

Rising skill premium and the dynamics of optimal capital and labor taxation

YI-CHAN TSAI

Department of Economics, National Taiwan University

C. C. YANG

Institute of Economics, Academia Sinica, National Chengchi University and Fengchia University

HSIN-JUNG YU

Central Bank, Republic of China(Taiwan)

With capital-skill complementarity, the secular decline in the price of capital equipment due to equipment-specific technological progress (ESTP) keeps pushing up the demand for skilled relative to unskilled labor and raising the skill premium. This paper quantitatively characterizes the dynamics of optimal taxation in response. Two main results emerge, regardless of whether the Ramsey (1927) or the Mirrlees (1971) approach is adopted. First, a tax on capital equipment corrects the “pecuniary externalities” caused by ESTP. The correction prescribes a downward or an upward adjustment of tax rates over time, depending on whether ESTP takes place at an accelerated or a decelerated pace. Second, both Ramsey and Mirrlees approaches prescribe an increasing marginal tax rate on labor income over time. Interestingly, we find that the prescribed pattern of optimal taxation resembles the empirical decline in capital taxes and the increase in labor taxes observed in the United States. In particular, despite the significant rise in the skill premium, the welfare gains of tax reform toward optimal Ramsey taxes are modest and small.

KEYWORDS. Skill premium, Optimal taxation, Capital-skill complementarity, Equipment-specific technological progress.

JEL CLASSIFICATION. C61, E22, E62, H21.

1. INTRODUCTION

It is known that the skill premium (i.e., the mean of weekly wages for college graduates relative to high school graduates) in the U.S. labor market has soared dramatically since the early 1980s.¹ This phenomenon has aroused serious concerns because it gives

Yi-Chan Tsai: yichantsai@ntu.edu.tw

C. C. Yang: ccyang@econ.sinica.edu.tw

Hsin-Jung Yu: hjyu@mail.cbc.gov.tw

We thank a co-editor and two anonymous referees, as well as Yili Chien, Bill Dupor, Yiting Li, Alexander Monge-Naranjo, and participants at various seminars and conferences for useful comments. Research grants of the Higher Education Sprout Project from the Ministry of Education (Grant No. 110L900201) and the Ministry of Science and Technology (MOST 110-2634-F-002-045-) in Taiwan are gratefully acknowledged.

¹ We will provide details on the time series of the skill premium later.

rise to greater income inequality and increasing disparity of economic well-being between skilled and unskilled workers. In this paper, we address the following question: how should taxation be set dynamically in response to the rising skill premium?

Let capital equipment be more substitutable for, or less complementary to, unskilled than skilled labor in production. This circumstance is referred to in the literature as “capital-skill complementarity.”² A critical implication of capital-skill complementarity is that a higher stock of capital equipment will raise the marginal product of skilled labor relative to unskilled labor. The U.S. economy has witnessed a steady, dramatic decline in the relative price of capital equipment because of “equipment-specific technological progress” (ESTP).³ In an influential paper, [Krusell et al. \(2000\)](#) found that, with a plausible difference in the elasticities of substitution between capital equipment and skilled versus unskilled labor that supports capital-skill complementarity, the rising skill premium since the early 1980s can be explained well by the secular cheapening of capital equipment associated with ESTP. Their finding holds even though there was a substantial increase in the relative supply of college graduates during the sample period they studied.

Despite the success of [Krusell et al. \(2000\)](#) in explaining the rising skill premium, no one has looked at the policy consequences of the rise. Building on the work of [Krusell et al. \(2000\)](#), this paper quantitatively characterizes the dynamics of optimal taxation in response.

There is a large body of literature investigating how a government should set taxes on capital versus labor.⁴ Of them, the contribution of [Werning \(2007\)](#) is most closely related to our paper. Rather than use the representative-household framework that abstracts from income inequality and is typical in the literature, he explicitly modeled distributional concerns by envisioning a heterogeneous-household economy. [Werning \(2007\)](#) emphasized that the need for the imposition of distorting taxes naturally arises from the trade-off between redistribution and efficiency in the heterogeneous-household framework.

Households in our economy are either skilled or unskilled. As such, it might appear that our model is just a special case of the one considered by [Werning \(2007\)](#). However, there is a difference: while all types of labor are equally complementary to capital equipment in the standard neoclassical production model in [Werning \(2007\)](#), skilled labor is more complementary to capital equipment than unskilled labor in our setting. This difference is of critical importance in light of the [Krusell et al. \(2000\)](#) finding that capital-skill complementarity plays a key role in explaining the rising skill premium. Indeed, our policy prescriptions deviate significantly from those prescribed by [Werning \(2007\)](#).

[Werning \(2007\)](#) adopted both the [Ramsey \(1927\)](#) and the [Mirrlees \(1971\)](#) approach to optimal taxation. When instantaneous household preferences are separable and isoelastic, [Werning \(2007, Proposition 2 for Ramsey taxation and Proposition 6 for Mirrleesian taxation\)](#) prescribed (i) a zero capital tax rate with no intertemporal distortion, except

² The idea builds on the work of [Griliches \(1969\)](#), which suggests that capital and skill are intrinsically complementary.

³ We will provide details on the time series of ESTP later.

⁴ See [Ljungqvist and Sargent \(2018, chapter 16\)](#) for a literature review.

for the initial period; and (ii) perfect labor tax smoothing across time and states. These policy prescriptions are aligned with the classical results that capital should go untaxed (Chamley (1986); Judd (1985)) and that intratemporal distortions on labor should be smoothed over time and states (Barro (1979); Lucas and Stokey (1983)).⁵

We adopt both the Ramsey and the Mirrlees approach as Werning (2007) did. Two main results emerge, regardless of the Ramsey or the Mirrlees approach. First, a tax on capital equipment corrects the “pecuniary externalities” caused by ESTP. The correction prescribes a downward or an upward adjustment of tax rates, depending on whether ESTP takes place at an accelerated or a decelerated pace over time. Second, both Ramsey and Mirrlees approaches prescribe an increasing marginal tax rate on labor income over time. Both results are in sharp contrast to those of Werning (2007).

The intuition for the dynamic pattern of optimal capital tax rates is as follows. As the extent to which ESTP obsolesces the old capital equipment accelerates (decelerates), the interest rate in competitive equilibrium will become higher (lower) to compensate for the loss (recon in the gain). To internalize a given level of the pecuniary externalities caused by ESTP, the imposed capital tax rate, which serves as an ad valorem tax on the interest rate, should then be adjusted downward (upward) accordingly. As to the dynamic pattern of optimal labor tax rates, the Ramsey solution is to remedy the worsening distribution as a result of secular ESTP, while the Mirrlees solution is to meet or relax the incentive compatibility constraints.

Interestingly, we find that the prescribed pattern of optimal taxation resembles the empirical decline in capital taxes and the increase in labor taxes in U.S. taxes. In particular, despite the significant rise in the skill premium, the welfare gains of tax reform toward optimal Ramsey taxes are modest and small.

Related work

Flug and Hercowitz (2000) provided evidence for capital-skill complementarity in production for a large panel of countries and investigated its effects on the relative wage and employment of skilled labor. Hornstein et al. (2005) and Quadriani and Ríos-Rull (2015) reviewed the literature on the effects of technical change on labor market inequalities. In their reviews, the work of Krusell et al. (2000) features significantly. The sample of Krusell et al. (2000) covers the 1963-1992 period. Maliar et al. (2020) extended the sample period to 2017, finding that the capital-skill complementarity framework remains remarkably successful in explaining the data. Costinot and Werning (2018) studied optimal technology regulation in a static environment. Among other things, they asked the question of whether taxes imposed on the use of new technologies should be raised or cut as the process of automation deepens and inequality increases. We relate our findings to theirs in the paper.

Although there is a large body of literature on capital-skill complementarity, to our knowledge, Jones et al. (1997), Slavík and Yazici (2014), and Angelopoulos et al. (2015) are the only three papers addressing the normative issue of optimal taxation in the presence of capital-skill complementarity.⁶

⁵ See also Zhu (1992) and Chari et al. (1994).

⁶ He and Liu (2008) and Angelopoulos et al. (2014) addressed the quantitative effects of some hypothetical tax-policy changes with capital-skill complementarity.

Jones et al. (1997) used capital-skill complementarity as an example to highlight a point: if two types of labor, skilled and unskilled, must be taxed at the same rate, the startling finding by Chamley (1986) and Judd (1985) that capital should go untaxed in the steady state may no longer hold when there is capital-skill complementarity. Assuming that labor's being skilled or unskilled is observable in a big family, Angelopoulos et al. (2015) studied tax smoothing in a business cycle model with capital-skill complementarity and endogenous skill formation. Both Jones et al. (1997) and Angelopoulos et al. (2015) studied their problems using the representative-household framework, and thereby both abstract from the redistributive role of taxation. By contrast, our study is in the context of rising wage inequality between skilled and unskilled workers.

Slavík and Yazici (2014) adopted the Mirrlees approach to taxation and basically adopted the same capital-skill complementarity form of production as suggested by Krusell et al. (2000). Their focus was on the differential taxation of capital income based on capital type, finding that capital equipment should be taxed at a higher rate than capital structures. This result violates the prescription of the classical work of Diamond and Mirrlees (1971) that tax systems should maintain production efficiency. We introduce ESTP, which is absent in Slavík and Yazici (2014). Moreover, we consider the Ramsey as well as the Mirrlees approach to optimal taxation.

The rest of the paper is organized as follows. Section 2 introduces the benchmark model. Sections 3-4 address the Ramsey approach, while Sections 5-6 address the Mirrlees approach. Section 7 considers an extension of the benchmark model, and Section 8 concludes.

2. BENCHMARK MODEL

We consider a dynamic economy based on the work of Werning (2007). The main departures are (i) the production technology is not of the standard neoclassical type but takes the form of empirically plausible capital-skill complementarity as suggested by Krusell et al. (2000), and (ii) there are no stochastic aggregate shocks; instead, the economy faces variation in the relative price of capital equipment because of ESTP. To simplify the exposition and to abstract from complication, it is assumed that people are endowed with perfect foresight regarding ESTP.

We first consider the Ramsey approach to taxation in Section 3-4 and turn to the Mirrlees approach in Sections 5-6.⁷ Time is discrete and the horizon is infinite, indexed by $t = 0, 1, 2, \dots$. The economy consists of heterogeneous households, a representative firm, and the government. We describe each of them in turn.

2.1 Households

Households are divided into two types: the skilled (s) and the unskilled (u). We assume that types are permanently fixed and that both types have the size of unit measure. We also assume that the labor productivity of the skilled and that of the unskilled are both

⁷ See Golosov et al. (2007) and Kocherlakota (2010, p.3) for discussions on the differences between the Mirrlees and the Ramsey.

equal to unity. Thus, the difference between the skilled and the unskilled in the benchmark model is driven solely by whether they are complementary to or substitutable for capital equipment. We relax these assumptions in the extension.

All households share the same preferences and their lifetime utility equals

$$V(i) = \sum_{t=0}^{\infty} \beta^t U(c_{it}, 1 - n_{it}), \quad i \in \{s, u\}, \quad (1)$$

where $\beta \in (0, 1)$ is the discount factor, c_{it} consumption, and n_{it} “raw” labor supply (hours worked). The period utility function $U(\cdot)$ is weakly concave and continuously twice differentiable with the properties of $U_c \geq 0$, $U_{cc} \leq 0$, $U_n \leq 0$, $U_{nn} \leq 0$, and $U_{cn} = 0$. Households own capital and rent it to the representative firm for its use. Capital structures and capital equipment owned by type i households are k_{ist} and k_{iet} , and their laws of motion are governed by

$$k_{ist+1} = (1 - \delta_s)k_{ist} + I_{ist}, \quad \text{given } k_{is0} \quad (2)$$

$$k_{iet+1} = (1 - \delta_e)k_{iet} + q_t I_{iet}, \quad \text{given } k_{ie0}, \quad (3)$$

where δ_s and δ_e are the depreciation rates for structures and equipment; I_{ist} and I_{iet} are the structures investment and equipment investment at time t . The key difference between capital structures and capital equipment lies in the presence of the variable q_t in the law of motion of capital equipment. Following the literature, we interpret the variable q_t in (3) as ESTP that enhances the productivity of newlyformed capital equipment relative to prior vintages of capital equipment. Its inverse, $1/q_t$, then represents the relative price of capital equipment; see [Hornstein et al. \(2005\)](#).

Households have the following flow budget constraint $\forall t$:

$$c_{it} + I_{ist} + I_{iet} + b_{it+1} = w_{it}n_{it} + r_{st}k_{ist} + r_{et}k_{iet} + (1 + r_{bt})b_{it} - \mathfrak{S}_t(\cdot), \quad i \in \{s, u\}, \quad (4)$$

where b_{it} is the risk-free, one-period noncontingent government bonds held by type i households and r_{bt} is their rate of return (b_{i0} is given); r_{st} and r_{et} denote the pre-tax rental rates of capital structures and equipment, w_{it} the pre-tax wage rate received by type i households, and $\mathfrak{S}_t(\cdot)$ is the tax function imposed by the government.

In line with the Ramsey approach and following [Werning \(2007\)](#), we specify $\mathfrak{S}_t(\cdot)$ a priori with

$$\mathfrak{S}_t(\cdot) = \tau_{Kst}(r_{st} - \delta_s)k_{ist} + \tau_{Ket}\left(r_{et} - \frac{\delta_e}{q_t}\right)k_{iet} + \tau_{Lt}w_{it}n_{it} - T_t,$$

where τ_{Kst} and τ_{Ket} are the (effective) flat tax rates imposed on structures and equipment capital income at time t , and (τ_{Lt}, T_t) is the linear tax schedule imposed on labor income at time t .⁸ Following [Slavík and Yazici \(2014\)](#), we allow for the possibility that

⁸The price of capital equipment goes down over time, so we implicitly assume that there is no tax deduction for negative capital gains.

$\tau_{Kst} \neq \tau_{Ket}$. As documented by Gravelle (2011), Slavík and Yazici (2014), and CBO (2014), the effective marginal tax rates on returns to capital assets in the U.S. tax code vary considerably depending on the capital type. These tax differentials are created mainly through tax depreciation allowances that differ from true economic depreciation.

Using (2)-(3) and the specification of $\mathfrak{F}_t(\cdot)$ above, one can rewrite (4) as

$$c_{it} + k_{ist+1} + \frac{k_{iet+1}}{q_t} + b_{it+1} = (1 - \tau_{Lt})w_{it}n_{it} + [1 + (1 - \tau_{Kst})(r_{st} - \delta_s)]k_{ist} \quad (5)$$

$$+ \left[\frac{1}{q_t} + (1 - \tau_{Ket}) \left(r_{et} - \frac{\delta_e}{q_t} \right) \right] k_{iet} + (1 + r_{bt})b_{it} + T_t,$$

which would reduce to the familiar household flow budget constraint if $q_t \equiv 1$; namely, if ESTP were absent.

