Rising skill premium and dynamics of optimal taxation:

When capital-skill complementarity and investment-specific technological change are present simultaneously

Yi-Chan Tsai∗
National Taiwan University

C.C. Yang
Academia Sinica

Hsin Jung Yu
National Chengchi University
Feng Chia University

December 25, 2018

Abstract

With capital-skill complementarity in production, the secular cheapening of capital due to investment-specific technological change keeps pushing up the demand for skilled relative to unskilled labor and makes the skill premium on the rise. This paper quantitatively characterizes the dynamics of optimal taxation in response. Two main results emerge, no matter whether adopting the Ramsey (1927) or the Mirrlees (1971) approach. First, capital income should be taxed but, over time, it is optimal to cut capital tax rates in order to accommodate improved technology. Second, the planner substitutes labor tax for capital tax and implements an increasing progressivity of labor tax over time to remedy the worsening redistribution resulting from cutting capital tax rates. Both results are in sharp contrast to the work of Werning (2007), which prescribed that capital income should go untaxed and labor tax should be perfectly smoothed over time.

∗Corresponding author, Department of Economics, National Taiwan University. Phone: 886-2-33668314, Fax: 886-2-2365-9128, E-mail: yichantsai@ntu.edu.tw.
1 Introduction

“This ongoing process of machine substitution for routine human labor complements educated workers who excel in abstract tasks ... Simultaneously, it devalues the skills of workers, typically those without postsecondary education, who compete most directly with machinery in performing routine-intensive activities.” Autor (2014, p. 846)

Figure 1 plots the time series of the log skill premium (i.e., the mean of the natural logarithm of weekly wages for college graduates relative to high school graduates) in the U.S. labor market from 1963 to 2012. One feature clearly stands out from the figure: The skill premium has soared dramatically since the early 1980s. This significantly upward trend in the skill premium has aroused serious concerns because it gives rise to greater income inequality and increasing disparity of economic well-being between skilled workers (college graduates) and unskilled workers (high school graduates).

In this paper we address the question: How should taxation be set dynamically in response to the rising skill premium?

Figure 1. Log skill premium, 1963-2012

Note: Data are from Autor (2014).

1 The definition of the skill premium follows the convention. The data are from Autor (2014). We explain the data in more detail later.

2 Katz and Murphy (1992) is the seminal work on the issue. For a recent review, see Autor (2014).
To address the question, the first priority is to know the fundamental cause of the rise in the skill premium. A leading explanation in the literature is so-called “skill-biased technical change,” which is commonly referred to as a latent trend that favors skilled over unskilled workers by raising skilled labor’s relative productivity and hence its relative demand. In an influential work, Krusell et al. (2000) argued that the tangible force driving the latent trend lies in the simultaneous presence of “capital-skill complementarity” and secular “investment-specific technological change.”

Consider a production technology under which capital is more substitutable for, or less complementary to, unskilled labor than skilled labor. This property is known as “capital-skill complementarity.” A critical implication of capital-skill complementarity is that a higher stock of capital 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 due to investment-specific technological change; see Figure 2. With a plausible difference in the elasticities of substitution between capital and skilled vs. unskilled labor that supports capital-skill complementarity, Krusell et al. (2000) found that the rising skill premium since the early 1980s can be well explained through the secular cheapening of capital associated with investment-specific technological change, even though there was a substantial increase in the relative supply of college skills during the sample period they studied.

In light of the important finding of Krusell et al. (2000), a sensible and interesting policy question naturally arises: How should a government set tax policy dynamically in response to the simultaneous presence of capital-skill complementarity and secular investment-specific technological change? Employing an empirically plausible capital-skill complementarity form of production function suggested by Krusell et al. (2000), this paper quantitatively characterizes the dynamics of optimal taxation in response.

There is a large literature investigating how a government should set taxes on capital versus labor in the face of aggregate shocks. Of them, the contribution of Werning (2007)
is most closely related to our paper. Unlike the typical representative-household framework that abstracts from income inequality in the literature, he explicitly modeled distributional concerns by envisioning a heterogeneous-household economy in which household types in terms of their labor productivity are permanently fixed. Werning (2007) emphasized in this heterogeneous-household framework that the need for imposing distortionary taxes naturally arises from the trade-off between redistribution and efficiency.

Household types in our economy are either skilled or unskilled, and they are permanently fixed as well.\(^8\) As such, it appears that our model is just a special case considered by Werning (2007). However, there is a crucial difference: While all types of labor are equally complementary to capital in the standard neoclassical production in Werning (2007), skilled labor is more complementary to capital than unskilled labor as in Krusell et al. (2000) 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 skill premium on the rise. Indeed, in the face of a specific aggregate shock – investment-specific technological change – in our context, 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 the instantaneous household preferences are separable and isoelastic,

\(^8\)In the extension we allow for exogenous variation in the proportion of household types over time.
Werning (2007, Proposition 2 for Ramsey taxation and Proposition 6 for Mirrlees taxation) prescribed in the face of aggregate shocks: (i) a zero capital tax rate with no intertemporal distortion all the time, except for the initial period; (ii) perfect labor tax smoothing (a constant labor tax rate 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)).

We adopt both the Ramsey and the Mirrlees approach to optimal taxation as Werning (2007) did. In the simultaneous presence of capital-skill complementarity and the secular cheapening of capital due to investment-specific technological change, two main results emerge, no matter whether adopting the Ramsey or the Mirrlees approach. First, capital income should be taxed but, over time, it is optimal to cut capital tax rates in order to accommodate improved technology. Second, the planner substitutes labor tax for capital tax and implements an increasing progressivity of labor tax over time to remedy the worsening redistribution resulting from cutting capital tax rates. Both results are in sharp contrast to the work of Werning (2007), which prescribed that capital should go untaxed and labor tax should be perfectly smoothed over time. The reasons why capital income should be taxed at a point in time are different between Ramsey and Mirrlees. Nevertheless, in the face of secular cheapening of capital, both adopt the policy of cutting capital tax rates in order to reap the gains of improved production technology. We elaborate on these findings after presenting them.

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 Quadrini and Ros-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) featured significantly. Fisher (2006) found an acceleration in the decline in the relative price of capital since the early 1980s and explored its implications for business cycles. Karabarbounis and Neiman (2014) documented a significant decline in the labor share globally since the early 1980s, occurring within the large majority of countries and industries. They inferred that the decline in the relative

9A difference between the Ramsey and the Mirrlees is that while the former requires all households to face the same marginal labor tax rate at all times, the latter allows the same marginal labor tax rate to vary across different household types.

10See also Zhu (1992) and Chari et al. (1994).
price of capital accounts for roughly half of the decline in the global labor share. Costinot and Werning (2018) studied optimal technology regulation in a static environment. Among others, 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 finding to their result later in the paper.

Although there is a large 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.\footnote{He and Liu (2008) and Angelopoulos et al. (2014) addressed the quantitative effects of some hypothetical tax-policy changes with capital-skill complementarity. Guerreiro et al. (2017) and Thuemmel (2018) specifically focused on robots and addressed the question of whether robots should be taxed in static settings.}

Jones et al. (1997) utilized capital-skill complementarity as an example to highlight a point: If the tax rates on skilled and unskilled labor are required to be equal, the startling finding by Chamley (1986) and Judd (1985) that capital should go untaxed in the steady state may no longer hold when production takes the form of capital-skill complementarity. While Jones et al. (1997) focused on the steady-state property of optimal taxation with capital-skill complementarity, we focus on the transitional dynamics of optimal taxation in the face of secular investment-specific technological change.

Assuming that labor 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. One of their main findings is that with capital-skill complementarity the cyclical properties of optimal labor taxes significantly depend on whether the relative supply of skill is restricted or flexible. In contrast to the setting of Angelopoulos et al. (2015) that the relative price of capital follows the stationary AR(1) process, we solve for the transitional dynamics of optimal taxation in the face of the secular fall in the relative price of capital because of investment-specific technological change. Both Jones et al. (1997) and Angelopoulos et al. (2015) addressed their problems in the representative-household framework and thereby both abstract from the redistributive role of taxation. In contrast, we are 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). Households are either skilled or unskilled and their focus was on the differential taxation of capital income based on capital type. They found that capital equipment should be taxed at a higher rate than capital structures, violating the prescription of the classical work of Diamond and Mirrlees (1971) that tax systems should maintain production efficiency.
The intuition behind this result, as they explained, is that households ignore the adverse effect of increasing capital equipment relative to capital structures on the skill premium and, therefore, there is a tendency for households to save too much through equipment relative to structures. Taxing capital equipment at a higher rate is to remedy this tendency. We make no distinction between structures and equipment in capital as in the standard literature but introduce investment-specific technological change, 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 Section 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 similar to Werning (2007). The main departures are: (i) 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) aggregate shocks are specific in terms of variation in the relative price of capital due to investment-specific technological change. To simplify exposition and abstract from complication, it is assumed that people are endowed with perfect foresight regarding investment-specific technological change.

