "Quantifying the Welfare Gains From Flexible Dynamic Income Tax Systems," Working paper.

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

Golosov, M., M. Troshkin, and A. Tsyvinski (2013): "Redistribution and Social Insurance," Working Paper 17642.

Gravelle, J. G. (2011): "Reducing Depreciation Allowances to Finance a Lower Corporate Tax Rate," *National Tax Journal*, 64(4), 1039–1054.

Green, E. (1987): "Lending and the Smoothing of Uninsurable Income," in *Contractual Arrangements for Intertemporal Trade*, ed. by E. C. Prescott, and N. Wallace. Minneapolis: University of Minnesota Press.

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

He, H., and Z. Liu (2008): "Investment-Specific Technological Change, Skill Accumulation, and Wage Inequality," *Review of Economic Dynamics*, 11(2), 314–334.

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.

Huggett, M., and J. C. Parra (2010): "How Well Does the U.S. Social Insurance System Provide Social Insurance?," *Journal of Political Economy*, 118(1), 76–112.

Huggett, M., G. Ventura, and A. Yaron (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923–2954.

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

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

Kocherlakota, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73(5), 1587–1621.

——— (2010): *The New Dynamic Public Finance*. Princeton University Press.

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.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(114), 175–208.

Naito, H. (1999): "Re-examination of Uniform Commodity Taxes under a Non-linear Income Tax System and its Implication for Production Efficiency," *Journal of Public Economics*, 71(2), 165–188.

Rogerson, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

Sadka, E. (1976): "On Income Distribution, Incentive Effects and Optimal Income Taxation," *Review of Economic Studies*, 43(2), 261–267.

Seade, J. K. (1977): "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, 7(2), 203–235.

Slavik, C., and H. Yazici (2014): "On the Consequences of Eliminating Capital Tax Differentials," Working paper.

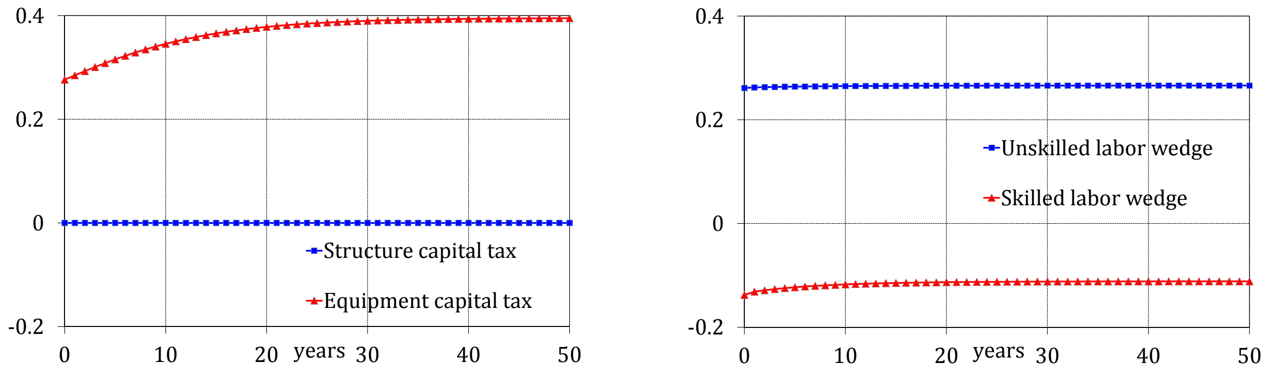
Stiglitz, J. E. (1982): "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 17(2), 213–240.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Consumption and Risk Sharing over the Life Cycle," *Journal of Monetary Economics*, 51(3), 609–633.

Werning, I. (2007): "Optimal Fiscal Policy with Redistribution," *Quarterly Journal of Economics*, 122(3), 925–967.

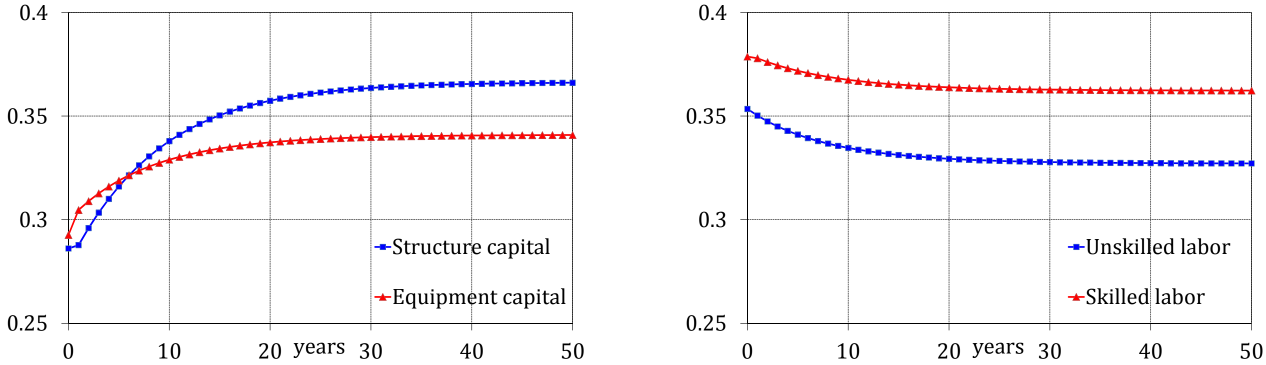
List of Figures

Figure 1: Optimal Taxes/Wedges at the Constrained Efficient Allocation



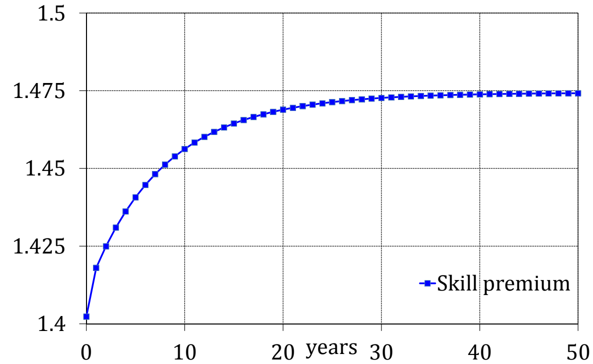
This figure shows the paths of optimal taxes (wedges) along the transition to the new steady state at the solution to the social planning problem.