Given b_{i0} , k_{is0} , and k_{ie0} , the household problem is to choose $\{c_{it}, n_{it}, k_{ist+1}, k_{iet+1}, b_{it+1}\}$ to maximize the lifetime utility (1), subject to the laws of motion (2)-(3) and a sequence of budget constraints (5). The resulting FOCs (first-order conditions) are given by

$$-U_{n,t}(i) = U_{c,t}(i)(1 - \tau_{Lt})w_{it}, \quad (6a)$$

$$U_{c,t}(i) = \beta U_{c,t+1}(i) [1 + (1 - \tau_{Kst+1})(r_{st+1} - \delta_s)], \quad (6b)$$

$$U_{c,t}(i) = \beta U_{c,t+1}(i) \left[\frac{q_t}{q_{t+1}} + q_t (1 - \tau_{Ket+1}) \left(r_{et+1} - \frac{\delta_e}{q_{t+1}} \right) \right], \quad (6c)$$

$$U_{c,t}(i) = \beta U_{c,t+1}(i) [1 + r_{bt+1}], \quad (6d)$$

where $U_{c,t}(i) \equiv \partial U(c_{it}, 1 - n_{it}) / \partial c_{it}$, and $U_{n,t}(i) \equiv \partial U(c_{it}, 1 - n_{it}) / \partial n_{it}$. These FOCs describe the type i household's optimal behavior in the face of factor prices $\{r_{bt}, r_{st}, r_{et}, w_{st}, w_{ut}\}$ and tax policy $\{\tau_{Lt}, \tau_{Kst}, \tau_{Ket}, T_t\}$. From the perspective of households, capital and government bonds are perfect substitutes. Thus, the no-arbitrage condition implies that the post-tax returns of capital given in (6b)-(6c) and the return on government bonds given in (6d) must be equal to each other in equilibrium. That is,

$$1 + (1 - \tau_{Kst+1})(r_{st+1} - \delta_s) = \frac{q_t}{q_{t+1}} + q_t (1 - \tau_{Ket+1}) \left(r_{et+1} - \frac{\delta_e}{q_{t+1}} \right) = 1 + r_{bt+1}.$$

Let us define the period 0 price of consumption at time t as:

$$p_t = \prod_{s=1}^t \frac{1}{(1 + r_{bs})}.$$

We normalize $p_0 = 1$. Imposing the no-Ponzi scheme, we can solve the flow budget constraint forward and obtain a household's lifetime budget constraint:

$$\sum_{t=0}^{\infty} p_t [c_{it} - (1 - \tau_{Lt})w_{it}n_{it}] = A_{i0} + T, \quad (7)$$

where $A_{i0} = [1 + (1 - \tau_{Ks0})(r_{st} - \delta_s)]k_{is0} + \left[\frac{1}{q_0} + (1 - \tau_{Ke0})(r_0 - \frac{\delta_e}{q_0})\right]k_{ie0} + b_{i0}$ denotes the initial wealth held by the type i household at time 0, and $T = \sum_{t=0}^{\infty} p_t T_t$ is the present value of the lump-sum transfers $\{T_t\}$ as defined in Werning (2007).

2.2 Firms

There is a representative firm producing the final good with a production function taking the following form at time t :

$$Y_t = F(K_{st}, K_{et}, N_{st}, N_{ut}) = K_{st}^\alpha \left[\mu N_{ut}^\sigma + (1 - \mu) \left[\lambda K_{et}^\rho + (1 - \lambda) N_{st}^\rho \right]^{\frac{\sigma}{\rho}} \right]^{\frac{1-\alpha}{\sigma}}, \quad (8)$$

where Y_t denotes output, K_{st} capital structures, K_{et} capital equipment, N_{st} the skilled labor input, and N_{ut} the unskilled labor input, with $\sigma, \rho < 1$. All of these terms are aggregate variables, with $N_{st} = n_{st}$, $N_{ut} = n_{ut}$, $K_{st} = \sum_{i \in \{s, u\}} k_{ist}$, and $K_{et} = \sum_{i \in \{s, u\}} k_{iet}$ in the benchmark model. This four-factor production function is the same as that in [Krusell et al. \(2000\)](#). A key feature of this production function is that it allows for different elasticities of substitution between capital equipment and the two types of labor. In particular, the elasticity of substitution between capital equipment and unskilled labor equals $1/(1 - \sigma)$, while the elasticity of substitution between capital equipment and skilled labor equals $1/(1 - \rho)$. The so-called capital-skill complementarity arises if $\sigma > \rho$.

All markets are competitive, and we let the final good be the numéraire. Subject to the production technology (8), the representative firm maximizes its profit

$$\Pi_t = Y_t - w_{st}N_{st} - w_{ut}N_{ut} - r_{st}K_{st} - r_{et}K_{et}, \quad (9)$$

where w_{st} , w_{ut} , r_{st} , and r_{et} are the factor prices for N_{st} , N_{ut} , K_{st} , and K_{et} , respectively.

The FOCs for the representative firm are given by $\frac{\partial Y}{\partial K_s} = r_s$, $\frac{\partial Y}{\partial K_e} = r_e$, $\frac{\partial Y}{\partial N_s} = w_s$, and $\frac{\partial Y}{\partial N_u} = w_u$. The skill premium in the benchmark model is defined as

$$\xi \equiv \frac{w_s}{w_u} = \frac{(1 - \mu)(1 - \lambda)}{\mu} \left[\lambda \left(\frac{K_e}{N_s} \right)^\rho + (1 - \lambda) \right]^{\frac{\sigma - \rho}{\rho}} \left(\frac{N_u}{N_s} \right)^{1 - \sigma}. \quad (10)$$

Using $\ln(1 + y) \approx y$, one can express the skill premium defined in (10) as

$$\ln \xi \approx \lambda \left(\frac{\sigma - \rho}{\rho} \right) \left(\frac{K_e}{N_s} \right)^\rho + (1 - \sigma) \ln \left(\frac{N_u}{N_s} \right) + \text{constant}. \quad (11)$$

Note that $\frac{\partial \xi}{\partial K_s} = 0$ since the skill premium as defined in (10) does not depend on capital structures directly. Note also that $\frac{\partial \xi}{\partial K_e} > 0$ because of the feature of capital-skill complementarity with $\sigma > \rho$ ($\frac{\partial \xi}{\partial K_e} = 0$ would hold if $\sigma = \rho$). [Krusell et al. \(2000\)](#) called the first component of (11) the “capital-skill complementarity effect” and the second component the “relative quantity effect.” The first effect indicates that, given $\sigma > \rho$, faster growth in capital equipment relative to skilled labor input will raise the skill premium, as it increases the relative demand for skilled labor. The second effect indicates that,

given $\sigma < 1$, faster growth in skilled relative to unskilled labor input will reduce the skill premium as it increases the relative supply of skilled labor.

2.3 Government

The government is required to finance an exogenous stream of government expenditures $\{G_t\}$, obey its budget constraints, and fully commit to its fiscal policy (taxes imposed and debts issued), given the initial government bond $B_0 = b_{s0} + b_{u0}$.

The government flow budget constraint is given by

$$G_t + 2T_t + (1 + r_{bt})B_t = \begin{bmatrix} \tau_{Lt}(w_{st}N_{st} + w_{ut}N_{ut}) + B_{t+1} \\ \tau_{Kst}(r_{st} - \delta_s)K_{st} + \tau_{Ket}\left(r_{et} - \frac{\delta_e}{q_t}\right)K_{et} \end{bmatrix}, \forall t, \quad (12)$$

where B_{t+1} is the amount of risk-free one-period noncontingent government bonds issued at time t . Note that we have $2T_t$ rather than T_t in (12). This is because both types of households have the size of unit measure in our setup. We impose the no-Ponzi scheme on the government so that government debt must be fully repaid by future primal surpluses (taxes collected net of government expenditures).

We can sum the budget constraints of both types of households and the government to obtain the aggregate resource constraint at time t for the economy:

$$Y_t = C_t + I_{st} + I_{et} + G_t, \quad (13)$$

where $C_t = \sum_{i \in \{s,u\}} c_{it}$, $I_{st} = \sum_{i \in \{s,u\}} I_{ist}$, and $I_{et} = \sum_{i \in \{s,u\}} I_{iet}$.

The definition of a competitive equilibrium for our model is standard. This completes the description of the benchmark model.

3. RAMSEY PROBLEM

We formulate the Ramsey problem in this section.

Different government policies result in different competitive equilibria. Given B_0 with $\{b_{i0}\}_{i \in \{s,u\}}$ and (K_{s0}, K_{e0}) with $\{k_{is0}, k_{ie0}\}_{i \in \{s,u\}}$, the Ramsey problem is to choose a competitive equilibrium to maximize the following social welfare function:

$$SWF = \sum_{i \in \{s,u\}} \psi^i V(i), \quad (14)$$

where $V(i)$ is given by (1), and $\psi^i \geq 0$ with $i \in \{s,u\}$ are the Pareto weights. Note that $\psi^s = \psi^u$ if the planner obeys the utilitarian criterion.

We adopt the primal approach to the Ramsey problem. In particular, we follow [Werning \(2007\)](#), who extended the primal approach of [Lucas and Stokey \(1983\)](#) from the representative-agent to the heterogeneous-agent framework.

[Werning \(2007\)](#) observed that inefficiencies in consumption and labor supply due to distorting linear taxation are all confined to the determination of their aggregates, which are represented by $\{C_t, n_{st}, n_{ut}\}$ with $n_{st} = N_{st}$ and $n_{ut} = N_{ut}$ in our benchmark

model. Following [Werning \(2007\)](#), consider a fictitious representative agent with utility $U^f(C_t, n_{st}, n_{ut}; \varphi)$ which solves the static subproblem:

$$U^f(C_t, n_{st}, n_{ut}; \varphi) = \max_{c_{st}, c_{ut}} \sum_{i \in \{s, u\}} \varphi^i U(c_{it}, 1 - n_{it}), \text{ subject to } \sum_{i \in \{s, u\}} c_{it} = C_t,$$

where $\varphi = \{\varphi^i\}_{i \in \{s, u\}}$ with $\sum_{i \in \{s, u\}} \varphi^i = 1$ are “market” weights. Let the solution of this static subproblem be

$$c_{it}^* = c_{it}(C_t, \varphi), \quad i = \{s, u\},$$

where household consumption is expressed as a function of aggregate consumption C_t and market weights φ . Note that c_{it}^* is independent of (n_{st}, n_{ut}) because $U_{cn} = 0$ by assumption.

The envelope condition for the above static subproblem gives

$$U_{C_t}^f = \varphi^i U_{c,t}(i), \quad U_{n,t}^f(i) = -\varphi^i U_{n,t}(i), \quad i = \{s, u\}, \quad (15)$$

where $U_{C_t}^f \equiv \partial U^f / \partial C_t$, and $U_{n,t}^f(i) \equiv \partial U^f / \partial n_{it}$. Using (15) and (6a)-(6c), we then have

$$\frac{-U_{n,t}^f(i)}{U_{C_t}^f} = (1 - \tau_{Lt}) w_{it}, \quad i = \{s, u\}, \quad (16)$$

$$\begin{aligned} \frac{U_{C_t}^f}{\beta U_{C_{t+1}}^f} &= [1 + (1 - \tau_{Kst+1})(r_{st+1} - \delta_s)] \\ &= \left[\frac{q_t}{q_{t+1}} + q_t (1 - \tau_{Ket+1}) \left(r_{et+1} - \frac{\delta_e}{q_{t+1}} \right) \right]. \end{aligned} \quad (17)$$

Utilizing the results from the fictitious representative-agent problem, we derive the implementability conditions from the household’s lifetime budget constraint (7):

$$\sum_{t=0}^{\infty} \beta^t \left[U_{C_t}^f c_{it}^* + U_{n,t}^f(i) n_{it}^* \right] = U_{C_0}^f (A_{i0} + T), \quad i \in \{s, u\}. \quad (18)$$

Given $U^f = U^f(C_t, n_{st}, n_{ut}; \varphi)$ and $c_{it}^* = c_{it}(C_t, \varphi)$, we have expressed the implementability conditions in terms of aggregate allocation $\{C_t, n_{st}, n_{ut}\}$ and market weights φ , similar to those in [Werning \(2007\)](#).

Because $w_{st} = w_{ut}$ need not hold in our model, we also need to impose the following restriction in the formulation of the Ramsey problem:

$$\frac{U_{n,t}(s)}{U_{c,t}(s) w_{st}} = \frac{U_{n,t}(u)}{U_{c,t}(u) w_{ut}}, \quad \forall t, \quad (19)$$

which guarantees that both the skilled and unskilled households face the same labor marginal tax rate at any point in time according to (6a). Using (15), the restriction (19) can be expressed as

$$\frac{U_{n,t}^f(s)}{U_{C_t}^f w_{st}} = \frac{U_{n,t}^f(u)}{U_{C_t}^f w_{ut}}, \quad \forall t,$$

which leads to

$$\log \frac{U_{n,t}^f(s)}{U_{n,t}^f(u)} - \log \xi_t = 0, \forall t, \quad (20)$$

where $\xi = w_s/w_u$ is the skill premium as given by (10).

Let $\Psi = \{\Psi^s, \Psi^u\}$, $\{\beta^t \Gamma_t\}$, and $\{\beta^t \Upsilon_t\}$ denote, respectively, the Lagrange multipliers on the implementability conditions (18), the resource constraints (13), and the restriction (20). Given initial capital structures K_{s0} and capital equipment K_{e0} , initial bond B_0 , the distributions of their holdings between the skilled and the unskilled, and initial capital tax rates τ_{Ks0} and τ_{Ke0} ,⁹ forming the Lagrangian for the Ramsey problem gives

$$\begin{aligned} L = & \max_{\{C_t, \varphi, n_{st}, n_{ut}, K_{st+1}, K_{et+1}, T\}} \sum_{t=0}^{\infty} \beta^t W(C_t, n_{st}, n_{ut}, \varphi) \\ & + \sum_{t=0}^{\infty} \beta^t \Gamma_t \left[F(K_{st}, K_{et}, N_{st}, N_{ut}) + \frac{(1-\delta_e)K_{et}}{q_t} - \frac{K_{et+1}}{q_t} \right. \\ & \quad \left. + (1-\delta_s)K_{st} - K_{st+1} - C_t - G_t \right] \\ & + \sum_{t=0}^{\infty} \beta^t \Upsilon_t \left[\log \frac{U_{n,t}^f(s)}{U_{n,t}^f(u)} - \log \xi_t \right] \\ & - U_{C_0}^f \sum_{i \in \{s, u\}} \Psi^i (A_{i0} + T), \end{aligned}$$

where the pseudo-utility function $W(\cdot)$ is defined by

$$W(C_t, n_{st}, n_{ut}, \varphi) = \sum_{i \in \{s, u\}} \left\{ \psi^i U(c_{it}^*, 1 - n_{it}^*) + \Psi^i \left[U_{C_t}^f c_{it}^* + U_{n,t}^f(i) n_{it}^* \right] \right\}.$$

Let $W_{C_t} \equiv \partial W / \partial C_t$ and $W_{n,t}(i) \equiv \partial W / \partial n_{it}$ with $i = \{s, u\}$. The resulting FOCs of the Ramsey problem for $t \geq 1$ are¹⁰

$$W_{C_t} = \Gamma_t, \quad (21)$$

$$-W_{n,t}(s) = \Gamma_t w_{st} + \frac{\Upsilon_t}{n_{st}} \left(\frac{(\partial U_{n,t}^f(s) / \partial n_{st}) n_{st}}{U_{n,t}^f(s)} - \frac{n_{st}}{\xi_t} \frac{\partial \xi_t}{\partial n_{st}} \right), \quad (22)$$

$$-W_{n,t}(u) = \Gamma_t w_{ut} - \frac{\Upsilon_t}{n_{ut}} \left(\frac{(\partial U_{n,t}^f(u) / \partial n_{ut}) n_{ut}}{U_{n,t}^f(u)} + \frac{n_{ut}}{\xi_t} \frac{\partial \xi_t}{\partial n_{ut}} \right), \quad (23)$$

⁹ The initial capital tax rates, τ_{Ks0} and τ_{Ke0} , should be a choice variable for the Ramsey planner. Werning (2007) allowed for unrestricted wealth taxation, showing that, at the optimum, the planner will implement a confiscatory rate for both τ_{Ks0} and τ_{Ke0} to confiscate all capital of households (and will also confiscate all bond holdings of households). This outcome seems extreme and unrealistic in the real world. We follow the standard practice in the literature to restrict the planner's ability to choose initial capital tax rates in the Ramsey problem. Specifically, we let both τ_{Ks0} and τ_{Ke0} correspond to the U.S. capital tax rate in the initial steady state in our quantitative study.