Households are either skilled or unskilled; however, the planner (a benevolent government) cannot impose taxes directly conditional on people’s skills (say, whether heads of households are college or high school graduates). As pointed out by Werning (2007) and Kocherlakota (2010, p.3), this ends up implying that the planner must treat households as if they were privately informed about their own skills. The setting prevents the planner from imposing discriminatory lump-sum taxes on household types to achieve the first-best.

We first consider the Ramsey approach to taxation and turn to the Mirrlees approach in Sections 5-6. A key difference between the two approaches is that while tax schemes are exogenously specified a priori in the Ramsey, the only restriction imposed on tax schemes by the Mirrlees is that taxes cannot condition directly on people’s skills (household types).\textsuperscript{12}

Time is discrete and the horizon is infinite, indexed by $t = 0, 1, 2, \ldots$. The economy consists of heterogeneous households, a representative firm, and the government. We describe them in turn.

\textsuperscript{12}See Golosov et al. (2007) and Kocherlakota (2010, p.3) for more discussions on the differences between the Mirrlees and the Ramsey approach.
2.1 Households

Households are divided into two types: the skilled \((s)\) and the unskilled \((u)\). We assume that both types have the size of unit measure, and that the labor productivity of the skilled and that of the unskilled are both 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. We relax these assumptions in the extension.

All households share the same preferences and their lifetime utility equals

\[
V(i) = \sum_{t=0}^{\infty} \beta^t U(c_{it}, 1 - n_{it}),
\]

where \(\beta \in (0, 1)\) is the discount factor, and \(c_{it}\) consumption and \(n_{it}\) “raw” labor supply (hours worked) with \(i \in \{s, u\}\). The period utility function \(U(.)\) is weakly concave and continuously twice differentiable with the properties of \(U_c \geq 0, U_{cc} \geq 0, U_n \leq 0, U_{nn} \leq 0,\) and \(U_{cn} = 0\).

Households own capital and rent it to the representative firm for the use. Capital owned by type \(i\) household is denoted \(k_{it}\) and its law of motion is governed by

\[
k_{i,t+1} = (1 - \delta)k_{it} + q_t I_{it}, \text{ given } k_{i0},
\]

where \(\delta\) is the depreciation rate and \(I_{it}\) investment at time \(t\). Following the literature, we interpret the variable \(q_t\) in (2) as the investment-specific technological change that enhances the productivity of newly formed capital relative to prior vintages of capital. Its inverse, \(1/q_t\), then represents the relative price of new capital; see Hornstein et al. (2005). From Figure 2, we see that \(1/q_t\) has been falling constantly since 1960s.

Households have the following flow budget constraint:

\[
c_{it} + I_{it} + b_{i,t+1} = w_{it} n_{it} + r_t k_{it} + (1 + r_{it}) b_{it} - \mathcal{T}_t, i \in \{s, u\}, \forall t,
\]

where \(b_{it}\) is the risk-free, one-period noncontingent government bond held by type \(i\) households at time \(t\) and \(r_{it}\) is its rate of return \((b_{i0}\) is given); \(r_t\) denotes the pre-tax rental rate of capital at time \(t\), \(w_{it}\) the pre-tax wage rate received by type \(i\) households at time \(t\), and \(\mathcal{T}_t(.)\) the tax function imposed by the government at time \(t\).

In the case of the Ramsey approach, we specify \(\mathcal{T}_t(.)\) a priori with

\[
\mathcal{T}_t(.) = \tau_{K,t} \left( r_t - \frac{\delta}{q_t} \right) k_{it} + \tau_{L,t} w_{it} n_{it} - T_t,
\]

where \(\tau_{K,t}\) is the flat tax rate imposed on capital income and \((\tau_{L,t}, T_t)\) is the linear tax
schedule imposed on labor income. We address more on this tax scheme when we come to the description of the government. Using (2) and the specification of $\mathcal{I}(t)$, one can rewrite (3) as

$$c_{it} + \frac{k_{i,t+1}}{q_{it}} + b_{i,t+1} = (1 - \tau_{L,t})w_{it}n_{it} + \left[\frac{1}{q_{it}} + (1 - \tau_{K,t}) \left( r_{t} - \frac{\delta}{q_{t}} \right) \right] k_{it}$$

$$+ (1 + r_{bt}) b_{it} + T_{r},$$

which would reduce to the familiar household budget constraint if $q_{t} = 1$. Given $b_{i0}$ and $k_{i0}$, the household problem is to choose $\{c_{it}, n_{it}, k_{i,t+1}, b_{i,t+1}\}$ to maximize the lifetime utility (1), subject to the laws of motion (2) and a sequence of budget constraints (4). The resulting FOCs (first-order conditions) are given by

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

$$U_{c,t}(i) = \beta U_{c,t+1}(i) \left[ \frac{q_{t}}{q_{t+1}} + q_{t} (1 - \tau_{K,t+1}) \left( r_{t+1} - \frac{\delta}{q_{t+1}} \right) \right],$$

$$U_{c,t}(i) = \beta U_{c,t+1}(i)[1 + r_{b,t+1}],$$

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 type $i$ household’s optimal behavior in the face of factor prices $\{r_{t}, w_{st}, w_{ut}\}$ and tax policy $\{\tau_{L,t}, \tau_{K,t}, T_{t}\}$. From the perspective of households, capital and government bond are perfect substitutes. Thus, no arbitrage condition implies that the post-tax returnsof capital given in (5b) and the return of government bond given in (5c) must be equal to each other in equilibrium. That is,

$$\frac{q_{t}}{q_{t+1}} + q_{t} (1 - \tau_{K,t+1}) \left( r_{t+1} - \frac{\delta}{q_{t+1}} \right) = 1 + r_{b,t+1}.$$

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:

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

8
where $A_{i0} = \left[ \frac{1}{q_0} + (1 - \tau_{K,0}) \left( r_0 - \frac{\delta}{q_0} \right) \right] k_{i0} + b_{i0}$ denotes the initial wealth held by 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 the production function of the form at time $t$:

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

(7)

where $Y_t$ denotes output, $K_t$ capital, $N_{st}$ skilled labor input, and $N_{ut}$ unskilled labor input, with $\sigma, \rho < 1$. All of them are aggregate variables with $N_{st} = n_{st}$, $N_{ut} = n_{ut}$, and $K_t = \sum_{i \in \{s, u\}} k_{it}$ in the benchmark model. This production function is qualitatively 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 and the two types of labor. In particular, the elasticity of substitution between capital and unskilled labor equals $1/(1 - \sigma)$, while the elasticity of substitution between capital and skilled labor equals $1/(1 - \rho)$. The so-called "capital-skill complementarity" in production arises if $\sigma > \rho$.

All markets are competitive and we let the final good be the numeraire. Subject to the production technology (7), the representative firm maximizes its profit

$$\Pi_t = Y_t - w_{st} N_{st} - w_{ut} N_{ut} - r_t K_t,$$

(8)

where $w_{st}$, $w_{ut}$, and $r_t$ are the factor prices for $N_{st}$, $N_{ut}$, and $K_t$, respectively.

The FOCs for the representative firm are given by $\frac{\partial Y}{\partial K} = r$, $\frac{\partial Y}{\partial N_{st}} = w_s$, and $\frac{\partial Y}{\partial N_{ut}} = w_u$.\(^{13}\)

The skill premium in the benchmark model is defined to be

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

(9)

Comparative statics give

$$\frac{\partial \xi}{\partial N_u} = (1 - \sigma) \frac{(1 - \mu)(1 - \lambda)}{\mu} B^{\frac{\sigma - \rho}{\rho}} N_s^{\rho - 1} N_u^{-\sigma} > 0 \text{ if } \sigma < 1,$$

\(^{13}\)An implicit assumption here is that although the government cannot distinguish between the two types of households directly, the firm can. This assumption is the same as that in Stiglitz (1982), who considered it to be quite a plausible assumption.
\[
\frac{\partial \xi}{\partial N_s} = \frac{(1 - \mu)(1 - \lambda)}{\mu N_u^{\sigma - 1}} B_{\sigma - 2}^{\rho - 2} N_s^{\rho - 2} [\lambda (\rho - 1) K^\rho + (1 - \lambda) (\sigma - 1) N_s^\rho] < 0 \text{ if } \sigma, \rho < 1,
\]

\[
\frac{\partial \xi}{\partial K} = (\sigma - \rho) \lambda \frac{(1 - \mu)(1 - \lambda)}{\mu} B_{\rho - 2}^{\sigma - 2} \left( \frac{N_s^{\sigma - 1}}{N_u^{\sigma - 1}} \right) K^{\rho - 1} > 0 \text{ if } \sigma > \rho,
\]

where \( B = \lambda K^\rho + (1 - \lambda) N_s^\rho \). These results are intuitive. In particular, \( \frac{\partial \xi}{\partial K} > 0 \) because of the feature of capital-skill complementarity with \( \sigma > \rho \). Note that \( \frac{\partial \xi}{\partial K} = 0 \) would hold if \( \sigma = \rho \).