Figure 2: Factors of Production at the Constrained Efficient Allocation



This figure shows the paths of factors of production along the transition to the new steady state at the solution to the social planning problem.

Figure 3: Skill Premium at the Constrained Efficient Allocation



This figure shows the path of the skill premium w_s/w_u along the transition to the new steady state at the solution to the social planning problem.

List of Tables

Table 1: Benchmark Parameters

Parameter	Symbol	Value	Source
Relative risk aversion	σ	2	
Inverse Frisch elasticity	γ	1	
Structure capital depreciation rate	δ_s	0.056	GHK
Equipment capital depreciation rate	δ_e	0.124	GHK
Share of structure capital in output	α	0.117	KORV
Measure of elasticity of substitution between equipment capital K_e and unskilled labor L_u	η	0.401	KORV
Measure of elasticity of substitution between equipment capital K_e and skilled labor L_s	ρ	-0.495	KORV
Tax on labor income	τ_y	0.27	HSV
Overall tax on structure capital income	τ_s	0.422	Gravelle (2011)
Overall tax on equipment capital income	τ_e	0.371	Gravelle (2011)
Government consumption	G/Y	0.16	NIPA
Relative supply of skilled workers	π_s/π_u	0.778	U.S. Census
Share of skilled workers' wealth	ζ	0.686	U.S. Census

This table reports the benchmark parameters that are taken directly from the data or the literature. The acronyms GHK, KORV, and HSV stand for Greenwood, Hercowitz, and Krusell (1997), Krusell, Ohanian, Ríos-Rull, and Violante (2000), and Heathcote, Storesletten, and Violante (2010), respectively. NIPA stands for the National Income and Product Accounts.

Table 2: Benchmark Calibration Procedure

Parameter	Symbol	Value	Target	Data and SCE	Source
Discount factor	β	0.985	Capital-to-output ratio	2.9	NIPA and FAT
Disutility of labor	ϕ	67.8	Labor supply	1/3	
Production function parameter	ω	0.477	Labor share	2/3	NIPA
Production function parameter	ν	0.657	Skill premium $\frac{w_s}{w_u}$	1.8	HPV

This table reports our benchmark calibration procedure. The production function parameters ν and ω control the income share of equipment capital, skilled and unskilled labor in output. The acronym HPV stands for Heathcote, Perri, and Violante (2010). NIPA stands for the National Income and Product Accounts, and FAT stands for the Fixed Asset Tables.

Table 3: Steady-State Comparison of Wedges

		SCE	SPP
Tax (wedge) on equipment capital	τ_e	37.10%	39.54%
Tax (wedge) on structure capital	τ_s	42.20%	0.00%
Tax (wedge) on unskilled labor	$\tau_y(u)$	27.00%	26.60%
Tax (wedge) on skilled labor	$\tau_y(s)$	27.00%	-11.14%

This table compares the tax rates in the steady-state competitive equilibrium (column SCE) and wedges at the steady state of the solution to the social planning problem (column SPP).

Table 4: Steady-State Comparison of Allocations

	SCE	SPP
K_e/K_s	1.02	0.93
L_s/L_u	0.82	1.11
w_s/w_u	1.80	1.47

This table compares allocations in the steady-state competitive equilibrium (column SCE) and at the steady state of the solution to the social planning problem (column SPP). K_e/K_s denotes the equipment-to-structure capital ratio, L_s/L_u denotes the skilled-to-unskilled labor ratio and w_s/w_u denotes the skill premium.

Table 5: Sensitivity Analysis

	Benchmark			Benchmark		
σ	1	2	4	2	2	2
γ	1	1	1	0.5	1	2
Optimal taxes						
τ_e	24.39%	39.54%	54.84%	49.23%	39.54%	22.88%
τ_s	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
$\tau_y(u)$	18.77%	26.60%	33.75%	26.43%	26.60%	24.90%
$\tau_y(s)$	-5.29%	-11.14%	-21.23%	-16.46%	-11.14%	-5.08%
Difference in welfare gains						
Opt. DTC vs. current DTC	0.24%	0.19%	0.21%	0.20%	0.19%	0.22%
Opt. NDTC vs. current DTC	0.23%	0.14%	0.07%	0.10%	0.14%	0.21%

This table reports the results of the sensitivity analysis. In the first column, σ is intertemporal elasticity of substitution, γ is the inverse Frisch elasticity of labor supply, τ_e is the equipment capital tax (wedge), τ_s is the structure capital tax (wedge), $\tau_y(u)$ is the labor wedge of the unskilled agents, $\tau_y(s)$ is the labor wedge of the skilled agents. For all taxes (wedges), the table reports their steady state values. Opt. DTC refers to *optimal differential taxation of capital*, i.e., a reform that reforms both labor and capital taxes. Current DTC refers to *current differential taxation of capital*, i.e., a tax reform that reforms labor taxes, but leaves capital taxes at their current values. Opt. NDTC refers to *optimal nondifferential taxation of capital*, i.e., a reform in which the planner is free to adjust labor taxes, but must set the tax rate on equipment capital equal to the tax rate on structure capital.