¹⁰ It is known that the FOCs for $t = 0$ differ from those for $t \geq 1$ in the Ramsey problem. The FOC with respect to φ is not reported here but is used in our numerical analysis.

$$\frac{\Gamma_t}{q_t} = \beta \left[\Gamma_{t+1} \left(r_{et+1} + \frac{1 - \delta_e}{q_{t+1}} \right) - \Upsilon_{t+1} \left(\frac{1}{\xi_{t+1}} \frac{\partial \xi_{t+1}}{\partial K_{et+1}} \right) \right], \quad (24)$$

$$\Gamma_t = \beta \Gamma_{t+1} (r_{st+1} + 1 - \delta_s), \quad (25)$$

$$\sum_{i \in \{s, u\}} \Psi^i = 0, \quad (26)$$

where the form of the last FOC with respect to T is qualitatively the same as equation (17) derived by Werning (2007). Together with the implementability conditions (18), the resource constraints (13) and the restriction (20), the FOCs of the Ramsey problem characterize the optimal allocation of the competitive equilibrium under the optimal fiscal policy $\{\tau_{Lt}, \tau_{Kst+1}, \tau_{Ket+1}, T, B_t\}$. Note that the planner chooses $T = \sum_{t=0}^{\infty} p_t T_t$ rather than the sequence $\{T_t\}$ in the Ramsey problem. As pointed out by Werning (2007), this makes the mix between $\{T_t\}$ and $\{B_t\}$ to smooth taxation become indeterminate at the optimum. In view of this indeterminacy, we shall focus on the characterization of optimal tax rates $\{\tau_{Lt}, \tau_{Kst+1}, \tau_{Ket+1}\}$.

The optimal intertemporal condition for capital structures (25) is standard and identical to that in the laissez faire competitive equilibrium. This immediately implies that saving decisions on capital structures should not be distorted. The nature of capital structures in our model is not different from that of capital in Werning (2007); in particular, there is no capital-skill complementarity. Werning (2007) prescribed that capital should go untaxed in his setting. It thus comes as no surprise that capital structures should go untaxed in our setting. Given the secular ESTP and capital-skill complementarity between capital equipment and skilled labor, we confine our analysis to the taxation of income from capital equipment and labor.

In the next section, we first quantitatively characterize the dynamics of Ramsey taxation resulting from the optimal allocation and then consider a simplified model to explain the mechanism underlying the quantitative results.

4. DYNAMICS OF RAMSEY TAXATION

This section quantitatively characterizes the dynamics of Ramsey taxation in the face of ESTP $\{q_t\}$. The optimal solution we obtain maximizes social welfare along the transition between the initial steady state and an endogenously determined final steady state. Note that a balanced growth path does not exist with the production function suggested by Krusell et al. (2000) if ESTP $\{q_t\}$ exhibits a trend; see He and Liu (2008). Therefore, the typical solution methods that involve log-linearizing around a balanced growth path are not applicable to our model. We instead compute the transitional dynamics from the initial steady state to the new steady state by adopting a nonlinear solution method.

It is important to recognize that the quantitative results of the benchmark model are not for the purpose of matching with the data, but for the purpose of illustrating the underlying mechanism of our model. As for the matching with the data, only the quantitative results of the extended model should be taken seriously.

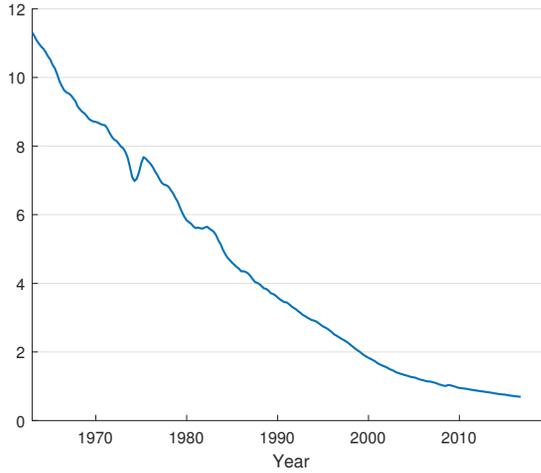


Figure 1.: Price of capital equipment relative to consumption, 1963-2016

Note: index 2009=1, seasonally adjusted, quarterly data.

Source: Federal Reserve Bank Economic Data (FRED), series PERIC, 1963-2016.

4.1 Calibration

First, we briefly describe the time-series data: ESTP $\{q_t\}$. [Gordon \(1990\)](#) is the seminal work on measuring ESTP. [DiCecio \(2009\)](#) constructed the relative price of capital equipment by chainweighting the deflator for equipment and software from NIPA. DiCecio's (2009) data sequence is updated in the Federal Reserve Economic Data (FRED) and is shown in Figure 1. The time series on ESTP $\{q_t\}$ are simply the reciprocal of the time series in the figure. It is seen that $1/q_t$ has been falling and hence q_t has been increasing constantly since the 1960s.

To obtain q_t beyond the data shown in Figure 1, we first compute the average growth rate of q_t from 2010 to 2016 and let it serve as the growth rate of q_t from 2016 to 2017. We then follow He and Liu's (2008) method by assuming that the growth rates of $\{q_t\}$ after 2016 slow down linearly to zero from 2017 to 2047, reaching a steady state at 2047, and then remain constant from 2047 to 2142. In this way, we construct a time series sequence of $\{q_t\}$ for a length of 180 years consisting of 54 years of data (1963-2016) from FRED and 126 years of artificial data (2017-2142) with 2047 to 2142 being the new steady state. Given the availability of relevant data and the fact that the q_t series are relatively stationary before 1963, we choose the year 1963 as our initial steady state so as to line up with the skill premium data shown later. We normalize $q_{t=1963} \equiv q_0 = 1$.

To compute the transition dynamics in response to ESTP $\{q_t\}$, we calibrate the model parameters to match some key features of the U.S. economy. We take one period in the model to be one calendar year in the data. The details are as follows.

Most parameter values of production are taken directly from [Krusell et al. \(2000\)](#). Therefore, we set $\alpha = 0.117$, $\sigma = 0.401$, and $\rho = -0.495$, so that the elasticity of substitution between capital and skilled labor is about $\frac{1}{1-\rho} \approx 0.67$, while that between capital

and unskilled labor is about $\frac{1}{1-\sigma} \approx 1.67$. The depreciation rates of structures and equipment are set as $\delta_s = 0.05$ and $\delta_e = 0.125$.

For the preferences on the household side, we specify the period utility function $U(\cdot)$ in (1) to be

$$U(c, 1 - n) = u(c) + v(1 - n) = \frac{c^{1-\gamma_c}}{1-\gamma_c} + \chi \frac{(1-n)^{1-\gamma_n}}{1-\gamma_n}. \quad (27)$$

This form of separable isoelastic preferences is common in the literature, and it also facilitates direct comparison with the results in [Werning \(2007\)](#). We set γ_c and γ_n to be 1.5 and 3, respectively, and the discount factor β to be 0.98. These are standard in the literature.

Using national account statistics as a primary source, [McDaniel \(2007\)](#) calculated series of average tax rates on labor income and capital income for 15 OECD countries for the period 1950-2003.¹¹ Since the year 1963 serves as our initial steady state, we simply let McDaniel's (2007) calculated tax rates in 1963 be the initial U.S. tax rates. As to the ratio of government expenditures to GDP, we set it to 17.5%, which is the average between 1963 and 2017 from the Bureau of Economic Analysis's National Income and Product Accounts (NIPA) tables. We stick to this ratio throughout the period we study.

There are three parameters that remain to be calibrated, which are μ , λ , and χ . Parameters μ and λ are related to the production function (8), and χ is the relative weight between consumption and leisure in the utility function (27). We calibrate the values of these parameters to match the following three moment conditions of the U.S. economy in 1963 (the initial steady state):

1. The skill premium ξ , when defined as the average annual wage of college graduates relative to that of high-school graduates, equals 1.474 (Autor, 2014).
2. The average income share of capital, which includes both capital structures and capital equipment, i.e., $\frac{rK}{Y}$, is around 0.3 in 1963 (OECD.Stat).
3. The ratio of consumption to GDP in 1963 is equal to 0.6 (NIPA).

We also consider the other moment conditions of the U.S. economy in 1963:

1. The capital-output ratio, $\frac{K}{Y}$, is about 2.69 (NIPA).
2. The ratio of gross domestic investment to GDP in 1963 is equal to 0.162 (NIPA).
3. The ratio of structure investment to equipment investment in 1963 is equal to 0.486 (NIPA).

As can be seen from Table 1, the above match does fairly well on these three additional moment conditions.

Finally, we need to specify the distribution of capital between skilled and unskilled workers in the initial steady state. We match it to the fraction of the unskilled's wealth in the total wealth. Since the data on this ratio are available only from 1989 on, we simply

¹¹ The tax series have been updated to 2013 by McDaniel.

use the 1989 datum, which is equal to 0.57.¹² Although admittedly unsatisfactory, it is the earliest datum we can find with regard to this ratio.

Table 2 summarizes all of our parameter values. The resulting steady-state capital structures and capital equipment in competitive equilibrium will serve as the initial one for our dynamic economy.

Table 1.: Initial moment conditions

Target	Model	Data
Skill premium	1.474	1.474
Capital income share	0.3	0.3
Consumption-output ratio	0.6	0.6
Capital-output ratio	2.64	2.69
Gross investment-output ratio	0.221	0.162
Ratio of investment structure to equipment	0.485	0.486

Table 2.: Parameter values

Parameters	Symbol	Value
Parameter values set		
Discount factor	β	0.98
Elasticity of intertemporal substitution	$1/\gamma_c$	0.67
Elasticity of leisure	$1/\gamma_n$	0.33
Elasticity of substitution, N_u and K_e	$1/(1-\sigma)$	1.67
Elasticity of substitution, N_s and K_e	$1/(1-\rho)$	0.67
Depreciation rate of K_s	δ_s	0.05
Depreciation rate of K_e	δ_e	0.125
Capital tax rate	τ_{K0}	0.341
Labor tax rate	τ_{L0}	0.142
Income share of K_s	α	0.117
Government expenditure to GDP ratio	G/Y	0.175
Parameter values calibrated		
Income share of N_u	μ	0.333
Income share of K_e	λ	0.348
Utility weight of leisure	χ	2.53

On the basis of the calibrated parameters and the time-series data of ESTP $\{q_t\}$, we compute the transitional dynamics of optimal taxation using a nonlinear solution

¹² The data are from the U.S. Census Bureau, Asset ownership of households.

method in the spirit of [Conesa and Krueger \(1999\)](#) and [He and Liu \(2008\)](#). The details of the method are relegated to Online Appendix A.

4.2 Quantitative results

Given q_t at a point in time, we obtain $\tau_{Ket+1} > 0$ and $\tau_{Lt} > 0$ at the optimum; see Figure 2. Our central question is this: how will this optimal tax structure at a point in time vary over time in response to increasing q_t ?

Figure 2 reports the dynamics of the optimal tax rate on equipment capital income, τ_{Ket+1} , and that of the optimal marginal tax rate on labor income, τ_{Lt} , for $\psi^u/\psi^s = 1, 2$ and 5.¹³ Two features stand out. First, optimal $\{\tau_{Ket+1}\}$ basically displays a declining trend over time before 1999; however, this declining trend is reversed and turns into an increasing trend over time after 1999.¹⁴ Second, optimal $\{\tau_{Lt}\}$ displays an increasing trend over time all the way.

Let $\Delta q_{t+1} = q_{t+1} - q_t$. The left panel of Figure 3 plots the trajectories of $\{\frac{\Delta q_{t+1}}{q_t}\}$ and optimal $\{\tau_{Ket+1}\}$ as $\psi^u/\psi^s = 1$. The right panel of Figure 3 shows the smoothed-curve representation of the trajectories after applying the Hodrick-Prescott filter. Taken together, the figure shows that, after the oil crisis of 1973-74,¹⁵ while optimal $\{\tau_{Ket+1}\}$ are adjusted downward as ESTP takes place at an accelerated pace (i.e., $\frac{\Delta q_{t+1}}{q_t}$ is increasing over time before 1999), they are adjusted upward as ESTP takes place at a decelerated pace (i.e., $\frac{\Delta q_{t+1}}{q_t}$ is decreasing over time after 1999). This salient characteristic of optimal $\{\tau_{Ket+1}\}$ remains robust with variation in Pareto weights $\psi^u/\psi^s \geq 1$.

When q_t is fixed, there is no ESTP. The problem facing the planner is simply with regard to how to divide a fixed “pie” (associated with a fixed q_t) between the skilled and the unskilled. However, as q_t is increasing over time, ESTP takes place constantly. Then, the problem facing the planner is not only about how to divide a fixed “pie” but also about how to exploit the improved technology to make the “pie” bigger. The planner can choose either to lean against ESTP by raising τ_{Ket+1} or to lean toward ESTP by lowering τ_{Ket+1} . Our quantitative finding is that technology wins out and the planner adopts the leaning-toward policy if ESTP takes place at an accelerated pace, while technology loses out and the planner adopts the leaning-against policy if ESTP takes place at a decelerated pace.

In a recent paper, [Costinot and Werning \(2018\)](#) studied optimal technology regulation. Among other things, they conducted a comparative statics in a simple economy

¹³ Figure 2 shows that, given q_t at a point in time, both τ_{Ket+1} and τ_{Lt} become higher in general as ψ^u/ψ^s increases. This outcome is intuitive. As the Pareto weight of the unskilled gets larger relative to that of the skilled, Ramsey taxation prescribes both a higher τ_{Ket+1} to have a lower skill premium and a higher τ_{Lt} to achieve a larger redistribution.

¹⁴ The huge spike of τ_{Ket+1} shown in Figure 2 is due to a dramatic change in q_t during the oil crisis of 1973-74; see Figure 1.