Using the formula \( \ln(1 + y) \approx y \), one can express the skill premium defined in (9) as

\[
\ln \xi \approx \text{constant} + \lambda \left( \frac{\sigma - \rho}{\rho} \right) \left( \frac{K}{N_s} \right)^\rho + (1 - \sigma) \ln \left( \frac{N_u}{N_s} \right).
\] (10)

Krusell et al. (2000) called the first component of (10) the “capital-skill complementarity effect,” and the second component the “relative quantity effect.” The first effect indicates that given \( \sigma > \rho \), a faster growth in capital relative to skilled labor input will enhance the skill premium, as it increases the relative demand for skilled labor. The second effect indicates that, given \( \sigma < 1 \), a faster growth in the skilled relative to the unskilled labor input will reduce the skill premium as it increases the relative supply of skilled labor.

In the face of increasing \( q_t \) (investment-specific technology change), we will examine how the two allocations in (10) – \( K/N_s \) and \( N_u/N_s \) – evolve over time under optimal taxation.

### 2.3 Government

Following Werning (2007), the government imposes the following taxes at time \( t \): a flat tax rate \( \tau_{K,t} \) on capital income, and a linear tax schedule on labor income, which is composed of the lump-sum transfer \( T_t \) and the marginal tax rate \( \tau_{L,t} \). With skill inequality across households and the availability of \( T_t \), Werning (2007) noted that distributional concerns play a key role in the determination of the optimal tax rate \( \tau_{L,t} \), since (p. 927) “a positive tax rate ensures that more productive, richer workers bear a heavier tax burden and alleviate that of less productive, poorer workers.” Put differently, one can legitimately interpret a higher \( \tau_{L,t} \) of the linear tax schedule as representing a more progressivity of labor taxation.

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 amount of government bond \( B_0 = b_s0 + b_u0 \).
The flow government budget constraint is given by

$$G_t + 2T_t + (1 + r_t) B_t = \tau_{L,t}(w_{st}N_{st} + w_{ut}N_{ut}) + \tau_{K,t} \left( r_t - \frac{\delta}{q_t} \right) K_t + B_{t+1}, \forall t, \quad (11)$$

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 (11). 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 the government debt must be fully repaid by future primal surpluses (taxes collected net of government purchases and transfers).

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:

$$Y_t = C_t + I_t + G_t,$$  \quad (12)

where $C_t = \sum_{i \in \{s,u\}} c_{it}$ and $I_t = \sum_{i \in \{s,u\}} I_{it}$.

### 2.4 Competitive equilibrium

The following definition of competitive equilibrium is standard.

**Definition 1** Given the initial government bond $B_0$, the initial capital $K_0$, their holdings $\{b_{i0}\}_{i \in \{s,u\}}$ and $\{k_{i0}\}_{i \in \{s,u\}}$, and a sequence of investment-specific technological changes $\{q_t\}$ and government expenditures $\{G_t\}$, a competitive equilibrium is a sequence of fiscal policy $\{\tau_{L,t}, \tau_{K,t}, T_t, B_t\}$, market prices $\{r_t, w_{ut}, w_{st}\}$ and non-negative quantities $\{c_{it}, n_{it}, b_{i,t+1}, k_{i,t+1}\}_{i \in \{s,u\}}$ such that:

1. Given $\{q_t\}, \{G_t\}, \{\tau_{L,t}, \tau_{K,t}, T_t\}$ and $\{r_t, w_{ut}, w_{st}\}$, both skilled and unskilled households maximize their lifetime utility subject to the law of motion (2) and budget constraints (4).

2. Given $\{q_t\}, \{G_t\}, \{\tau_{L,t}, \tau_{K,t}, T_t\}$ and $\{r_t, w_{ut}, w_{st}\}$, the representative firm maximizes its profit (8) subject to the production technology (7).

3. The government’s budget constraint given by (11) is satisfied $\forall t$. 

4. All markets clear $\forall t$:

\[ K_t = \sum_{i \in \{s,u\}} k_{it}, \quad C_t = \sum_{i \in \{s,u\}} c_{it}, \quad B_t = \sum_{i \in \{s,u\}} b_{it}, \]

\[ N_{st} = n_{st}, \quad N_{ut} = n_{ut}, \]

\[ Y_t = C_t + I_t + G_t. \]

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 $B_0$, $K_0$, \( \{b_{i0}\}_{i \in \{s,u\}} \) and \( \{k_{i0}\}_{i \in \{s,u\}} \), the Ramsey problem is to choose a competitive equilibrium to maximize the following social welfare function:

\[ SWF = \sum_{i \in \{s,u\}} \psi^i V(i), \quad (13) \]

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

We adopt the primal approach to the Ramsey problem as in Werning (2007), who extended the primal approach of Lucas and Stokey (1983) from the representative-agent to the heterogeneous-agent framework. As explained in Ljungqvist and Sargent (2012, chapter 16), a key step in the primal Ramsey approach is to derive the so-call implementability condition. This step consists of two parts. First, use the FOCs of households and firms to express taxes and prices as functions of the allocation. Second, substitute these expressions for taxes and prices in the household’s present-value budget constraint. Since a common marginal labor tax rate is imposed on labor income of both the skilled and the unskilled at any point in time in our model, we also need to account for this extra restriction in the use of the primal approach. The details are given below.

Following Werning (2007), consider a fictitious representative-agent with utility $U_f^i (C_t, n_{st}, n_{ut}; \varphi)$ which solves the static subproblem:

\[ U_f^i (C_t, n_{st}, n_{ut}; \varphi) = \max_{c_{st}, c_{ut}} \sum_{i \in \{s,u\}} \varphi^i U(c_{it}, 1 - n_{it}), \text{ subject to } \sum_{i \in \{s,u\}} c_{it} = C_t, \]
where $\varphi = \{\varphi^i\}_{i \in \{s,u\}}$ are “market” weights. From the above static subproblem, we obtain

$$U^f_{c_t} = \varphi^i U_{c_t}(i), \quad U^f_{n_t}(i) = \varphi^i U_{n_t}(i), \quad i = \{s,u\},$$

(14)

where $U^f_{c_t} \equiv \partial U^f / \partial C_t$ and $U^f_{n_t}(i) \equiv \partial U^f / \partial n_{it}$. Werning (2007) observed that inefficiencies due to distortive linear taxation are all confined to the determination of aggregates allocations, which are represented by $\{C_t, n_{st}, n_{ut}\}$ in our benchmark model. Using (14) and (5a)-(5b), we then have

$$-\frac{U^f_{n_t}(i)}{U^f_{c_t}} = (1 - \tau_{L,t})w_{it}, \quad i = \{s,u\},$$

(15)

$$\frac{U^f_{c_t}}{\beta U^f_{c_{t+1}}} = \left[ \frac{q_t}{q_t} + q_t (1 - \tau_{K,t+1}) \left( r_{t+1} - \frac{\delta}{q_t} \right) \right],$$

(16)

and household consumption can be expressed as

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

where $c^*_it$ is independent of $(n_{st}, n_{ut})$ because $U_{cn} = 0$ by assumption.

Utilizing the results from the fictitious representative-agent problem, we derive the implementability conditions from (6) as in Werning (2007):

$$\sum_{t=0}^{\infty} \beta^t \left[ U^f_{c_t} c^*_it + U^f_{n_t}(i)n_{it} \right] = U^f_{c_0}(A_{i_0} + T), \quad i \in \{s,u\}.$$  

(17)

The skilled and the unskilled are perfectly substitutable in terms of their effective labor supply in Werning (2007). This is not true in our setting. Thus, unlike Werning (2007), we also need to impose the following restriction in the formulation of the Ramsey problem:

$$\frac{U_{n_t}(s)}{U_{c_t}(s)w_{st}} = \frac{U_{n_t}(u)}{U_{c_t}(u)w_{ut}}, \forall t,$$

which guarantees that both the skilled and the unskilled household face the same marginal labor tax rate at any point in time. Using (14), the above restriction can be expressed as

$$\frac{U^f_{n_t}(s)}{U^f_{c_t}w_{st}} = \frac{U^f_{n_t}(u)}{U^f_{c_t}w_{ut}}, \forall t,$$
which leads to
\[
\log \frac{U_{n,t}^f(s)}{U_{n,t}^f(u)} - \log \xi_t = 0, \forall t,
\] (18)
where \( \xi = w_s/w_u \) is the skill premium as given by (9).

Let \( \theta = \{\theta^s, \theta^u\}, \{\beta^t_1\}, \) and \( \{\beta^t_Y_t\} \) denote, respectively, the Lagrange multipliers on the implementability conditions (17), the resource constraints (12), and the extra restriction (18). Given initial capital \( K_0 \), initial bond \( B_0 \), the distributions of their holdings between the skilled and the unskilled, and initial capital tax rate \( \tau_{K,0} \), forming the Lagrangian for the Ramsey problem gives

\[
L = \max_{\{C_t, n_{st}, n_{ut}, K_{t+1}, \varphi, T\}} \sum_{t=0}^{\infty} \beta^t W(C_t, n_{st}, n_{ut}, \varphi, \theta) \\
+ \sum_{t=0}^{\infty} \beta^t \Gamma_t \left[ F(K_t, N_{st}, N_{ut}) + \frac{(1-\delta)}{q_t} K_t - C_t - \frac{K_{t+1}}{q_t} - G_t \right] \\
+ \sum_{t=0}^{\infty} \beta^t \Upsilon_t \left[ \log \frac{U_{n,t}^f(s)}{U_{n,t}^f(u)} - \log \xi_t \right] \\
- U_{C_0}^f \sum_{i \in \{s, u\}} \theta^i (A_{i0} + T),
\]

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

\[
W(C_t, n_{st}, n_{ut}, \varphi, \theta) = \sum_{i \in \{s, u\}} \left\{ \psi^i U(c^s_{it}, n_{it}) + \theta^i \left[ U_{C_t}^f c^s_{it} + U_{n,t}^f(i)n_{it} \right] \right\}.
\]

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

\footnote{The initial capital tax rate, \( \tau_{K,0} \), 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 \( \tau_{K,0} \) to confiscate all capitals of households (and will also confiscate all of bond holdings by 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 of choosing \( \tau_{K,0} \) in the Ramsey problem. Specifically, we let \( \tau_{K,0} \) correspond to the U.S. capital tax rate in the initial steady state in our quantitative study later.}
Ramsey problem for \( t \geq 1 \) are \(^{15}\)

\[
W_{C_t} = \Gamma_t, \tag{19}
\]

\[-W_{n,t}(s) = \Gamma_t w_{st} + \frac{Y_t}{n_{st}} \left( \frac{\partial U_{n,t}^f(s)/\partial n_{st}}{U_{n,t}^f(s)} - \frac{n_{st}}{\xi_t} \frac{\partial \xi_t}{\partial n_{st}} \right), \tag{20}\]

\[-W_{n,t}(u) = \Gamma_t w_{ut} - \frac{Y_t}{n_{ut}} \left( \frac{\partial U_{n,t}^f(u)/\partial n_{ut}}{U_{n,t}^f(u)} + \frac{n_{ut}}{\xi_t} \frac{\partial \xi_t}{\partial n_{ut}} \right), \tag{21}\]

\[
\frac{\Gamma_t}{q_t} = \beta \left[ \Gamma_{t+1} \left( r_{t+1} + \frac{1 - \delta}{q_{t+1}} \right) - \Upsilon_{t+1} \left( \frac{1}{\xi_{t+1}} \frac{\partial \xi_{t+1}}{\partial K_{t+1}} \right) \right], \tag{22}\]

\[
\sum_{i \in \{s,u\}} \theta^i = 0, \tag{23}\]

where the last FOC is with respect to \( T \) and its form is qualitatively the same as equation (17) derived by Werning (2007). Together with the implementability conditions (17), the resource constraints (12), and the restriction (18), the FOCs of the Ramsey problem characterize the allocation of competitive equilibrium under the optimal fiscal policy \( \{\tau_{L,t}, \tau_{K,t}, T, B_t\} \).

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 indetermination, we focus on the characterization of optimal tax rates \( \{\tau_{L,t}, \tau_{K,t}\} \).

From (15)-(16) and (19)-(22), \( \tau_{K,t+1} \) and \( \tau_{L,t} \) at the optimum satisfy:\(^{16}\)

\[
\tau_{K,t+1} = \frac{W_{C_t}}{W_{C_{t+1}} \left( 1 - \Delta_{K,t+1} \right)} - \frac{U_{C_t}^f}{U_{C_{t+1}}^f} \beta q_{t+1} \left( r_{t+1} - \frac{\delta}{q_{t+1}} \right), \tag{24}\]

\[
\tau_{L,t} = 1 + \frac{U_{n,t}^f(s)}{U_{C_t}^f w_{st}} = 1 - \frac{U_{n,t}^f(s) W_{C_t} \left( 1 + \Delta_{st} \right)}{U_{C_t}^f W_{n,t}(s)} \tag{25}\]

\[
= 1 + \frac{U_{n,t}^f(u)}{U_{C_t}^f w_{ut}} = 1 - \frac{U_{n,t}^f(u) W_{C_t} \left( 1 - \Delta_{ut} \right)}{U_{C_t}^f W_{n,t}(u)}, \tag{26}\]

\(^{15}\)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 \( \phi \) is not reported here, but it is used in the numerical analysis.

\(^{16}\)From (16), we have \( \frac{U_{C_t}^f}{\beta q_t U_{C_{t+1}}^f} = r_{t+1} + \frac{1 - \delta}{q_{t+1}} - \tau_{K,t+1} \left( r_{t+1} - \frac{\delta}{q_{t+1}} \right) \).

From (19) and (22), we have \( \frac{W_{C_t}}{\beta q_t W_{C,t+1} \left( 1 - \Delta_{K,t+1} \right)} = r_{t+1} + \frac{1 - \delta}{q_{t+1}} \).

Putting them together leads to (24).
where

\[ \Delta K_{t+1} = \frac{\Upsilon_{t+1}}{W_{Ct+1}} \left( \frac{1}{t_{t+1}} \right) \left( K_{t+1} + \frac{1}{t_{t+1}} \frac{\partial K_{t+1}}{\partial K_{t+1}} \right), \]

\[ \Delta s_t = \frac{\Upsilon_t}{W_C w_s n_s} \left( \frac{\partial U_{n,t}(s)/\partial n_{st}}{U_{n,t}(s)} - \frac{n_{st} \partial n_{st}}{n_{st}} \right), \]

\[ \Delta u_t = \frac{\Upsilon_t}{W_C w_u n_u} \left( \frac{\partial U_{n,t}(u)/\partial n_{ut}}{U_{n,t}(u)} + \frac{n_{ut} \partial n_{ut}}{n_{ut}} \right). \]

If the skilled and the unskilled were perfectly substitutable in terms of their effective labor supply as in Werning (2007), it would be unnecessary to impose the extra restriction (18) in the Ramsey problem. Given \( \Upsilon_t \equiv 0 \), we would have \( \Delta K_{t+1} = \Delta s_t = \Delta u_t = 0 \) at all times. Our derived tax-rate formulas, (24) and (25), would then reduce to those derived in Werning (2007). In particular, when \( U(.) \) takes the separable isoelastic form as expressed by equation (27) later, Werning (2007) showed that the resulting optimal \( \tau_{K,t+1} = 0 \) for all \( t \geq 0 \).

Even if skilled and unskilled labors are not perfect substitutes, the result of \( \Upsilon_t \equiv 0 \) would still be applicable if the planner were able to impose tax rates on labor income of the skilled and that of the unskilled separately.

Imposing a common tax rate on labor income of both the skilled and the unskilled could be viewed as a third-best if compared to the second-best of imposing separate tax rates (the first-best is to impose discriminatory lump-sum taxes directly on household types). Unfortunately, the second-best is not incentive compatible under the linear labor tax schedule, since all households will obviously choose to face lower rather than higher marginal labor tax rates. Thus, an interpretation of the third-best with \( \tau_{K,t} > 0 \) at the optimum (as we shall show quantitatively later) is that it serves as a way to imperfectly mimic the second-best situation where the planner could somehow be able to impose separate tax rates on labor income of the skilled and that of the unskilled separately.

The logic underlying the result of \( \tau_{K,t} > 0 \) at the optimum is closely related to the findings in Corlett and Hague (1953), Erosa and Gervais (2002), and Conesa et al. (2009). Corlett and Hague (1953) showed that since leisure cannot be taxed directly, it is optimal to impose higher tax rates on commodities that are complementary to leisure. Within a life-cycle framework, Erosa and Gervais (2002) found that a zero rather than nonzero capital tax would result at the optimum if labor taxes conditional on age were available. In offering an intuition for their quantitative result of finding a positive optimal capital tax rate, Conesa et al. (2009, p. 41) explained: “in a life-cycle model in which household labor supply changes with age, if the government cannot condition the tax function on age, it optimally uses the
capital income tax to mimic age-dependent labor income taxes.”

4 Dynamics of Ramsey taxation

This section quantitatively characterizes the dynamics of Ramsey taxation over time in the face of investment-specific technological changes \( \{q_t\} \). The optimal solution we obtain maximizes the 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 the investment-specific technological change \( q_t \) 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 a non-linear solution method.

4.1 Calibration

To carry out quantitative explorations, 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.

First, considering 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 in Figure 1. We normalize \( q_{t=1963} \equiv q_0 = 1 \).

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. McDaniel’s calculation focuses on the part of taxes and leaves out the part of transfers. As such, her 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. Browning and Johnson (1984) argued that only the net effect of taxes and transfers is crucial for redistribution, and they provided evidence in support of the hypothesis that a linear income tax can have distributional implications similar to those resulting from the actual tax plus transfer system. Adapting from a figure in Jonathan Heathcote (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. Figure 3 shows the evolution of McDaniel’s calculated tax rates on labor and capital income in the U.S. economy for the period 1963-2013.\(^{17}\) Since the

\(^{17}\)The tax series have been updated to 2013 by McDaniel.
year 1963 serves as our initial steady state, we simply let McDaniel (2007) calculated tax rates in 1963 be the initial U.S. tax rates.