¹⁵ The dramatic change in q_t during the oil crisis of 1973-74 causes the filtered $\{q_t\}$ in Figure 3 to decline rather than rise before the oil crisis. This in turn causes the filtered $\{\tau_{Ket+1}\}$ in Figure 3 to rise rather than decline before the oil crisis. However, the opposite pattern will arise before the oil crisis once we use the filtered $\{q_t\}$ instead of the actual $\{q_t\}$. As such, our illustration of Figure 3 focuses on the period after the oil crisis.

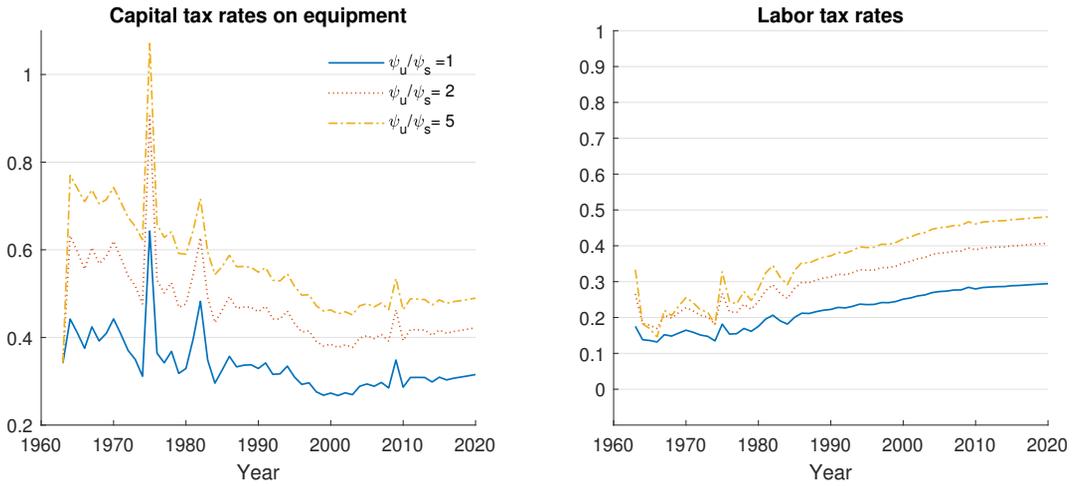


Figure 2.: Ramsey taxation

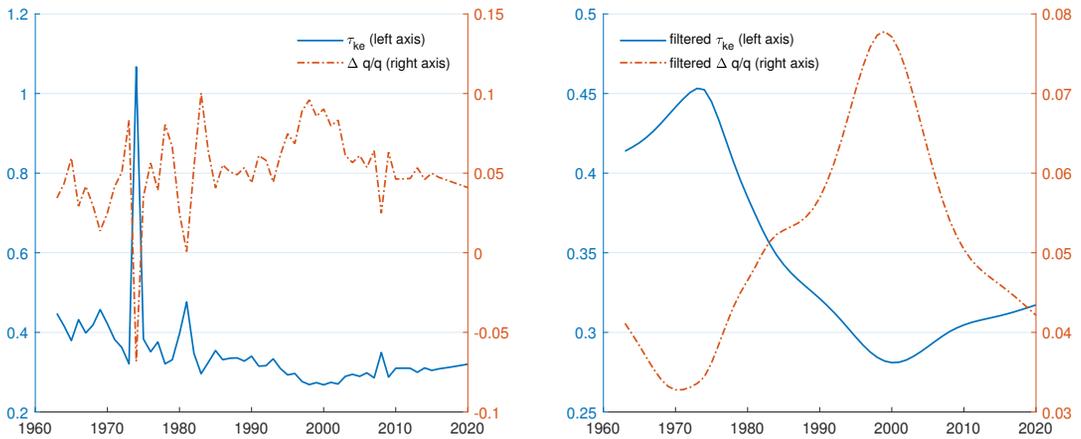


Figure 3.: Capital tax rates on equipment and pace of ESTP

to address the following question: if improvements in new technologies make the economy's inequality worse, should taxes imposed on the use of these improved technologies be raised or cut? They showed that while distributional concerns create a rationale for positive taxes on new technologies, the magnitude of these taxes may decrease rather than increase as the process of automation deepens and inequality increases. Our finding is consistent with their result, in that the planner may lower τ_{Ket+1} to lean toward ESTP.

In the setting where the relative wages between different households are fixed and unchanged over time, [Werning \(2007\)](#) explained his finding of perfect labor tax smoothing as follows (p. 927): “With heterogeneous workers and a lump-sum tax, it is distributional concerns that determine the optimal tax rate. Since the desired level of redistribution is pinned down by the constant distribution of relative skills across workers,

a constant tax rate is optimal.” On the basis of this explanation, one would expect the emergence of perfect labor tax smoothing in our setting as well. This is logical because, like that in [Werning \(2007\)](#), the distribution of relative skill across workers is permanently fixed and so remains constant over time in our setting. However, contrary to the expectation, we find that optimal τ_{Lt} is increasing over time. Why?

Because of capital-skill complementarity, the relative wage between the skilled and the unskilled (i.e., the skill premium) is increasing over time as a result of ESTP in our setting. Put differently, in terms of relative wages, the distribution between different households remains the same over time in the setting of [Werning \(2007\)](#), whereas the distribution between the skilled and the unskilled is worsening over time in our setting. This very difference explains why optimal τ_{Lt} remains constant over time in [Werning \(2007\)](#), whereas optimal τ_{Lt} is increasing over time in our setting: the planner implements an increasing marginal tax rate τ_{Lt} over time to remedy the worsening distribution as a result of secular ESTP, and this increasing τ_{Lt} arises despite the distribution of relative skill across workers, as in [Werning \(2007\)](#), being permanently fixed in our setting.

4.3 What if there is a different σ ?

How robust are the results shown in [Figure 2](#)? The production technology given by [\(8\)](#) features capital-skill complementarity if $\sigma > \rho$. The difference between σ and ρ is the key driver of our quantitative analysis. It is intuitive that the higher is the value of σ above ρ , the higher will be the degree of capital-skill complementarity in production.

Here we fix $\rho = -0.495$ as in our calibration but vary σ between $\sigma = \rho$ and $\sigma = 1$ (recall $\sigma = 0.401$ in our calibration). According to [\(11\)](#), the “capital-skill complementarity effect” (the first component of $\ln \xi$) will vanish if $\sigma = \rho$, whereas the “relative quantity effect” (the second component of $\ln \xi$) will vanish if $\sigma = 1$. We vary the value of σ between the two extremes, $\sigma = \rho$ and $\sigma = 1$, with the objective of checking the robustness of the pattern of Ramsey taxation in the face of ESTP as shown in [Figure 4](#).

[Figure 4](#) reports our finding with $\psi^u/\psi^s = 1$. It is seen that, except for the extreme case where $\sigma = \rho = -0.495$ so that capital-skill complementarity vanishes, the pattern of Ramsey taxation shown in [Figure 4](#) basically remains the same as that shown in [Figure 2](#).

4.4 A simplified model

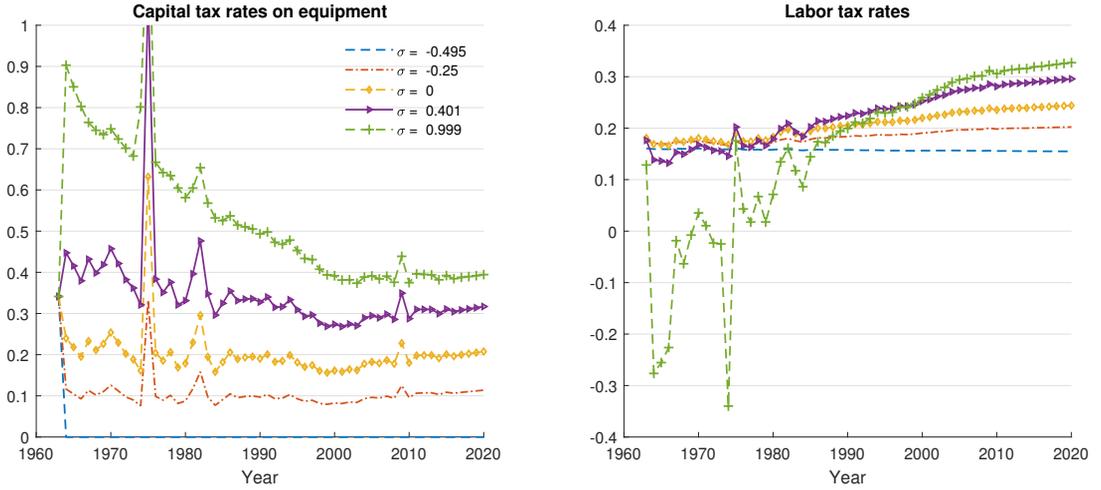
This subsection analytically considers a simplified model as a complement to our quantitative study. We seek to explain the mechanism underlying the quantitative results.

Let $\alpha = 0$, $\sigma = 1$ and $\rho = 0$ so that the production function [\(8\)](#) reduces to

$$Y_t = \mu N_{ut} + (1 - \mu) K_{et}^\lambda N_{st}^{1-\lambda}.$$

The elasticity of substitution between capital equipment and unskilled labor equals $1/(1 - \sigma) = \infty$ as $\sigma = 1$. We have

$$w_{ut} = \mu, \tag{28}$$

Figure 4.: Ramsey taxation under different σ

$$w_{st} = (1 - \mu)(1 - \lambda) \left(\frac{K_{et}}{N_{st}} \right)^\lambda = \mu \xi_t, \quad (29)$$

$$\xi_t = \frac{w_{st}}{w_{ut}} = \frac{(1 - \mu)(1 - \lambda)}{\mu} \left(\frac{K_{et}}{N_{st}} \right)^\lambda, \quad (30)$$

where the second equality of w_{st} has invoked (30). Equation (30) shows that variation in ξ_t is completely determined by variation in $\frac{K_{et}}{N_{st}}$ and there is no “relative quantity effect.”¹⁶ We also let $\delta_e = 0$.

Let the period utility function take the quasilinear form: $U(c, n) = c - \chi \frac{n^2}{2}$. The household FOC (6a) gives

$$1 - \tau_{Lt} = \chi \frac{N_{st}}{w_{st}} = \chi \frac{N_{ut}}{w_{ut}} \Rightarrow N_{st} = N_{ut} \xi_t, \quad (31)$$

while the household FOC (6c) gives

$$\frac{1}{q_t} = \beta \left[\frac{1}{q_{t+1}} + (1 - \tau_{Ket+1}) r_{et+1} \right]. \quad (32)$$

Note that the household’s Euler equation (32) would reduce to the standard one (under the quasilinear utility) if $q_t = q_{t+1} = 1$; namely, if ESTP were absent.

We formulate the Ramsey problem and derive the resulting FOCs in Online Appendix B. Given that the period utility function takes the quasilinear form, we let $\psi^s < \psi^u$ to create a motive for the planner’s redistribution toward the unskilled.

¹⁶ Given that the skill premium ξ_t is increasing over time even though $\frac{N_{ut}}{N_{st}}$ is decreasing (see Figure C.1 in Online Appendix C), it is clear that the “capital-skill complementarity effect” dominates the behavior of ξ_t .

We obtain

$$1 - \tau_{Lt} = \frac{(\psi^s + \psi^u)(1 + \frac{\xi_t^2}{1 + \lambda})}{2(\psi^s + \frac{\xi_t^2}{1 + \lambda}\psi^u)}, \quad (33)$$

which implies

$$\frac{\partial(1 - \tau_{Lt})}{\partial \xi_t} = \frac{\xi_t}{1 + \lambda} \frac{(\psi^s + \psi^u)(\psi^s - \psi^u)}{\left[\psi^s + \frac{\xi_t^2}{1 + \lambda}\psi^u\right]^2} < 0,$$

since $\psi^s < \psi^u$. This analytical result is consistent with our quantitative finding, indicating that as ξ_t is increasing over time as a result of secular ESTP, the planner implements an increasing marginal tax rate τ_{Lt} over time to remedy the worsening distribution between the skilled and the unskilled. Note that we would have a constant τ_L over time at the optimum as in [Werning \(2007\)](#) if there were no capital-skill complementarity. This would-be result arises because the skill premium ξ_t in (33) would remain unchanged over time. Note also that if $\psi^s = \psi^u$ were to hold (i.e., there were no motive for the planner's redistribution toward the unskilled), we would have $\frac{\partial(1 - \tau_{Lt})}{\partial \xi_t} = 0$ at the optimum.

We also obtain¹⁷

$$1 - \tau_{Ket+1} = \frac{1}{1 + \frac{x_{t+1}}{\frac{1}{\beta q_t} - \frac{1}{q_{t+1}}}}, \quad (34)$$

where

$$\begin{aligned} x_{t+1} &= \frac{2}{\psi^s + \psi^u} \Upsilon_{t+1} N_{ut+1} \frac{\partial \xi_{t+1}}{\partial K_{et+1}} \\ &= \left(\frac{(1 - \mu)(1 - \lambda)}{\mu}\right)^{\frac{1}{\lambda}} \frac{\lambda \mu}{1 + \lambda} \frac{\psi^u - \psi^s}{\xi_{t+1}^{-1 + \frac{1}{\lambda}} \psi^s + \frac{1}{1 + \lambda} \xi_{t+1}^{1 + \frac{1}{\lambda}} \psi^u}. \end{aligned} \quad (35)$$

The planner's Euler equation is given by¹⁸

$$\frac{1}{q_t} = \beta \left[\frac{1}{q_{t+1}} + r_{et+1} - x_{t+1} \right]. \quad (36)$$

The optimal tax formula (34) is derived from the alignment of the household's Euler equation (32) with the planner's.

The presence of the term $x_{t+1} > 0$ in (34) involves a feature analogous to the so-called pecuniary externalities: there exists a positive effect of ESTP on the skill premium

¹⁷ According to (34) and (35), $\tau_{Ket+1} > 0$ at the optimum requires both $\Upsilon_{t+1} > 0$ (the skilled and the unskilled face the same marginal labor tax rate τ_{Lt}) and $\frac{\partial \xi_{t+1}}{\partial K_{et+1}} > 0$ (capital-skill complementarity). This corresponds to the result shown numerically by [Jones et al. \(1997\)](#).

¹⁸ It is derived from (24) under the simplified model.

due to capital-skill complementarity (see (35)), but the households fail to internalize this effect in their decision on the accumulation of capital equipment. This failure to internalize results in overaccumulation of capital equipment in the laissez faire equilibrium.¹⁹ To correct the overaccumulation, the planner imposes $\tau_{Ket+1} > 0$ to enforce $r_{et+1} > (\frac{1}{\beta q_t} - \frac{1}{q_{t+1}})$ in (32) and induce a higher r_{et+1} than the laissez faire equilibrium. Note that we would have $\tau_{Ket+1} = 0$ at the optimum as in [Werning \(2007\)](#) if there were no capital-skill complementarity. No capital-skill complementarity would imply $\frac{\partial \xi_{t+1}}{\partial K_{et+1}} = 0$, which would in turn imply the absence of “pecuniary externalities” with $x_{t+1} = 0$ according to (35).