The ratio of government expenditures to GDP is set to 20.5%, which is the average between year 1963 and 2017 from NIPA. We stick to this ratio throughout the period we study.

![Figure 3. Tax rates in the U.S. calculated by McDaniel (2007)](image)

The period utility function $U(.)$ in (1) is specified to be

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

(27)

where $\gamma_c$ a relative risk aversion parameter, $\chi > 0$ a weight between consumption and leisure, and $\gamma_n$ related to the Frisch elasticity of labor supply, which is equal to $\frac{v'}{n_v v''} = \frac{1-n}{\gamma_n n_l}$.

This separable isoelastic specification of household preferences is commonly employed in the literature and it also facilitates comparison with the results in Werning (2007).

For the parameters associated with preferences, we set $\gamma_c$ and $\gamma_n$ to be 1.5 and 2, and the discount factor $\beta$ to be 0.98. These are standard in the literature. On the production side, based on Greenwood et al. (1997), we set the depreciation rate for capital to be a weighted average between structures and equipment, which equals $\delta = 0.096$. The parameter $\sigma$ is chosen to be 0.401, and the parameter $\rho$ is equal to $-0.495$. As a result, 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 $\frac{1}{1-\sigma} \approx 1.67$. Both are consistent with the findings in
Krusell et al. (2000).\textsuperscript{18}

There are three parameters that remain to be calibrated, which are $\mu$, $\lambda$ and $\chi$. Parameters $\mu$ and $\lambda$ are related to the production function (7), and $\chi$ is the relative weight between consumption and leisure in the utility function (27). We calibrate the values of these parameters to match the following moment conditions of the U.S. economy in 1963 (the initial steady state):\textsuperscript{19}

1. The skill premium $\xi$, when defined as the average annual wage of college graduates relative to that of high-school graduates, equals 1.474 (Autor, 2014).

2. The capital output ratio, $\frac{K}{Y}$, is about 2.27 (NIPA).

3. The ratio of consumption (excluding durable goods) to GDP in 1963 is equal to 0.6 (NIPA).

4. 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).

5. The ratio of gross domestic investment to GDP in 1963 is equal to 0.247 (NIPA).

Finally, we need to specify the distribution of capital between the skilled and the unskilled workers in the initial steady state. We match it to the ratio of the unskilled’s wealth relative to the skilled’s. Since the data on this ratio are only available from 1989 on, we simply use the 1989 datum, which is equal to 0.57.\textsuperscript{20} Although admittedly unsatisfactory, it is the earliest datum we can find with regard to it.

Table 1 summarizes all of our parameter values that are directly set, while Table 2 reports the results of our moment matching. The resulting steady-state capital in competitive equilibrium will serve as the initial capital for our dynamic economy.

We apply the obtained parameter values above to both the Ramsey and the Mirrlees approach in our quantitative study.

### 4.2 Time-series data

In the extended model (Section 7) there are five sets of time-series data we need to use in our quantitative study: The tax rates $\{\tau_t\}$, the skill premium $\{\xi_t\}$, the investment-specific technological change $\{q_t\}$, and $\{x_t\}$ and $\{z_t\}$ with $x_t = z_t \cdot \pi_t$, where $z_t$ represents the

\textsuperscript{18} We check the robustness of our results by considering other values of $\sigma$ and $\rho$ in some extreme examples later and in the Online Appendix.

\textsuperscript{19} We target the first three moment conditions. The other two moment conditions serve as a robust check.

\textsuperscript{20} The data are from the U.S. Census Bureau, Asset ownership of households.
Table 1: Parameter values set

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor</td>
<td>$\beta$</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Elasticity of intertemporal substitution</td>
<td>$\gamma_c$</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Elasticity of leisure</td>
<td>$\gamma_n$</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Elasticity of $N_u$ and $N_s$, $K$ composite</td>
<td>$\sigma$</td>
<td>0.401</td>
<td>Krusell et al. (2000)</td>
</tr>
<tr>
<td>Elasticity of $N_s$ and $K$</td>
<td>$\rho$</td>
<td>-0.495</td>
<td>Krusell et al. (2000)</td>
</tr>
<tr>
<td>Depreciation rate of $K$</td>
<td>$\delta$</td>
<td>0.096</td>
<td>Greenwood et al. (1997)</td>
</tr>
<tr>
<td>Capital tax rate</td>
<td>$\tau_K$</td>
<td></td>
<td>McDaniel (2007)</td>
</tr>
<tr>
<td>Labor income tax rate</td>
<td>$\tau_s, \tau_u$</td>
<td></td>
<td>McDaniel (2007)</td>
</tr>
<tr>
<td>Government expenditure to GDP ratio</td>
<td>$G$</td>
<td>0.175</td>
<td>NIPA</td>
</tr>
</tbody>
</table>

Table 2: Parameter values calibrated

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
<th>Target</th>
<th>Model (Data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income share of $N_u$</td>
<td>$\mu$</td>
<td>0.327</td>
<td>Skill premium</td>
<td>1.474 (1.474)</td>
</tr>
<tr>
<td>Income share of $K$</td>
<td>$\lambda$</td>
<td>0.567</td>
<td>Capital-output ratio</td>
<td>2.27 (2.27)</td>
</tr>
<tr>
<td>Utility weight of leisure</td>
<td>$\chi$</td>
<td>7.9</td>
<td>Consumption-output ratio</td>
<td>0.60 (0.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Capital income share</td>
<td>0.288 (0.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gross investment-output ratio</td>
<td>0.218 (0.247)</td>
</tr>
</tbody>
</table>

productivity of the skilled’s “raw” labor relative to that of the unskilled’s and $\pi_t$ the number of members in the skilled household relative to the unskilled household.\footnote{In the extended model we assume that there are several members in a household, and that all members of the skilled household are skilled workers while all members of the unskilled household are unskilled workers.} Given $x_t = z_t \cdot \pi_t$, we shall focus on $\{x_t\}$ and $\{z_t\}$ instead of $\{z_t\}$ and $\{\pi_t\}$. Recall that we abstract from varying $\{x_t\}$ and $\{z_t\}$ and let $x_t = 1$ and $z_t = 1$ all the time in the benchmark model.

We briefly describe how we have obtained these time-series data.

**Tax rates $\{\tau_t\}$** The tax rate series are obtained directly from McDaniel (2007). We have already described them. These tax series will be viewed as the representation of the U.S. tax system in our model. For the years after 2013, we let the tax rates remain the same as those in the year 2013 so that the economy can converge to the new steady state. In computing the competitive equilibrium of the U.S. economy, they are applied uniformly to both the skilled and the unskilled.

**Skill premium $\{\xi_t\}$** 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 con-
stant the 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 representation in our Figure 1.22

**Investment-specific technological change \{q_t\}** Gordon (1990) is the seminal work on measuring investment-specific technological change. DiCecio (2009) constructed the relative price of capital by chainweighting the deflator for equipment and software from NIPA. DiCecio (2009)'s data sequence is updated at Federal Reserve Economic Data (FRED), a database maintained by the Federal Reserve Bank of St. Louis. The time series on the relative price of capital shown in Figure 2 are directly taken from the data. Given the investment-specific technological change \(q_t\), the relative price of capital is equal to \(1/q_t\). Thus, the time series on investment-specific technological change \(\{q_t\}\) are simply the reciprocal of the time series shown in Figure 2.

To obtain \(q_t\) beyond the data shown in Figure 2, we first compute the average growth rate of \(q_t\) from 2010 to 2014 and let it serve as the growth rate of \(q_t\) from 2014 to 2015. We then follow He and Liu (2008) method by assuming that the growth rates of \(\{q_t\}\) beyond year 2014 slow down linearly to zero from 2015 to 2072 and reach a steady state at 2072 and then remain constant from 2072 to 2142. In this way, we construct a time series sequence of \(\{q_t\}\) for a length of 180 years consisting of 52 years of data (1963-2014) from FRED and 128 years of artificial data (2015-2142) with 2072 to 2142 being the new steady state.

\(\{x_t\}\) and \(\{z_t\}\) Given the “deep” parameters of the model as we have calibrated, one can compute the transitional dynamics of competitive equilibrium of the U.S. economy from the initial steady state to a new steady state, once \(\{\tau_t\}\), \(\{q_t\}\), \(\{x_t\}\) and \(\{z_t\}\) are given. We obtain the sequences \(\{x_t\}\) and \(\{z_t\}\) according to the following algorithm:

1. Starting with \(\{x_t(1)\}\), compute the transitional dynamics of competitive equilibrium and thereby obtain from the model the allocation \(\{\hat{n}_{st}, \hat{n}_{ut}, \hat{K}_t, \hat{N}_{st}, \hat{N}_{ut}\}\).