Given that the price for a unit of capital equipment is equal to $\frac{1}{q_t}$ at time t , the term $\frac{1}{\beta q_t} - \frac{1}{q_{t+1}}$ in (34) represents the extent to which ESTP obsolesces the old capital equipment. We have from (34):

$$\tau_{Ket+1} \approx \frac{x_{t+1}}{\frac{1}{\beta q_t} - \frac{1}{q_{t+1}}}. \quad (37)$$

The alignment of the household’s Euler equation (32) with the planner’s (36) also gives

$$\tau_{Ket+1} = \frac{x_{t+1}}{r_{et+1}}. \quad (38)$$

The price mechanism does not fully work in the presence of externalities. To correct the price mechanism, equation (38) shows that the planner imposes an ad valorem tax (with tax rate τ_{Ket+1}) on the price r_{et+1} to internalize the “pecuniary externalities” x_{t+1} . Contrasting (38) with (37) reveals that the price r_{et+1} in our setting is basically determined by the term $\frac{1}{\beta q_t} - \frac{1}{q_{t+1}}$, i.e., the extent to which ESTP obsolesces the old capital equipment. In fact, in the laissez faire equilibrium with $\tau_{Ket+1} = 0$, the equality $r_{et+1} = \frac{1}{\beta q_t} - \frac{1}{q_{t+1}}$ holds exactly; see (32). To summarize, we have the result: as the extent to which ESTP obsolesces the old capital equipment increases (decreases) so as to make the price r_{et+1} become higher (lower), the imposed ad valorem tax rate τ_{Ket+1} should be adjusted lower (higher) in order to internalize a given level of the “pecuniary externalities” x_{t+1} .

Note that $q_{t+1} = q_t(1 + \frac{\Delta q_{t+1}}{q_t})$ and, hence, $d(\frac{\Delta q_{t+1}}{q_t})/dt > 0$ implies $d(\frac{1}{\beta q_t} - \frac{1}{q_{t+1}})/dt > 0$. Thus, given x_{t+1} , (37) prescribes that $d(\tau_{Ket+1})/dt < 0$ if $d(\frac{\Delta q_{t+1}}{q_t})/dt > 0$, but $d(\tau_{Ket+1})/dt > 0$ if $d(\frac{\Delta q_{t+1}}{q_t})/dt < 0$. This explains our quantitative finding that τ_{Ket+1} should be adjusted downward or upward according to whether ESTP takes place at an accelerated pace (i.e., $\frac{\Delta q_{t+1}}{q_t}$ is increasing over time) or at a decelerated pace (i.e., $\frac{\Delta q_{t+1}}{q_t}$ is decreasing over time).

Figure 3 shows that while increasing $\{\frac{\Delta q_{t+1}}{q_t}\}$ is associated with a significant downward adjustment of optimal τ_{Ket+1} , decreasing $\{\frac{\Delta q_{t+1}}{q_t}\}$ is associated with only a modest upward adjustment of optimal τ_{Ket+1} . The result of $\frac{\partial \tau_{Ket+1}}{\partial \xi_{t+1}} < 0$ at the optimum

¹⁹ See [Dávila et al. \(2012\)](#) for more on pecuniary externalities. Changes in the amount of aggregate savings influence the return to capital and the wage and, hence, may help or hurt the low-wealth poor in incomplete-market economies. The paper finds that, depending on the factor composition of the income of the low-wealth poor, over- or underaccumulation of capital can arise by the utilitarian criterion.

(through $\frac{\partial \tau_{Ket+1}}{\partial x_{t+1}} > 0$ and $\frac{\partial x_{t+1}}{\partial \xi_{t+1}} < 0$ according to (34)-(35)) explains why this asymmetry arises.²⁰ As ξ_t is increasing over time as a result of secular ESTP, this result reinforces the increasing $\left\{ \frac{\Delta q_{t+1}}{q_t} \right\}$ on the downward adjustment of optimal τ_{Ket+1} , but it offsets the decreasing $\left\{ \frac{\Delta q_{t+1}}{q_t} \right\}$ on the upward adjustment of optimal τ_{Ket+1} . Our quantitative result shows that the direct effect of ESTP on τ_{Ket+1} via variation in $\frac{1}{\beta q_t} - \frac{1}{q_{t+1}}$ dominates its induced effect on τ_{Ket+1} via variation in x_{t+1} .

We find that the planner employs $\{\tau_{Ket+1}\}$ to internalize the “pecuniary externalities” caused by ESTP and $\{\tau_{Lt}\}$ to remedy the worsening distribution resulting from ESTP. However, it should be noted that the only reason there is a need to correct the pecuniary externalities is because the planner wants to redistribute. Indeed, it is clear from (35) that if $\psi^u = \psi^s$ were to hold, there would be no motive for the planner’s redistribution toward the unskilled and we would have $x_{t+1} = 0$ and thereby $\tau_{Ket+1} = 0$ according to (34). Note also that τ_{Ket+1} functions as an ad valorem tax is important. If τ_{Ket+1} somehow were to function as a unit tax and be independent of changes in the price r_{et+1} , it is clear that optimal τ_{Ket+1} would not display the feature we have just described.

5. MIRRLEES PROBLEM

We turn to the Mirrlees approach and formulate the Mirrlees problem in this section. Unlike the Ramsey approach, there are no *a priori* restrictions placed on the tax scheme except that taxes imposed cannot be conditioned directly on household types. As noted by Werning (2007) and Kocherlakota (2010, p.3), this restriction ends up implying that the planner must de facto treat households as being privately informed about their own skills even if this may not be true. Following Mirrlees (1971) and most of the subsequent literature, individual labor income earned, $w_t n_t$, and consumption, c_t , (and so saving) are assumed to be observable by the planner. Since household types are either skilled or unskilled, our model is basically a dynamic extension of Stiglitz (1982) with the inclusion of capital-skill complementarity and ESTP.²¹

5.1 Incentive compatibility constraints

Following the Mirrlees approach, we consider a social planner (a benevolent government) who offers each household a contract with commitment.²² As is standard, households are assumed to have no outside opportunities available in the face of taxation.

²⁰ It can be shown that $\frac{\partial x_{t+1}}{\partial \xi_{t+1}}$ has the same sign as $\frac{\partial^2 \xi_{t+1}}{\partial K_{et+1}^2}$. Since $\frac{\partial^2 \xi_{t+1}}{\partial K_{et+1}^2} < 0$ so that the law of diminishing returns applies, we have $\frac{\partial x_{t+1}}{\partial \xi_{t+1}} < 0$.

²¹ In Stiglitz (1982, equation (16a)), the skill premium ξ depends on $\frac{N_u}{N_s}$ and the substitutability between N_u and N_s . Here ξ also depends on $\frac{K_e}{N_s}$ and the substitutability between K_e and N_s relative to that between K_e and N_u (captured by ρ relative to σ); see (11).

²² As time passes, information about household types (skilled or unskilled) may be revealed. The social planner is assumed to commit to the contract without exploiting information revelation as time passes. It is known that the society is better off with such a commitment than without it; see, for example, Laffont and Tirole (1988).

Since whether households are skilled or unskilled is the private information of households, the revelation principle allows us to restrict attention to contracts with a direct mechanism that relies on truthful reports of the households' types. Thus, each household reports its skill type and receives an allocation $\{c_{it}, n_{it}\}$ as a function of this report such that the allocation is required to satisfy the incentive-compatibility (IC) constraints:

$$V(s) \geq V^u(s), \quad (39)$$

$$V(u) \geq V^s(u), \quad (40)$$

with

$$V^u(s) = \sum_{t=0}^{\infty} \beta^t U(c_{ut}, 1 - \frac{n_{ut}}{\xi_t}),$$

$$V^s(u) = \sum_{t=0}^{\infty} \beta^t U(c_{st}, 1 - \xi_t n_{st}),$$

where $V^j(i)$ with $i \neq j$ denotes the type i household's level of utility derived from deceptively mimicking the type j household. The social planner by assumption can observe a household's earnings but cannot tell whether the household in question is skilled or unskilled. As a result, the minimal labor input that must be expended for the skilled (the unskilled) to mimic the other type is equal to $\frac{w_{ut}n_{ut}}{w_{st}} = \frac{n_{ut}}{\xi_t}$ (resp. $\frac{w_{st}n_{st}}{w_{ut}} = \xi_t n_{st}$), which appears in $V^u(s)$ (resp. $V^s(u)$).

The IC constraints characterized by (39)-(40) are key elements of the Mirrlees approach. The inequality $V(s) \geq V^u(s)$ of (39) dictates that the skilled weakly prefer the allocation designated for them to that for the unskilled. Likewise, the inequality $V(u) \geq V^s(u)$ of (40) dictates that the unskilled weakly prefer the allocation designated for them to that for the skilled. We focus on the normal case where $V(s) = V^u(s)$ and $V(u) > V^s(u)$; namely, the skilled mimic the unskilled rather than the other way around. We confirm numerically that the normal case holds in our study.

5.2 Constrained efficient allocation

Our numerical solutions focus on solving for constrained efficient allocations and their implied optimal marginal tax rates. The so-called "constrained" is attributed to the presence of the IC constraints (39)-(40) relative to their absence.

Given $\{K_{e0}, K_{s0}\}$, the Mirrlees problem is to choose the allocation $\{c_{it}, n_{it}, K_{st+1}, K_{et+1}\}$ with $i \in \{s, u\}$ so as to maximize an *SWF* defined in (14), subject to the resource constraints (13) and the IC constraint (39) with equality. Kocherlakota (2010, section 3.2) showed that the set of allocations $\{c_{it}, n_{it}, K_{st+1}, K_{et+1}\}$ that are achievable by the society under the Mirrlees approach is exactly the one that satisfies the IC constraints and the resource constraints.

Let $U(c, 1 - n) = u(c) + v(1 - n)$, and let Λ and $\{\beta^t \Gamma_t\}$ be the multipliers on the IC constraint (39) and resource constraints (13), respectively. Then the FOCs for the planner problem are given by

$$c_{it} : u'(c_{st}) = \frac{\Gamma_t}{\psi^s + \Lambda}, \quad u'(c_{ut}) = \frac{\Gamma_t}{\psi^u - \Lambda}, \quad (41)$$

$$n_{ut} : \psi^u v'(1 - n_{ut}) = \Gamma_t w_{ut} + \Lambda v'(1 - \frac{n_{ut}}{\xi_t}) \left[\frac{1}{\xi_t} - \frac{n_{ut}}{\xi_t^2} \frac{\partial \xi_t}{\partial n_{ut}} \right], \quad (42)$$

$$n_{st} : [\psi^s + \Lambda] v'(1 - n_{st}) = \Gamma_t w_{st} - \left[\Lambda v' \left(1 - \frac{n_{ut}}{\xi_t} \right) \frac{n_{ut}}{\xi_t^2} \right] \frac{\partial \xi_t}{\partial n_{st}}, \quad (43)$$

$$K_{st+1} : \Gamma_t = \beta \Gamma_{t+1} [F_{K_{st+1}} + (1 - \delta_s)], \quad (44)$$

$$K_{et+1} : \frac{\Gamma_t}{q_t} = \beta \left\{ \Gamma_{t+1} \left[r_{et+1} + \frac{(1 - \delta_e)}{q_{t+1}} \right] - \Delta_{t+1} \right\}, \quad (45)$$

where $\Delta_{t+1} = \left[\Lambda v' \left(1 - \frac{n_{ut+1}}{\xi_{t+1}} \right) \frac{n_{ut+1}}{\xi_{t+1}^2} \right] \frac{\partial \xi_{t+1}}{\partial K_{et+1}}$. Note that $\Delta_{t+1} > 0$ because $\frac{\partial \xi_{t+1}}{\partial K_{et+1}} > 0$.

The role of $\frac{\partial \xi_{t+1}}{\partial K_{et+1}} > 0$ in the FOC (45) of Mirrleesian taxation is in essence the same as the role of $\frac{\partial \xi_{t+1}}{\partial K_{et+1}} > 0$ in the FOC (24) of Ramsey taxation. They both stem from ESTP with capital-skill complementarity. However, $\Upsilon_{t+1} > 0$ in (24) (the skilled and the unskilled face the same marginal labor tax rate) has been replaced by $\Lambda > 0$ in (45) (the binding of the IC constraint). This replacement is natural in light of the observation that IC constraints are crucial in Mirrleesian taxation and that the IC constraints (39)-(40) directly depend on ξ_t .

We denote the constrained efficient allocation resulting from (41)-(45) plus the resource constraints (13) and the equality of (39) by $\{c_i^*, n_i^*, K_s^*, K_e^*\}$ with $i \in \{s, u\}$. Following the idea of Slavík and Yazıcı (2014), one can implement the constrained efficient allocation $\{c_i^*, n_i^*, K_s^*, K_e^*\}$ as part of a competitive equilibrium via a time-varying flat tax rate on capital structures income $\tau_{K_s}(t + 1)$, a time-varying flat tax rate on equipment capital income $\tau_{K_e}(t + 1)$, and a time-varying nonlinear tax schedule on labor income, in which $\tau_s(t)$ and $\tau_u(t)$ are, respectively, the implicit marginal tax rates on the labor incomes of the skilled and the unskilled.²³

The optimal intertemporal condition for capital structures (44) is standard and identical to that in the laissez-faire competitive equilibrium or to the one without the imposition of IC constraints. This immediately implies that saving decisions on capital structures are not distorted at the optimum and hence we have the tax rate $\tau_{K_s}(t + 1) = 0$ all the time. As such, like Ramsey taxation, we confine our analysis to the taxation of equipment capital income and labor income.

²³ A tax system is said to implement the constrained efficient allocation if the constrained efficient allocation arises as part of a competitive equilibrium under the tax system.

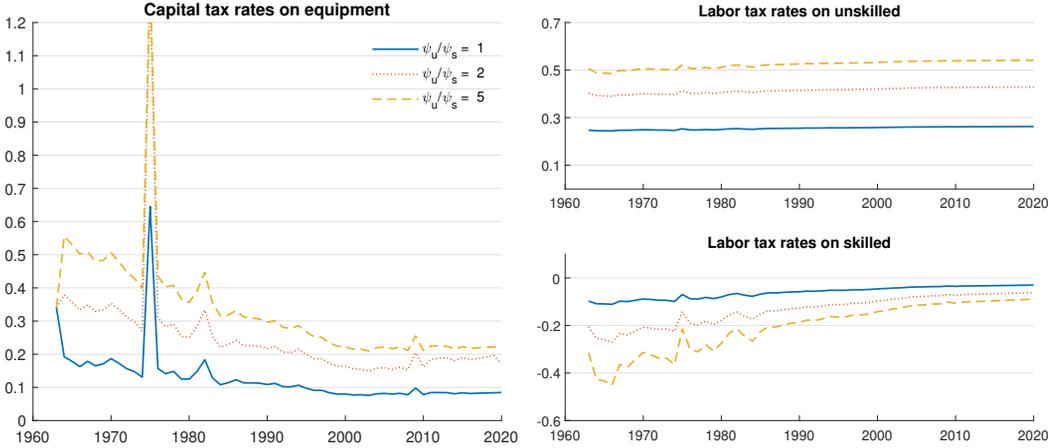


Figure 5.: Mirrleesian taxation

In the next section, we first quantitatively characterize the dynamics of Mirrleesian taxation resulting from the constrained efficient allocation and then consider a simplified model to explain the mechanism underlying the quantitative results.

6. DYNAMICS OF MIRRLEESIAN TAXATION

This section reports the dynamics of Mirrleesian taxation in the face of ESTP $\{q_t\}$.