2. Using the real-world data \(\{\xi_t\}\) and the allocation obtained from step 1, we calculate \(\{z_t\}\) by applying the formula (40) in the extended model, that is,

\[
z_t = \frac{\xi_t}{\left(\frac{(1-\mu)(1-\lambda)}{\mu}\right) \left[\lambda \left(\frac{\hat{K}_t}{\hat{N}_{st}}\right)^{\rho} + (1 - \lambda)\right]^{\sigma\frac{\sigma}{\sigma-1}} \left(\frac{\hat{N}_{ut}}{\hat{N}_{st}}\right)^{1-\sigma}}.
\]

22The data are available from Autor’s website.
3. Let $H_{st}$ (resp. $H_{ut}$) denote the total amount of hours worked by the skilled (resp. the unskilled) in the data at time $t$.\footnote{The data are directly from Autor (2014).} Using the obtained $\{\hat{n}_{st}, \hat{n}_{ut}\}$ from step 1, we calculate

$$\pi_t = \frac{(H_{st}/H_{ut})}{(\hat{n}_{st}/\hat{n}_{ut})},$$

where $\pi_t$ denotes the resulting number of members in the skilled household relative to that in the unskilled household at time $t$.

4. Given $\{z_t\}$ obtained from step 2 and $\{\pi_t\}$ obtained from step 3, we calculate

$$x_t(2) = z_t \cdot \pi_t.$$
5. Iterate until \( \{x_t(2)\} \approx \{x_t(1)\} \).

Figure 4 reports the resulting \( \{x_t\} \) and \( \{z_t\} \). Figure 5 shows the match between the real-world data \( \{\xi_t\} \) and the model data on the skill premium under the U.S. tax system. Note that the match is very well.

On the basis of the calibrated parameters and the obtained time-series data, we compute the transitional dynamics of optimal taxation using a nonlinear solution method in the spirit of Conesa and Krueger (1999) and He and Liu (2008). The details of the algorithm are relegated to the Appendix.

### 4.3 Quantitative results

Given \( q_t \) at a point in time, we obtain \( \tau_{K,t} > 0 \) and \( \tau_{L,t} > 0 \) at the optimum; see Figure 6. As explained earlier, the imposition of \( \tau_{K,t} > 0 \) is to imperfectly mimic the second-best situation under the third-best, while that of \( \tau_{L,t} > 0 \) is attributed to distributional concerns. Our central question is: How will this optimal tax structure at a point in time vary over time in the face of increasing \( q_t \)?

Figure 6 reports the dynamics of the optimal tax rate on capital income, \( \tau_{K,t} \), and that of the optimal marginal tax rate on labor income, \( \tau_{L,t} \). A feature clearly stands out: In the face of increasing \( q_t \), \( \{\tau_{K,t}\} \) displays a declining trend over time but \( \{\tau_{L,t}\} \) displays an increasing trend over time.\(^{24}\) Consider the utilitarian \( \psi^u/\psi^s = 1 \) between year 1965 and 2015. The

\(^{24}\)Note that optimal capital tax rates are rather high initially. This is a well-known feature of Ramsey taxation. As explained by Erosa and Gervais (2001), taxing the return on asset holdings that are accumulated...
optimal $\tau_{K,t}$ has decreased by about 32% (from $\tau_{K,t} = 0.50$ to 0.34), while the optimal $\tau_{L,t}$ has increased by about 62.5% (from $\tau_{L,t} = 0.14$ to 23.2%). This salient feature remains true across variation in Pareto weights $\psi^u/\psi^s \geq 1$. It is interesting to observe from McDaniel (2007) that, between year 1965 and 2013, the U.S. $\tau_{K,t}$ has decreased by about 17.6% (from $\tau_{K,t} = 0.33$ to 0.27), while the U.S. $\tau_{L,t}$ has increased by about 56.7% (from $\tau_{L,t} = 0.13$ to 0.21).

Figure 6 also shows that, with a fixed $q_t$ at a point in time, both $\tau_{K,t}$ and $\tau_{L,t}$ becomes higher as $\psi^u/\psi^s$ gets larger. This outcome is anticipated, in that a higher $\tau_{K,t}$ results in a lower skill premium and a higher $\tau_{L,t}$ implies a more redistribution. Werning (2007) explained his finding of perfect labor tax smoothing (i.e., the optimal marginal labor-income tax rate remains constant over time and unresponsive to aggregate shocks) 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 as well in our setting. This is logical because, like that in Werning (2007), the distribution of relative skill across workers is perfectly fixed and so remains constant over time in our setting. However, contrary to the expectation, we find that optimal $\tau_{L,t}$ is increasing over time. Why?

The optimal tax structure at any point in time balances distributional concerns against prior to date zero perfectly imitates a nondistortionary lump-sum tax; thus, the optimal tax rates will be set at confiscatory rates if unrestricted.
efficiency. Although distributional concerns directly related to the distribution of relative skill do remain unchanged over time in our setting, the efficiency aspect of the economy is changing in the face of increasing $q_t$.

When $q_t$ is fixed, the corresponding investment-specific technology is fixed too. The problem facing the planner is with regard to how to divide a fixed “pie” (associated with a fixed $q_t$) between the skilled and the unskilled.

As $q_t$ increases over time, the investment-specific technology is improving over time. Then the problem facing the planner is not only about how to divide a “pie” but also about whether to exploit improved technology to make the “pie” bigger. The planner can choose to either lean against the increasing $q_t$ by depressing the skill premium or lean toward the increasing $q_t$ by letting the skill premium increase. Our finding of cutting rather than raising $\tau_{K,t}$ over time at the optimum clearly indicates that the planner prefers the lean-toward to the lean-against policy. Indeed, the lean-toward policy takes the active form of cutting $\tau_{K,t}$ rather than the passive form of maintaining a constant $\tau_{K,t}$.

However, choosing this active lean-toward policy will surely exacerbate the redistributive aspect of the economy since it raises the skill premium through capital-skill complementarity. To remedy the worsening redistribution resulting from cutting $\tau_{K,t}$, the planner counterbalances it with an increasing progressivity of labor tax (i.e., a higher $\tau_{L,t}$) over time. This increasing progressivity of labor tax arises despite the distribution of relative skill across workers is perfectly fixed in our setting as in Werning (2007).

In a recent paper Costinot and Werning (2018) studied optimal technology regulation. Among others, they conducted a comparative statics in a simple economy to address the question: If improvements in new technologies make the economy’s inequality worse off, 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 instead of increase as the process of automation deepens and inequality increases. Although their model is static and its associated mechanism is different from ours, both their result and our finding point to the adoption of cutting rather than raising capital tax rates in the face of improved technology.

The right-hand panel of Figure 7 reports the evolution of $K/N_s$ and $N_u/N_s$ over time in the face of increasing $q_t$ (we include the case of laissez faire (no tax) in the figure). According to (10), while $K/N_s$ is the sole allocation that determines the capital-skill complementarity effect of the skill premium, $N_u/N_s$ is the sole allocation that determines the relative quantity effect of the skill premium. The evolutions of $K/N_s$ and of $N_u/N_s$ shown in the right-hand panel of Figure 7 explains why the skill premium shown in the left-hand panel of Figure 7 is on the rise.
An illustrated example  To showcase the main force underlying the declining optimal capital tax rates over time, we simplify the tax formula (24) by choosing specific parameter values for production technology and household preferences: \( \sigma = 1, \rho = 0, \gamma_c = 0, \gamma_n = 1, \chi = 1 \). As such, the production function (7) reduces to
\[
Y_t = \mu N_{ut} + (1 - \mu) K_t^\lambda N_{st}^{1-\lambda},
\]
while the utility function (27) reduces to
\[
U(c, n) = c + \log(1 - n).
\]
The skill premium at time \( t + 1 \) then becomes
\[
\xi_{t+1} = \frac{(1 - \mu)(1 - \lambda)}{\mu} \left( \frac{K_{t+1}}{N_{s,t+1}} \right)^\lambda,
\]
which shows that variation in \( \xi_{t+1} \) is completely determined by variation in \( \frac{K_{t+1}}{N_{s,t+1}} \).

From (24), we obtain a simplified formula for optimal capital tax rates:
\[
\tau_{K,t+1} = \frac{\lambda Y_{t+1}}{W_C \cdot (r_{t+1} - \frac{\delta}{q_{t+1}}) K_{t+1}},
\]
where \( W_C \) becomes a constant. If we let \( \delta \to 0 \), the above tax formula further reduces to
\[
\tau_{K,t+1} \approx \frac{Y_{t+1}}{W_C \cdot (1 - \mu) \left( \frac{K_t}{N_{st}} \right)^\lambda N_{s,t+1}},
\]
where we have utilized the FOC of the firm, \( r_t = (1 - \mu) \lambda \left( \frac{K_t}{N_{st}} \right)^\lambda \).

Numerical results show that the dynamic patterns exhibited in Figures 6-7 all remain qualitatively intact under the above simplified production technology and household preferences; see the Online Appendix. Importantly, it is found that, the “capital-skill complementarity effect” coined by Krusell et al. (2000) and embedded in the term \( \frac{K_{t+1}}{N_{s,t+1}} \) in both (28) and (29), is the key force that drives the capital tax rate down and the skill premium up simultaneously in the face of increasing \( q_t \) at the optimum.

To gain further insights into the results of Figures 6, we ask two “what if” questions under \( \psi^u/\psi^s = 1 \).
What if $q_t = 1$ at all times? That is, what would be if there were no investment-specific technological change? Figures 8 reports the results of $q_t = 1$ at all times and compare them with those of varying $\{q_t\}$. It is clear from the figure that neither declining $\tau_{K,t}$ nor increasing $\tau_{L,t}$ would arise at the optimum if $q_t = 1$ at all times.

What if $\sigma = \rho$ in the production technology (7)? That is, what would be if there were no capital-skill complementarity in production? For simplicity, we focus on the case where $\sigma = \rho = 0$ so that the production technology (7) reduces to $Y = N_0^\mu N_s^{(1-\mu)(1-\lambda)} K^{(1-\mu)\lambda}$,
a Cobb-Douglas production function. Note that $\frac{\partial c}{\partial K} = 0$ if $\sigma = \rho$ were to hold. Figure 9 reports the findings. It is shown that capital should go untaxed and that labor tax should be perfectly smoothed over time. Both results are consistent with Proposition 2 in Werning (2007), in which all types of labor are equally complementary to capital in production.

5 Mirrlees problem

We turn to the Mirrlees approach and formulate the Mirrlees problem in this section. Unlike the Ramsey approach, there are no \textit{a priori} restrictions placed on the tax scheme, except that taxes imposed cannot condition directly on household types. As noted by Werning (2007) and Kocherlakota (2010, p. 3), this restriction ends up implying that the planner de facto must treat households as being privately informed about their own skills even if this may not be true. Following the seminal work of 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 the celebrated work of Stiglitz (1982).

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.

As is standard, households are assumed to have no outside opportunities available in the face of taxation. 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’ type. Thus, each household reports its skill type and receives an allocation $\{c_t, n_t\}$ as a function of this report such that the allocation is required to satisfy the incentive-compatibility (IC) constraints:

$$V(s) \geq V^u(s), \quad (30)$$

$$V(u) \geq V^s(u), \quad (31)$$

with

\[As time passes, information about household types (skilled or unskilled) may be revealed. The social planner is assumed to commit to the contract without exploiting information revelation as time passes. It is known that the society is better off with such a commitment than without it; see, for example, Laffont and Tirole (1988).\]
\[ V^u(s) = \sum_{t=0}^{\infty} \beta^t U(c_{ut}, 1 - \frac{n_{ut}}{\xi_t}), \]
\[ V^s(u) = \sum_{t=0}^{\infty} \beta^t U(c_{st}, 1 - \xi_t n_{st}), \]

where \( V^j(i) \) with \( i \neq j \) denotes type \( i \) household’s level of utility derived from deceptively mimicking 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. \( \frac{w_{ut} n_{ut}}{w_{st}} = \xi_t n_{st} \)), which appears in \( V^u(s) \) (resp. \( V^s(u) \)).

The IC constraints characterized by (30)-(31) are key elements of the Mirrlees approach. The inequality \( V(s) \geq V^u(s) \) of (30) 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 (31) 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 (30)-(31) relative to their absence.

Given initial capital \( K_0 \), the Mirrlees problem is to choose the allocation \( \{c_{it}, n_{it}, K_t\} \) with \( i \in \{s, u\} \) so as to maximize a SWF defined in (13), subject to the resource constraints (12) and the IC constraint (30) with equality. Kocherlakota (2010, section 3.2) showed that the set of allocations \( \{c_{it}, n_{it}, K_t\} \) that are achievable by the society under the Mirrlees approach are exactly the ones that satisfy the IC constraints and the resource constraints.

Let \( U(c, 1 - n) = u(c) + v(1 - n) \), and let \( \Lambda \) and \( \{\beta^t \Gamma_t\} \) be the multipliers on the IC constraint (30) and resource constraints (12), respectively. Then the FOCs for the planner problem are given by

\[ c_{it} : \quad u'(c_{st}) = \frac{\Gamma_t}{\psi_s + \Lambda}, \quad u'(c_{ut}) = \frac{\Gamma_t}{\psi_u - \Lambda}, \]
\[ n_{ut} : \psi u' (1 - n_{ut}) = \Gamma_t w_{ut} + \Lambda v' (1 - n_{ut}) \left[ \frac{1}{\xi_t} - \frac{n_{ut}}{\xi_t^2} \frac{\partial \xi_t}{\partial n_{ut}} \right], \quad (33) \]

\[ n_{st} : [\psi^s + \Lambda] v' (1 - n_{st}) = \Gamma_t w_{st} - \left[ \Lambda v' (1 - n_{ut}) \frac{n_{ut}}{\xi_t^2} \right] \frac{\partial \xi_t}{\partial n_{st}}, \quad (34) \]

\[ K_t : \Gamma_t = \beta \left\{ \Gamma_{t+1} \left[ r_{t+1} + \frac{(1 - \delta)}{q_{t+1}} \right] - \Delta_{t+1} \right\}, \quad (35) \]

where \( \Delta_{t+1} = \left[ \Lambda v' (1 - \frac{n_{ut+1}}{\xi_{t+1}}) \frac{n_{ut+1}}{\xi_{t+1}^2} \right] \frac{\partial \xi_{t+1}}{\partial K_{t+1}} \). The optimal condition for capital (35) is different from the standard one due to the presence of the term \( \Delta_{t+1} \), which would be equal to zero if \( \frac{\partial \xi_{t+1}}{\partial K_{t+1}} = 0 \) (i.e., if there were no capital-skill complementarity with \( \sigma = \rho \)). We denote the constrained efficient allocation resulting from (32)-(35) plus the resource constraints (12) and the equality of (30) by \((c^*_i, n^*_i, K^*)\) with \( i \in \{s, u\} \).

Following the literature on optimal taxation, we define wedges or distortions as the differences between the marginal rates of substitution in preferences and the marginal rates of transformation in production. All the wedges would be equal to zero if distortionary government interventions were absent.

On the basis of (5b), the intertemporal wedge on capital, call it \( \varpi_{K_{t+1}} \), is defined to be

\[
\frac{u'(c_{it})}{q_{t+1} \beta u'(c_{it+1})} = \frac{1}{q_{t+1}} + \left( 1 - \varpi_{K_{t+1}} \right) \left( r_{t+1} - \frac{\delta}{q_{t+1}} \right),
\]

where \( i \in \{s, u\} \). If \( q_{t+1} = 1 \) (i.e., there were no investment-specific technological change), the above intertemporal wedge would reduce to the familiar one. Contrasting this intertemporal wedge with the optimal intertemporal condition (35) yields

\[
\left[ \frac{1}{q_{t+1}} + \left( 1 - \varpi_{K_{t+1}} \right) \left( r_{t+1} - \frac{\delta}{q_{t+1}} \right) \right] = \frac{u'(c_{it})}{q_{t+1} \beta u'(c_{it+1})} = \left[ r_{t+1} + \frac{(1 - \delta)}{q_{t+1}} - \frac{\Delta_{t+1}}{\Gamma_{t+1}} \right],
\]

which in turn yields at the optimum

\[
\varpi_{K_{t+1}} = \frac{\Delta_{t+1}}{(r_{t+1} - \frac{\delta}{q_{t+1}}) \Gamma_{t+1}}, \quad (36)
\]

where \( \Delta_{t+1} > 0 \) because \( \frac{\partial \xi_{t+1}}{\partial K_{t+1}} > 0 \); the term \( r_{t+1} - \frac{\delta}{q_{t+1}} \) represents the return on capital net of depreciation. The result of (36) indicates that a positive distortion should be imposed on capital.

On the basis of (5a), the intratemporal wedges on skilled and unskilled labor are defined by

\[
\varpi_{st} = 1 - \frac{v' (1 - n_{st})}{w_{st}} / u'(c_{st}); \quad \varpi_{ut} = 1 - \frac{v' (1 - n_{ut})}{w_{ut}} / u'(c_{ut}),
\]
which are standard. By the same token, we can back out these wedges at the optimum by contrasting them with the optimal intratemporal conditions (33) and (34):

\[
\varpi_{st} = \frac{\Lambda v' \left( 1 - \frac{n_{ut}}{\xi_t} \right) \left( \frac{n_{ut}}{\xi_t} \right) \partial \xi_t}{\psi s + \Lambda} w_{st} w' (c_{st}) < 0; \quad \varpi_{ut} = \frac{\Lambda}{\psi u \xi_t} \left[ 1 - \frac{v' \left( 1 - \frac{n_{ut}}{\xi_t} \right)}{w_{st} w' (c_{ut})} \right] \left[ 1 - \frac{N_{ut}}{\xi_t} \frac{\partial \xi_t}{\partial N_{ut}} \right],
\]

(37)

where \( \frac{\partial \xi}{\partial N_s} < 0 \) stems from the comparative statics of \( \xi \) with respect to \( N_s \); thus, \( \varpi_{st} < 0 \). The sign of \( \varpi_{ut} \) is not definite, depending on whether the term \( \frac{N_{ut}}{\xi_t} \frac{\partial \xi_t}{\partial N_{ut}} \) is elastic or not. However, our quantitative results later show that \( \varpi_{ut} > 0 \).