6.1 Quantitative results

To facilitate comparison, we use the same functions and calibrations as in Ramsey taxation.

Given q_t at a point in time, we find that $\tau_{Ke}(t+1) > 0$, $\tau_s(t) < 0$, and $\tau_u(t) > 0$ at the optimum; see Figure 5.²⁴ Like the Ramsey problem, our central question for the Mirrlees problem is: how will this optimal tax structure at a point in time vary over time in the face of an increasing q_t ?

Figure 5 reports the dynamics of the optimal tax rate on equipment capital income, $\tau_{Ke}(t+1)$, and that of the optimal marginal tax rates on skilled and unskilled labor, $\tau_s(t)$ and $\tau_u(t)$, for $\psi^u/\psi^s = 1, 2$ and 5.²⁵ Two features stand out. First, $\{\tau_{Ke}(t+1)\}$ basically

²⁴ The regressive labor taxation at the margin (i.e., $\tau_s(t) < 0$ and $\tau_u(t) > 0$) at the optimum is consistent with the finding in Stiglitz (1982). Note also that $\tau_{Ks}(t+1) = 0$ and $\tau_{Ke}(t+1) > 0$ is consistent with the finding in Slavík and Yazici (2014) that capital equipment should be taxed at a higher rate than capital structures.

²⁵ Figure 5 shows that, given q_t at a point in time, both $\tau_{Ke}(t+1) > 0$ and $\tau_u(t) > 0$ become higher while $\tau_s(t) < 0$ becomes lower at the optimum as ψ^u/ψ^s increases. This result is intuitive. As ψ^u/ψ^s increases, the planner assigns a higher allocation of consumption and leisure to the unskilled household relative to the skilled. This assignment induces a stronger incentive for the skilled to mimic the unskilled. To meet the IC constraint, $V(s) \geq V^u(s)$, the planner needs to further depress the skill premium at the given q_t . This explains why both $\tau_{Ke}(t+1) > 0$ and $\tau_u(t) > 0$ become higher while $\tau_s(t) < 0$ becomes lower at the optimum, since these tax adjustments will all depress the skill premium.

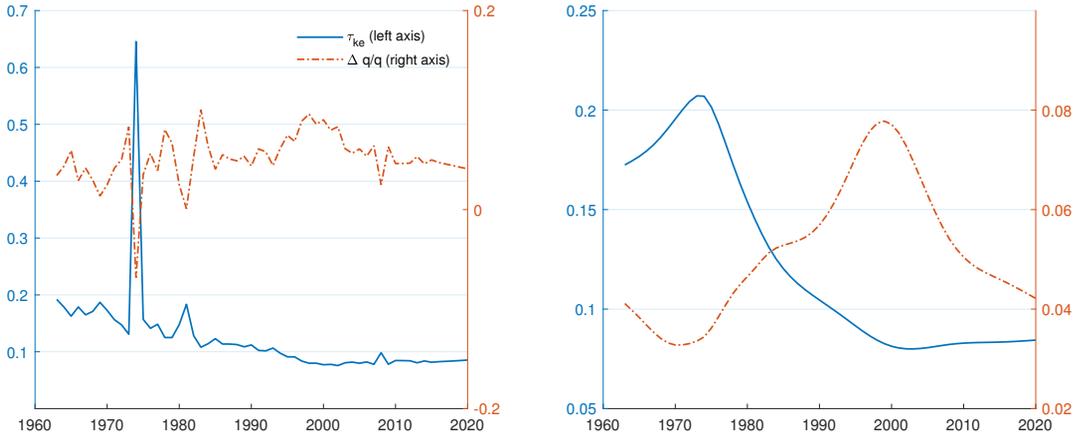


Figure 6.: Capital tax rates on equipment and pace of ESTP

displays a declining trend over time before 1999; however, this declining trend is reversed and turns into an increasing trend over time after 1999 (although this increasing feature of the trend is mild). Second, both $\{\tau_s(t)\}$ and $\{\tau_u(t)\}$ display an increasing trend over time.

The left panel of Figure 6 plots the trajectories of $\{\frac{\Delta q_{t+1}}{q_t}\}$ and optimal $\{\tau_{Ke}(t+1)\}$ as $\psi^u/\psi^s = 1$. The right panel of Figure 6 shows the smoothed-curve representation of the trajectories after applying the Hodrick-Prescott filter. Taken together, the figure shows that, after the oil crisis of 1973-74,²⁶ while optimal $\{\tau_{Ke}(t+1)\}$ are adjusted downward as ESTP takes place at an accelerated pace (i.e., $\frac{\Delta q_{t+1}}{q_t}$ is increasing over time before 1999), they are adjusted upward as ESTP takes place at a decelerated pace (i.e., $\frac{\Delta q_{t+1}}{q_t}$ is decreasing over time after 1999). This salient characteristic of optimal $\{\tau_{Ke}(t+1)\}$ remains robust with variation in Pareto weights $\psi^u/\psi^s \geq 1$. Thus, like Ramsey taxation, Mirrleesian taxation prescribes that tax rates on capital equipment should be adjusted downward or upward according to whether ESTP takes place at an accelerated or a decelerated pace.

Werning (2007, pp. 927-928) prescribed the following Mirrleesian labor taxation: “workers should face different marginal tax rates but that these should remain perfectly constant over time and unresponsive to shocks.” We find that the skilled and the unskilled should face different marginal tax rates as in Werning (2007) but that these marginal tax rates should be adjusted upward over time rather than permanently remain constant. We explain this result later.

6.2 What if there is a different σ ?

How robust are the results shown in Figure 5? To answer the question, we vary the value of σ between the two extremes, $\sigma = \rho$ and $\sigma = 1$, with the objective of checking the robustness of the pattern of Mirrleesian taxation in the face of ESTP as shown in Figure 5.

²⁶ Footnote 15 equally applies here.

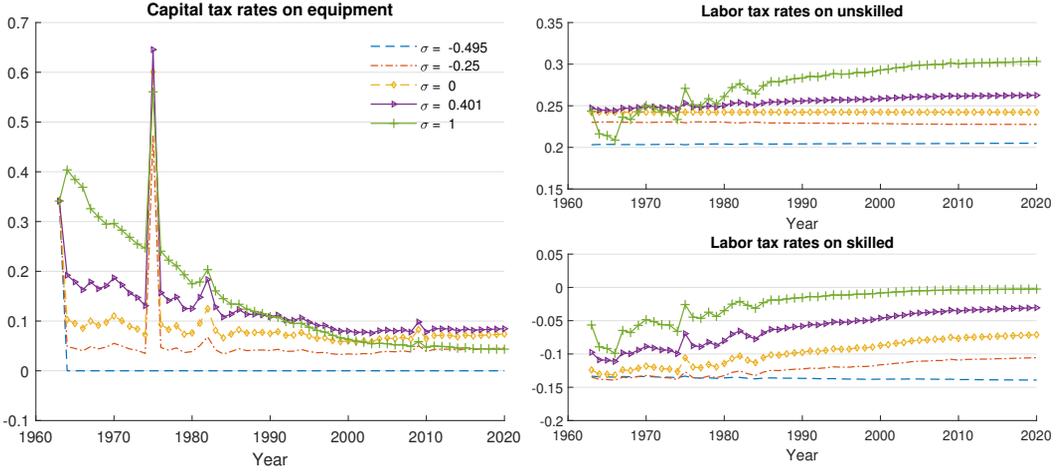
Figure 7.: Mirrleesian taxation under different σ

Figure 7 reports the results. Except for the extreme case where $\sigma = \rho = -0.495$, the pattern of Mirrleesian taxation shown in Figure 7 basically remains the same as that shown in Figure 5.

6.3 A simplified model

This subsection analytically considers a simplified model as a complement to our quantitative study. The setting is identical to that in Ramsey taxation. We seek to explain the mechanism underlying the quantitative results.

With Mirrleesian taxation, the household FOC (31) is replaced with

$$1 - \tau_s(t) = \chi \frac{N_{st}}{w_{st}}; 1 - \tau_u(t) = \chi \frac{N_{ut}}{w_{ut}}, \quad (46)$$

while the household FOC (32) remains the same, except for the notation replacement of τ_{Ket+1} with $\tau_{Ke}(t+1)$. We formulate the Mirrlees problem and derive the resulting FOCs in Online Appendix B.

We obtain

$$1 - \tau_u(t) = \frac{(\psi^u + \psi^s)/2}{\psi^u - \frac{\psi^u - \psi^s}{2\xi_t^2}}, \quad (47)$$

which implies $\tau_u(t) > 0$ and $\frac{\partial \tau_u(t)}{\partial \xi_t} > 0$. We obtain $\tau_s(t) < 0$ with

$$1 - \tau_s(t) = \frac{1}{2} \left[1 + \left(1 + 4 \frac{(\psi^u - \psi^s)\lambda(1 - \tau_u(t))^2}{(\psi^u + \psi^s)\xi_t^4} \right)^{\frac{1}{2}} \right]. \quad (48)$$

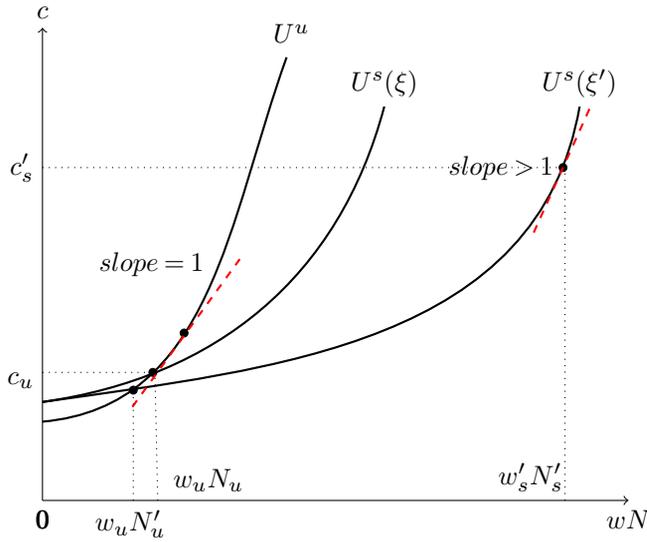


Figure 8.: $N'_u > N_u$ as $\xi' > \xi$

Using $1 - \tau_u(t)$ from (47), we have $\frac{\partial \tau_s(t)}{\partial \xi_t} > 0$. The analytical result that $\frac{\partial \tau_u(t)}{\partial \xi_t} > 0$ and $\frac{\partial \tau_s(t)}{\partial \xi_t} > 0$ is consistent with our quantitative finding that both $\tau_u(t)$ and $\tau_s(t)$ at the optimum are increasing over time. What is the underlying reason for the result?

First, regardless of (47) or (48), it is observed that the impact of ξ_t on $\tau_u(t)$ and that on $\tau_s(t)$ both invoke the term associated with $(\psi^u - \psi^s)$, which happens to be the multiplier of the IC constraint (see (B.13) in Online Appendix B). This suggests that the underlying reason for the increasing $\tau_u(t)$ and $\tau_s(t)$ has to do with the IC constraint. Note that if $\psi^s = \psi^u$ were to hold, the IC constraint would not bind and we would have $\tau_u(t) = \tau_s(t) = 0$ according to (47) and (48). This result is intuitive: in the absence of a redistributive motive, it is optimal for the Mirrlees planner to simply impose uniform lump-sum taxation rather than nonlinear taxation on labor income in our setting of the quilinear utility. Second, the relative quantity effect is absent according to (30) in our simplified model. It implies that we cannot appeal to the relative quantity effect to explain the result of $\frac{\partial \tau_u(t)}{\partial \xi_t} > 0$. It also implies that we must rely on the capital-skill complementarity effect to explain the result of $\frac{\partial \tau_s(t)}{\partial \xi_t} > 0$.

In the case of $\tau_u(t)$, lowering N_{ut} from the no-distortion level with $\tau_u(t) > 0$ at the optimum is a well-known result; see Stiglitz (1982). Note that the skilled wage rate $w_{st} = \mu \xi_t$ (see (29)), while the unskilled wage rate $w_{ut} = \mu$ (see (28)). As ESTP causes an increase in the skill premium ξ_t , w_{st} increases in ξ_t but w_{ut} remains unchanged. This implies that the indifference curve (in the (wN, c) plane) for the skilled becomes flatter, but that for the unskilled remains the same; see Figure 8. As a result, to satisfy the IC constraints, it is optimal to lower N_{ut} further with ESTP. This explains why $\frac{\partial \tau_u(t)}{\partial \xi_t} > 0$ at the optimum.

In the case of $\tau_s(t)$, subsidizing N_{st} with $\tau_s(t) < 0$ at the optimum is also a well-known result; see Stiglitz (1982). Over time, the planner could choose between lowering and raising the subsidy. Suppose that the planner chooses a higher subsidy so that, all

else equal, N_{st} becomes higher over time. This will induce a higher K_{et} over time, since N_{st} and K_{et} are complementary inputs and a higher N_{st} enhances the marginal product of K_{et} . The higher K_{et} results in a higher K_{et}/N_{st} because of capital-skill complementarity and, hence, a higher skill premium through the capital-skill complementarity effect. In the end, it will tighten the IC constraint and lead to greater distortion of allocations. By contrast, if the planner chooses a lower subsidy so that, all else equal, N_{st} becomes lower over time, the opposite result occurs: the IC constraint will be relaxed and allocations will be less distorted. This explains why $\frac{\partial \tau_s(t)}{\partial \xi_t} > 0$ at the optimum.

We also obtain

$$1 - \tau_{Ke}(t+1) = \frac{1}{1 + \frac{y_{t+1}}{\frac{1}{\beta q_t} - \frac{1}{q_{t+1}}}}, \quad (49)$$

where

$$\begin{aligned} y_{t+1} &= \frac{\psi^u - \psi^s}{\psi^u + \psi^s} \chi \frac{N_{ut+1}^2}{\xi_{t+1}^3} \frac{\partial \xi_{t+1}}{\partial K_{et+1}} \\ &= \frac{\mu \lambda (\psi^u - \psi^s)}{(\psi^u + \psi^s)} \left(\frac{(1-\lambda)(1-\mu)}{\mu} \right)^{\frac{1}{\lambda}} \frac{(\psi^u + \psi^s)}{\xi_{t+1}^{1/\lambda} [2\xi_{t+1}^2 \psi^u - (\psi^u - \psi^s)]} \frac{N_{ut+1}}{N_{st+1}}. \end{aligned} \quad (50)$$

Except for replacing x_{t+1} with y_{t+1} , there is no difference in essence between the formula (49) for Mirrleesian capital taxation and the formula (34) for Ramsey capital taxation. As in Ramsey taxation, if there were no capital-skill complementarity, we would have $y_{t+1} = 0$ and hence $\tau_{Ke}(t+1) = 0, \forall t \geq 0$ at the optimum according to (49)-(50).

7. EXTENSION

In the benchmark model we focus on how ESTP $\{q_t\}$ with capital-skill complementarity shapes the dynamics of optimal taxation. For the sake of focus, we assume that the skilled and the unskilled have the same size all the time and that the labor productivities of the skilled and unskilled are both equal to unity. These assumptions are relaxed in the extension.

The economy now consists of heterogeneous households identified by their innate talent θ . Ales et al. (2015) considered a talent-to-task assignment model *à la* Teulings (1995), Costinot and Vogel (2010), and Acemoglu and Autor (2011). Following the setting in their quantitative study, we pose:

1. Talent θ is distributed uniformly over $[0, 1]$. To facilitate our numerical analysis, we use $\Theta = \{\theta_j\}_{j=1}^M$ to approximate $\theta \in [0, 1]$ and let $\pi(\theta_j) = 1/M$ denote the fraction of talent θ_j households in the population.
2. The match of talent θ to task $i \in \{s, u\}$ gives rise to labor productivity $z_i(\theta) = \exp(a_i + b_i \theta)$. We normalize $a_u = b_u = 0$ and let $b_s > 0$ so that $z_s(\theta')/z_s(\theta) >$

$z_u(\theta')/z_u(\theta)$ for all $\theta' > \theta$, that is, talent θ' (a high type) relative to talent θ (a low type) has a comparative advantage in performing task $i = s$.²⁷

3. Each household chooses either $i = s$ or $i = u$ to work in.

Instead of $i \in \{s, u\}$, the model of Ales et al. (2015) features a continuum of tasks. They mainly conducted the comparative-statics analysis of Mirrleesian taxation with respect to technical change in $\{a_i, b_i\}$ in the U.S. economy between the two periods – the 1970s and 2000s. However, since they abstracted from modeling capital explicitly, there is no ESTP with capital-skill complementarity in their analysis.

In the extended model, the FOCs for the representative firm are still given by $\frac{\partial Y}{\partial K_s} = r_s$, $\frac{\partial Y}{\partial K_e} = r_e$, $\frac{\partial Y}{\partial N_s} = w_s$, and $\frac{\partial Y}{\partial N_u} = w_u$. For each unit of “raw” labor supply, those θ who choose $i = s$ earn wage $w_s z_s(\theta)$ while those θ who choose $i = u$ earn wage $w_u z_u(\theta)$. It is clear that the task i chosen by talent θ will be $i = \arg \max\{w_i z_i(\theta)\}_{i \in \{s, u\}}$. Let $\hat{\theta}$ denote the threshold such that those θ with $\theta < \hat{\theta}$ choose $i = u$ while those θ with $\theta \geq \hat{\theta}$ choose $i = s$. Let Θ_i denote the subset of Θ with $i \in \{s, u\}$.²⁸ The aggregate inputs K_s, K_e, N_s and N_u need to be redefined:

$$K_s = \sum_{\theta_j \in \Theta} k_s(\theta_j) \pi(\theta_j), K_e = \sum_{\theta_j \in \Theta} k_e(\theta_j) \pi(\theta_j),$$

$$N_s = \sum_{\theta_j \in \Theta_s} z_s(\theta_j) n_s(\theta_j) \pi(\theta_j), N_u = \sum_{\theta_j \in \Theta_u} z_u(\theta_j) n_u(\theta_j) \pi(\theta_j) = \sum_{\theta_j \in \Theta_u} n_u(\theta_j) \pi(\theta_j),$$

where $k_s(\theta_j)$ and $k_e(\theta_j)$ represent the capital structures and capital equipment supplies by θ_j households, and $z_i(\theta_j) \pi(\theta_j) n_i(\theta_j)$, $i \in \{s, u\}$, represents the effective labor supply in task i by θ_j households (note that $z_u(\theta_j) = 1$ for all θ_j since we normalize $a_u = b_u = 0$).

The skill premium defined in (10) is modified to become

$$\xi \equiv \frac{w_s}{w_u} z = \frac{(1 - \mu)(1 - \lambda)}{\mu} \left[\lambda \left(\frac{K_e}{N_s} \right)^\rho + (1 - \lambda) \right]^{\frac{\sigma - \rho}{\rho}} \left(\frac{N_u}{N_s} \right)^{1 - \sigma} z,$$

where

$$z = \frac{z_s}{z_u} \text{ with } z_s = \frac{N_s}{\sum_{\theta_j \in \Theta_s} n_s(\theta_j) \pi(\theta_j)}, z_u = \frac{N_u}{\sum_{\theta_j \in \Theta_u} n_u(\theta_j) \pi(\theta_j)} = 1.$$

²⁷ Given that $z_i(\theta) = \exp(a_i + b_i \theta)$, $i \in \{s, u\}$ and that $a_u = b_u = 0$, $z_s(\theta')/z_s(\theta) > z_u(\theta')/z_u(\theta)$ for all $\theta' > \theta$ if and only if $b_s > 0$.

²⁸ Let $\Theta = \{\theta_1, \theta_2\}$ and

$$z_s(\theta_2) = z_u(\theta_1) = 1,$$

$$z_s(\theta_1) = z_u(\theta_2) = 0.$$

Talent θ_2 will always choose task s , and talent θ_1 will always choose task u ; their labor productivities are both equal to unity. We are back to the benchmark model.

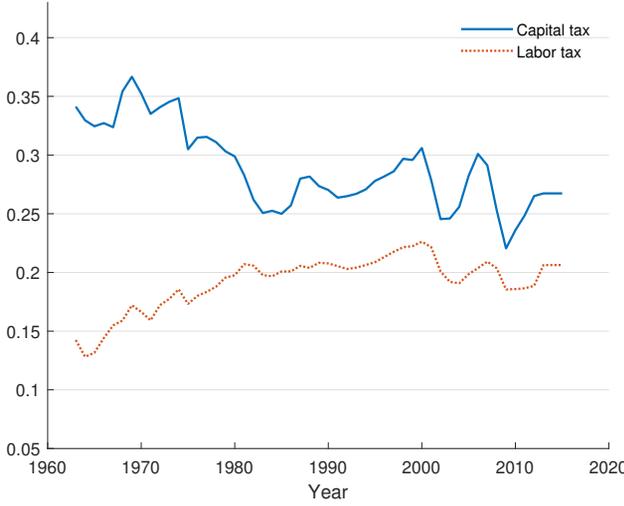


Figure 9.: Tax rates in the U.S. calculated by McDaniel (2007)

We then obtain

$$\ln \xi \approx \lambda \left(\frac{\sigma - \rho}{\rho} \right) \left(\frac{K_e}{N_s} \right)^\rho + (1 - \sigma) \ln \left(\frac{\sum_{\theta_j \in \Theta_u} n_u(\theta_j) \pi(\theta_j)}{\sum_{\theta_j \in \Theta_s} n_s(\theta_j) \pi(\theta_j) z_s(\theta_j)} \right) + \sigma \ln z + \text{constant}, \quad (51)$$

where the term $\ln z$ is absent in the benchmark model; see equation (11). [Krusell et al. \(2000\)](#) called the third component of (51) the “relative efficiency effect.” Given $\sigma > 0$ (the elasticity of substitution between two types of labor is greater than one in our calibration), an increase in z will raise the skill premium.

In the extended model, besides ESTP $\{q_t\}$, there are two sets of time-series data we need to use in our quantitative study: the tax rates $\{\tau_t\}$ and the skill premium $\{\xi_t\}$. We briefly describe them.

The tax rate series are obtained directly from [McDaniel \(2007\)](#). Her calculation focuses on taxes and ignores transfers. As such, the obtained average tax rates can be viewed as the marginal tax rates of a linear income tax system; see [McDaniel \(2007\)](#) for a formal argument. By adapting a figure in [Heathcote et al. \(2017\)](#), [Bhandari et al. \(2017\)](#) showed that a linear tax schedule can approximate actual tax and transfer programs of the U.S. economy pretty well. The tax series obtained by [McDaniel \(2007\)](#) will be viewed as the data representation of the U.S. tax system; see Figure 9.²⁹

[Acemoglu and Autor \(2011\)](#) used data sources including the March CPS to calculate the college/high-school skill premium for full-time, full-year workers for the period 1963-2008. Their approach is sophisticated in that they managed to hold constant the

²⁹ For the years after 2013, we let the tax rates remain the same as those in 2013 so that the economy can converge to a new steady state. In computing the competitive equilibrium of the U.S. economy, these rates are applied uniformly to both the skilled and the unskilled.

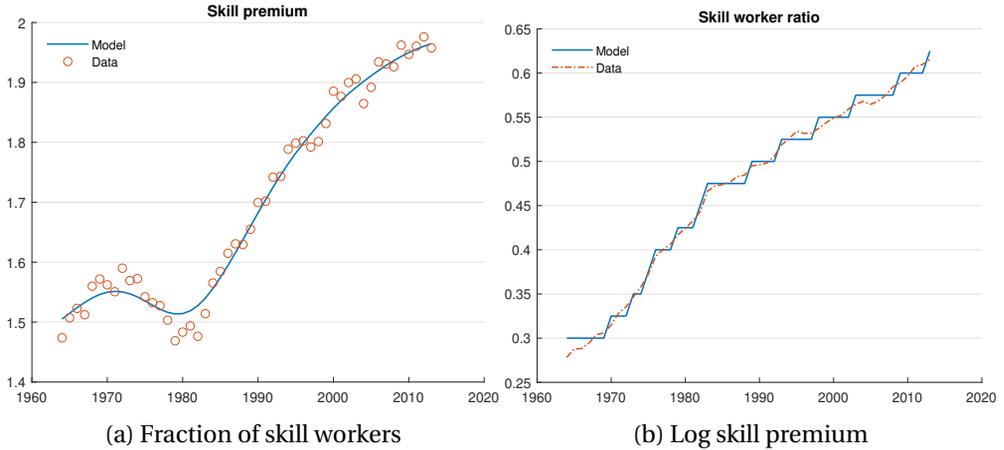


Figure 10.: Model vs. data

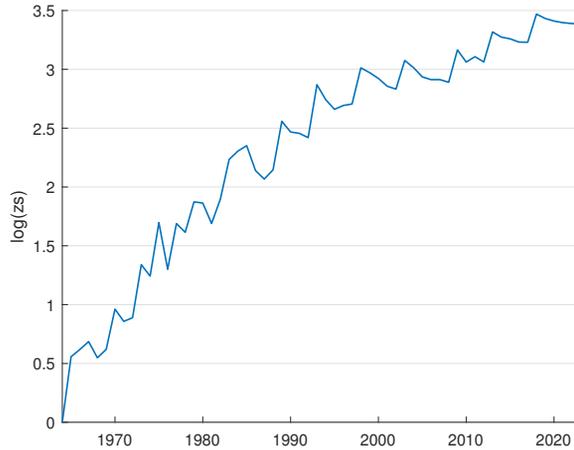


Figure 11.: Labor productivity $\log z$ (1963=1)

relative employment shares of the demographic group (including gender, education, and potential experience) across all years of their sample. Autor (2014) extended the data sequence to the year 2012, which is the data representation in Figure 10.³⁰

All households (indexed by θ_j) in the extended model need to choose task $i \in \{s, u\}$ at the beginning of each time period. Let $\Pi_s(t) \equiv \sum_{\theta_j \in \Theta_s(t)} \pi(\theta_j)$, namely, the fraction of households who choose $i = s$ at time t . From our setup above, we obtain $\hat{\theta}_t = \frac{1}{b_{st}} \left(\log \frac{w_{ut}}{w_{st}} - a_{st} \right)$. Thus, in the face of ESTP $\{q_t\}$, as long as the relative wage rate w_{ut}/w_{st} and the technology parameters a_{st} and b_{st} vary over time, the threshold $\hat{\theta}_t$ and hence the fraction $\Pi_s(t)$ will also vary over time. Given the “deep” parameters of household preferences and production technology as in the benchmark model, we compute the transitional dynamics of the competitive equilibrium of the U.S. economy from the

³⁰ The data are available from Autor’s website.

initial steady state to a new steady state in the extended model. We calibrate the series of the technology parameters, $\{a_{st}, b_{st}\}$, in the extended model to match the series of the skill premium and the fraction of the skilled, $\{\xi_t, \Pi_s(t)\}$, in the data.³¹ Figure 10 reports the match. Note that the match is rather good. Figure 11 reports the resulting $\{\log z_t\}$ from the match.

With $\{a_{st}, b_{st}\}$ at hand, we study how ESTP $\{q_t\}$ with capital-skill complementarity shapes the dynamics of optimal taxation in the extended model. The social welfare function in the extended model becomes

$$SWF = \sum_{\theta_j \in \Theta} \psi(\theta_j) \pi(\theta_j) \sum_{t=0}^{\infty} \beta^t [u(c_t(\theta_j)) + v(1 - n_t(\theta_j))], \quad (52)$$

where $\psi(\theta_j) \geq 0$ with $\theta_j \in \Theta$ are the Pareto weights and the period utility function $u(c) + v(1 - n)$ takes the form of (27). We focus on the case where $\psi(\theta_j) = 1$ for all $\theta_j \in \Theta$, i.e., the utilitarian criterion.

7.1 Dynamics of optimal taxation

7.1.0.1 Ramsey problem There are two types of households in the benchmark model and hence there are two corresponding implementability conditions in the Ramsey problem. In the extended model, the idea of formulating the Ramsey problem is not different from that in the benchmark model, except that all households (indexed by θ_j) now need to choose task $i \in \{s, u\}$ at the beginning of each time period and that there are 40 instead of 2 implementability conditions given that we set $\Theta = \{\theta_j\}_{j=1}^{40}$.

7.1.0.2 Mirrlees problem There are two types of households in the benchmark model and hence there are two IC constraints as given by (39)-(40). Given that $\Theta = \{\theta_j\}_{j=1}^{40}$ in the extended model, the corresponding IC constraints in the Mirrlees problem are more complicated. However, the idea of formulating the Mirrlees problem is not different from the benchmark model in essence. In particular, as in the benchmark model, we first assume that only the local downward IC constraints bind and then verify the validity of this assumption numerically. Of course, like Ramsey taxation, all households (indexed by θ_j) need to choose task $i \in \{s, u\}$ at the beginning of each time period in the extended model.

7.1.0.3 Mirrleesian labor taxation and average marginal tax rate Figure 12 reports the cross-section Mirrleesian labor marginal tax rates for the years 1964, 1988 and 2012 (other years yield similar results). All of the three cross sections display a hump-shaped pattern, showing that middle types of households face higher labor marginal tax rates than either low or high types of households.³² This hump-shaped pattern (in particular, the labor marginal tax rates are positive for the lowest type of households, but they are negative for the highest type of households) is similar to what Ales et al. (2015) found

³¹ We choose $M = 40$ in our match and verify numerically that $b_{st} > 0$ for all t .

³² With our normalization that $a_u = b_u = 0$, unskilled workers are in effect homogeneous. This explains why we obtain the flat part of the Mirrleesian labor marginal tax rate in Figure 12.

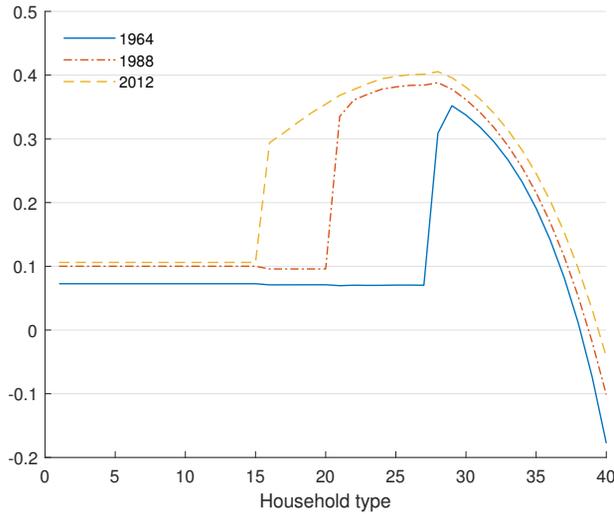


Figure 12.: Mirrleesian labor marginal tax rate

for their cross-section labor marginal tax rates; see their Table 2. The similarity is not surprising, in that our extended model builds on the talent-to-task model of [Ales et al. \(2015\)](#). Although we consider capital explicitly and address ESTP with capital-skill complementarity, there is no difference in essence between the two models as far as the intratemporal aspect is concerned. Figure 12 shows that all households face an increasing labor marginal tax rate over time. This feature is consistent with what we find in the benchmark model, where both the skilled and the unskilled face higher labor marginal tax rates over time in the face of secular ESTP.

All households face a single labor marginal tax rate, τ_{Lt} , at time t in Ramsey labor taxation. However, different types of households earn different levels of labor income and hence face different labor marginal tax rates at time t in Mirrleesian labor taxation. To facilitate comparison with a single labor marginal tax rate in Ramsey labor taxation, we summarize cross-section labor marginal tax rates faced by different households at time t in Mirrleesian labor taxation by a single rate, the so-called average marginal tax rate proposed by [Barro and Sahasakul \(1983\)](#). As they argued, this concept (the weighted average of marginal tax rates, where weights are equal to shares of labor income) is more relevant than the average tax rate in assessing the economic effects of taxation.

Figure 13 compares optimal taxes (including both Ramsey and Mirrlees) with U.S. taxes.³³ Over the period we study, optimal taxation prescribes a declining trend for capital tax rates and an increasing trend for labor marginal tax rates. It is interesting to observe that the pattern of the prescribed trends resembles the empirical decline in capital taxes and the increase in labor taxes in U.S. taxes.

Figure 14 compares the evolution of the skill premium under optimal taxes with that under U.S. taxes. The evolutions are close to each other for Ramsey and U.S. taxes. How-

³³ The labor marginal tax rate is represented by the average marginal tax rate in the case of Mirrleesian labor taxation. Given that both Ramsey and Mirrlees prescribe a zero tax on capital structures, we report optimal capital tax rates only for capital equipment.

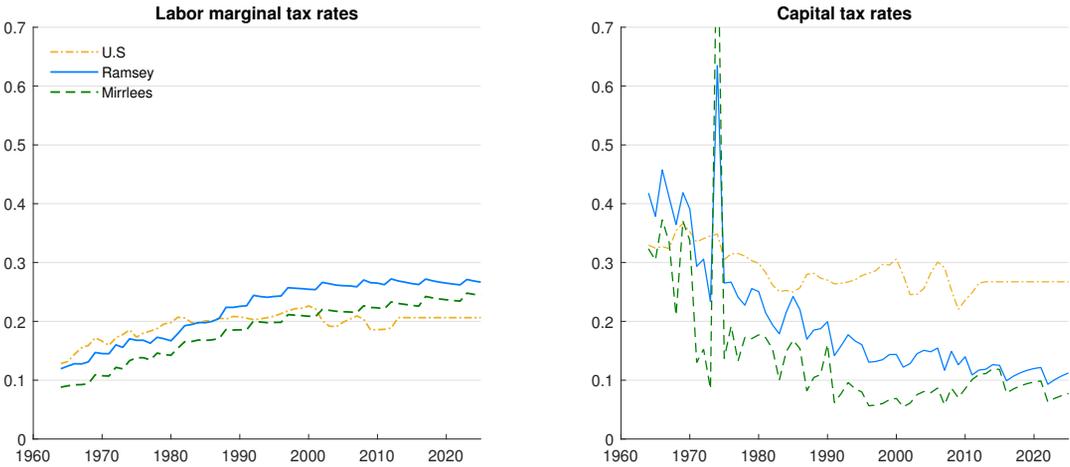


Figure 13.: Optimal tax rates vs. U.S. tax rates

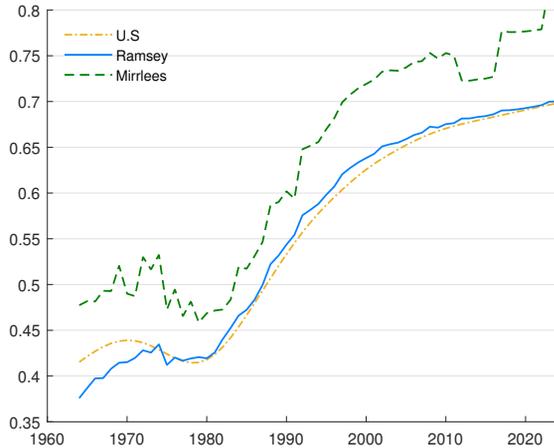


Figure 14.: Log skill premium

ever, optimal Mirrlessian taxes demand a higher skill premium than U.S. taxes. This latter result may not be surprising in view of the observation from Figure 13 that Mirrlessian taxation prescribes a lower level of tax rates than Ramsey taxation in general. Although the fraction of skilled workers is increasing over time as shown in Figure 10, the associated relative quantity effect on the skill premium (the second term of (51)) can be offset by the strong but opposite relative efficiency effect (the third term of (51)) as shown in Figure 11. Overall, Figure 14 reveals that the capital-skill complementarity effect embedded in the first term of (51) still dominates the evolution of the skill premium as in the benchmark model.³⁴

³⁴ It should be emphasized that the results shown in Figures 13 and 14 critically hinge on the utilitarian criterion with $\psi(\theta_j) = 1$ for all $\theta_j \in \Theta$ in the social welfare function (52). Different Pareto weights could lead to the US tax system being closer to the Ramsey and/or Mirrlees solution.

Table 3.: Welfare gains and decomposition

	Ramsey	Mirrlees
Total change (in percent)	0.39	4.22
Consumption	0.77	5.91
Level	0.50	3.23
Distribution	0.27	2.60
Leisure	-0.38	-1.60
Level	-0.42	-0.72
Distribution	0.04	-0.84

7.2 Welfare gains of tax reform

This subsection measures the welfare gains from tax reform in the context of the extended model. That is, we address the following question: given the social welfare function defined by (52), what are the welfare gains of switching from the U.S. tax system to optimal taxation?

To answer the question, we follow Lucas (1987) by considering a measure known as the consumption-equivalent variation (*CEV*) in the literature. Let $\{c, n\}$ denote the consumption-labor allocation resulting from the U.S. tax system. The *CEV* of switching from $\{c, n\}$ to the optimal allocation $\{c^*, n^*\}$ is implicitly defined by

$$SWF(c(1 + CEV), n) = SWF(c^*, n^*),$$

where *SWF* is given by (52) with $\psi(\theta_j) = 1$ for all $\theta_j \in \Theta$.

Following Conesa et al. (2009), we can decompose *CEV* into two parts: one stems from the change in consumption from c to c^* (denoted by CEV_c), and the other stems from the change in labor employment from n to n^* (denoted by CEV_n). It can be verified that $CEV \approx CEV_c + CEV_n$. Following Conesa et al. (2009), we can further decompose CEV_c and CEV_n , respectively, into a level effect and a distribution effect. The level effect measures welfare gains or losses from the change in the aggregate allocation, while the distribution effect measures welfare gains or losses from the redistribution of the allocation across different types of households.

Table 3 reports what we find regarding *CEV* and its decomposition. In the Ramsey case, $CEV = 0.39\%$, of which $CEV_c = 0.77\%$ and $CEV_n = -0.38\%$. Thus, (i) the replacement of the U.S. tax system with optimal Ramsey taxes generates a modest welfare improvement, and (ii) the major part of the welfare gains in terms of *CEV* comes from changes in consumption, not from changes in leisure; in fact, changes in leisure generate welfare loss. Given that the evolution of the skill premium under U.S. taxes closely follows that under optimal Ramsey taxes as shown in Figure 14, it is not surprising to find that there is only a modest improvement in welfare despite the significant rise in the skill premium.

In the Mirrlees case, $CEV = 4.22\%$, of which $CEV_c = 5.91\%$ and $CEV_n = -1.60\%$. Thus, Mirrleesian taxes generate a much more significant welfare improvement from tax

reform than Ramsey taxes. Ramsey taxation imposes a linear labor tax schedule, while Mirrleesian taxation imposes a nonlinear labor tax schedule. The flexibility of the labor tax schedule of Mirrleesian taxation allows it to achieve a higher level of welfare than Ramsey taxation.

8. CONCLUSION

In the simultaneous presence of capital-skill complementarity and the secular decline in the price of capital equipment, how should taxation be set dynamically in response to the rising skill premium? In this paper we attempt to answer the question. Two main results emerge, no matter whether we adopt the Ramsey or the Mirrlees approach. First, a tax on capital equipment corrects the “pecuniary externalities” caused by ESTP. The correction prescribes a downward or an upward adjustment of tax rates over time, depending on whether ESTP takes place at an accelerated or a decelerated pace. Second, both Ramsey and Mirrlees approaches prescribe an increasing marginal tax rate on labor income over time.

Should robots be taxed? This question has recently aroused the interest of economists; see, for example, [Guerreiro et al. \(2017\)](#), [Thuemmel \(2018\)](#), and [Acemoglu et al. \(2020\)](#). Our paper has not sought to answer this specific question. However, given that robots are part of capital equipment, one may view our paper as belonging to this line of the literature. As such, our derived results may be useful for the analysis of taxing robots as well. In particular, it is plausible to argue that the technological progress of robots would cause “pecuniary externalities” just like ESTP causes “pecuniary externalities” in our paper. If so, what are these “pecuniary externalities”? How should taxes be employed to correct them? How should taxes be adjusted over time in response to the technological progress of robots? Answering these and related questions seems interesting and important. We plan to pursue this research direction in the future.

REFERENCES

- Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: implications for employment and earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4, Chapter 12, pp. 1043–1171. Elsevier.
- Acemoglu, D., A. Manera, and P. Restrepo (2020). Does the US tax code favor automation? NBER Working Papers 27052, National Bureau of Economic Research, Inc.
- Ales, L., M. Kurnaz, and C. Sleet (2015). Technical change, wage inequality, and taxes. *American Economic Review* 105(10), 3061–3101.
- 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(C), 420–444.
- Angelopoulos, K., B. X. Fernandez, and J. R. Malley (2014). The distributional consequences of tax reforms under capital-skill complementarity. *Economica* 81(324), 747–767.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the other 99 percent. *Science* 344, 843–851.

Barro, R. J. (1979). On the determination of the public debt. *Journal of Political Economy* 87(5), 940–971.

Barro, R. J. and C. Sahasakul (1983). Measuring the average marginal tax rate from the individual income tax. *The Journal of Business* 56(4), 419–452.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2017). Public debt in economies with heterogeneous agents. *Journal of Monetary Economics* 91(C), 39–51.

CBO (2014). Taxing capital income: effective marginal tax rates under 2014 law and selected policy options. Reports 49817, Congressional Budget Office.

Chamley, C. (1986). Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica* 54(3), 607–622.

Chari, V. V., L. J. Christiano, and P. J. Kehoe (1994). Optimal fiscal policy in a business cycle model. *Journal of Political Economy* 102(4), 617–652.

Conesa, J. C., S. Kitao, and D. Krueger (2009). Taxing capital? Not a bad idea after all! *American Economic Review* 99(1), 25–48.

Conesa, J. C. and D. Krueger (1999). Social security reform with heterogeneous agents. *Review of Economic Dynamics* 2(4), 757–795.

Costinot, A. and J. Vogel (2010). Matching and inequality in the world economy. *Journal of Political Economy* 118(4), 747–786.

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.

Dávila, J., J. H. Hong, P. Krusell, and J.-V. Ríos-Rull (2012). Constrained efficiency in the neoclassical growth model with uninsurable idiosyncratic shocks. *Econometrica* 80(6), 2431–2467.

Diamond, P. A. and J. A. Mirrlees (1971). Optimal taxation and public production I: production efficiency. *American Economic Review* 61(1), 8–27.

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

Flug, K. and Z. Hercowitz (2000). Equipment investment and the relative demand for skilled labor: international evidence. *Review of Economic Dynamics* 3(3), 461–485.

Golosov, M., A. Tsyvinski, and I. Werning (2007). New dynamic public finance: a user's guide. In *NBER Macroeconomics Annual 2006, Volume 21*, pp. 317–388. National Bureau of Economic Research, Inc.

Gordon, R. J. (1990). *The Measurement of Durable Goods Prices*. NBER Books.

Gravelle, J. G. (2011). Reducing depreciation allowances to finance a lower corporate tax rate. *National Tax Journal* 64(4), 1039–1053.

Griliches, Z. (1969). Capital-skill complementarity. *Review of Economics and Statistics* 51(4), 465–468.

Guerreiro, J., S. Rebelo, and P. Teles (2017). Should robots be taxed? NBER Working Papers 23806, National Bureau of Economic Research, Inc.

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., K. Storesletten, and G. L. Violante (2017). Optimal tax progressivity: an analytical framework. *The Quarterly Journal of Economics* 132(4), 1693–1754.

Hornstein, A., P. Krusell, and G. L. Violante (2005). The effects of technical change on labor market inequalities. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1, Chapter 20, pp. 1275–1370. Elsevier.

Jones, L. E., R. E. Manuelli, and P. E. 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.

Kocherlakota, N. R. (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.

Laffont, J.-J. and J. Tirole (1988). The dynamics of incentive contracts. *Econometrica* 56(5), 1153–1175.

Ljungqvist, L. and T. J. Sargent (2018). *Recursive Macroeconomic Theory, fourth edition*. MIT Press Books.

Lucas, R. E. (1987). *Models of Business Cycles*. Wiley-Blackwell.

Lucas, R. E. and N. L. Stokey (1983). Optimal fiscal and monetary policy in an economy without capital. *Journal of Monetary Economics* 12(1), 55–93.

Maliar, L., S. Maliar, and I. Tsener (2020). Capital-skill complementarity and inequality: twenty years after. CEPR Discussion Papers 15228, C.E.P.R. Discussion Papers.

McDaniel, C. (2007). Average tax rates on consumption, investment, labor and capital in the OECD: 1950-2003. Working Paper.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38(2), 175–208.

Quadrini, V. and J. Ríos-Rull (2015). Inequality in macroeconomics. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 2, Chapter 14, pp. 1229–1302. Elsevier.

- Ramsey, F. (1927). A contribution to the theory of taxation. *Economic Journal* 37, 47–61.
- Slavík, C. and H. Yazici (2014). Machines, buildings, and optimal dynamic taxes. *Journal of Monetary Economics* 66(C), 47–61.
- Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics* 17(2), 213–240.
- Teulings, C. N. (1995). The wage distribution in a model of the assignment of skills to jobs. *Journal of Political Economy* 103(2), 280–315.
- Thuemmel, U. (2018). Optimal taxation of robots. Working Paper.
- Werning, I. (2007). Optimal fiscal policy with redistribution. *Quarterly Journal of Economics* 122(3), 925–967.
- Zhu, X. (1992). Optimal fiscal policy in a stochastic growth model. *Journal of Economic Theory* 58(2), 250–289.