Following the idea of Slavík and Yazici (2014), one can implement the constrained efficient allocation \((c^*_i, n^*_i, K^*)\) as part of a competitive equilibrium via a time-varying flat tax rate on capital income and a time-varying nonlinear tax schedule on labor income that satisfy:

\[
\tau_K(t + 1) = \varpi_{K_{t+1}};
\]

(38)

\[
\tau_s(t) = \varpi_{st}; \quad \tau_u(t) = \varpi_{ut},
\]

(39)

where \( \tau_K(t + 1) \) is the flat tax rate on capital income, and \( \tau_s(t) \) and \( \tau_u(t) \) are, respectively, the implicit marginal tax rates on labor income of the skilled and that of the unskilled implied by the nonlinear labor tax schedule.\(^{26}\)

Given \((c^*_i, n^*_i, K^*)\) with \( \varpi_{K_t} > 0, \varpi_{st} < 0 \) and \( \varpi_{ut} > 0 \) for any \( q_t > 0 \), we have \( \tau_K(t + 1) > 0, \tau_s(t) < 0 \) and \( \tau_u(t) > 0 \) for any \( q_t > 0 \) at the optimum.\(^{27}\) What is the intuition underlying this optimal tax structure at a point in time? We offer a brief explanation.

With \( V(s) = V^u(s) \) in equilibrium, a critical property of \( V^u(s) \) is that the skilled will need to expend a labor input of \( \frac{n^u}{\xi_t} \) in order to pretend to be the unskilled; see the definition of \( V^u(s) \) in (30). As such, given \( n^u_i \), the skill premium \( \xi_t \) precisely captures how easy or difficult it is for the skilled to mimic the unskilled: The higher the skill premium, the less labor input required and hence the easier the task of mimicking. Now suppose that the skill premium \( \xi_t \) increases. Then, all else being equal, the required labor input for mimicking will decrease, implying that the skilled will be able to enjoy more leisure and thereby obtain a higher \( V^u(s) \) from mimicking. This tightening of the IC constraint (30) due to an increase in the skill premium clearly worsens social welfare. It then suggests that a benevolent government will have incentives to impede the increase in the skill premium. However, how does a Mirrleesian

\(^{26}\)See the Online Appendix for the detail on implementation.

\(^{27}\)It is worth noting that 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).
planner accomplish the job?

Given the plausible parameter restriction $\sigma > \rho$ and $\sigma < 1$, we see from (10) that variation in the skill premium is determined by two effects: (i) the capital-skill complementarity effect – capital relative to the skilled labor input $(K/N_s)$; (ii) the relative quantity effect – the unskilled labor input relative to the skilled $(N_u/N_s)$. Thus, there are two natural channels for the planner to accomplish her job of impeding the increase in the skill premium: (i) via the capital-skill complementarity effect by imposing taxes on capital so as to retard its growth, and (ii) via the relative quantity effect by subsidizing skilled workers or taxing unskilled workers so as to increase the skilled’s labor input relative to the unskilled’s. Both channels relax the IC constraints through obstructing the increase in the skill premium. At the same time they determine the structure of optimal taxation at a point in time: A positive tax rate on capital; a positive marginal tax rate on unskilled labor, but a negative marginal tax rate on skilled labor. The result of a positive capital tax rate differs from the zero-tax prescription in Werning (2007). It suggests that, in the presence of capital-skill complementarity, it is always desirable for a benevolent government to relax the IC constraints by imposing taxes on capital so as to impede the increase in the skill premium.

Relaxing IC constraints via the manipulation of market prices to improve social welfare is not a new idea in the literature; see, for example, Naito (1999) in the absence of capital and Slavík and Yazici (2014) in the presence of capital. Our contribution to the literature is represented in the next section, where we explore how the optimal tax structure at a point in time will vary over time in the face of the secular cheapening of capital with capital-skill complementarity.

6 Dynamics of Mirrleesian taxation

This section reports the dynamics of Mirrlees taxation in the face of the series $\{q_t\}$. We let the period utility function $U(.)$ take the form of (27) as in Ramsey taxation.

Given $q_t$ at a point in time, we find that $\tau_K(t) > 0$, $\tau_s(t) < 0$ and $\tau_u(t) > 0$ at the optimum; see Figure 10. As explained earlier, this tax structure is to depress the skill premium so as to relax the IC constraints. Like the Ramsey approach, our central question is: How will this optimal tax structure at a point in time vary over time in the face of increasing $q_t$?

Given initial capital $K_0$, the left-hand panel of Figure 10 reports the dynamics of the optimal tax rate on capital income, $\tau_K(t)$, while the right-hand panel of Figure 10 reports the dynamics of the optimal marginal tax rates on skilled and unskilled labor, $\tau_s(t)$ and $\tau_u(t)$. Two features clearly stand out. First, $\{\tau_K(t)\}$ displays a declining trend over time. Second, both $\{\tau_s(t)\}$ and $\{\tau_u(t)\}$ display an increasing trend over time and this upward trend is much
Figure 10. Mirrleesian taxation

more significant for $\{\tau_s(t)\}$ than for $\{\tau_u(t)\}$. Consider the utilitarian $\psi^u/\psi^s = 1$ between year 1965 and 2015. The optimal $\tau_K(t)$ has decreased by about 31.9% (from $\tau_K(t) = 0.20$ to 0.14), while the optimal $\tau_s(t)$ has increased by about 54.8% (from $\tau_s(t) = -0.17$ to $-0.08$) and the optimal $\tau_u(t)$ has increased by about 4.2% (from $\tau_u(t) = 0.30$ to 0.33). These two salient features remain true across variation in Pareto weights $\psi^u/\psi^s \geq 1$.

Figure 10 also shows that, given $q_t$ at a point in time, both $\tau_K(t+1) > 0$ and $\tau_u(t) > 0$ become higher while $\tau_s(t) < 0$ becomes lower at the optimum as $\psi^u/\psi^s$ increases. This result is intuitive. As $\psi^u/\psi^s$ 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 need to further depress the skill premium. This explains why both $\tau_K(t+1) > 0$ and $\tau_u(t) > 0$ get higher while $\tau_s(t) < 0$ gets lower at the optimum, since these tax adjustments will all depress the skill premium.

Like the Ramsey planner, the Mirrlees planner chooses the active policy of cutting capital tax rates to accommodate improved technology. Since this policy raises the skill premium through capital-skill complementarity, it will exacerbate the redistributive aspect of the economy as in the Ramsey problem. In addition, it will tighten the IC constraint through the rising skill premium for the Mirrlees problem. Figure 10 shows how the Mirrlees planner adjusts labor taxes in response to these two consequences. First, the optimal marginal tax rate on unskilled labor, $\tau_u(t) > 0$, will be on the rise, though only slightly. The rising $\tau_u(t) > 0$ can depress the skill premium through the relative quantitative effect and thereby can relax the IC constraint tightened resulting from the cut in capital tax rates. Second, the optimal marginal tax rate on skilled labor, $\tau_s(t) < 0$, will be on the rise too. The rising
τ_s(t) < 0 can lessen the worsening redistribution of the economy via cutting the marginal subsidy to skilled labor.

In the Ramsey approach, we interpret the marginal-tax-rate component of the linear labor tax schedule as representing the progressivity of labor taxation. In the Mirrles approach with the nonlinear labor tax schedule, we let the difference between τ_s(t) and τ_u(t), i.e., τ_s(t) − τ_u(t), measure the progressivity of labor taxation. That is, a higher τ_s(t) − τ_u(t) represents a more progressivity of labor taxation at the margin. Given that τ_s(t) < 0 and τ_u(t) > 0 at the optimum, a higher τ_s(t) − τ_u(t) actually represents a less regressivity of labor taxation at the margin.

Since τ_s(t) increases substantially but τ_u(t) increases only slightly according to Figure 10, the regressivity of the nonlinear labor tax schedule at the margin, measured by τ_s(t) − τ_u(t), is decreasing to a significant extent over time.

The reason underlying the optimal tax structure at a point in time in Mirrlesian taxation is different from Ramsey taxation. Nevertheless, a plausible explanation seems equally applicable to their dynamics: The planner adopts the lean-toward policy to reap the gains of improved technology by cutting capital tax rates and counterbalances the resulting deterioration of redistribution via an increasing progressivity or a decreasing regressivity of labor tax.

The left-hand panel of Figure 11 reports the evolution of the skill premium over time in the face of increasing q_t, while the right-hand panel of Figure 11 reports the evolution of K/N_s and N_u/N_s over time. As in the Ramsey problem, we include the case of laissez faire (no tax) in the figure. Note that all the dynamic patterns shown in Figure 11 are qualitatively of no difference from those found in Figure 7 for the Ramsey problem.
Figure 12. Mirrleesian taxation: $q_t = 1$ vs. $q_t = \text{data}$

We choose the same specific parameter values for production technology and household preferences as the illustrated example in Ramsey taxation. Numerical results show that the dynamic patterns exhibited in Figures 10-11 all remain qualitatively intact under the simplified production technology and household preferences; see the Online Appendix.

To gain further insights into the results of Figure 10, as in the Ramsey approach, we ask two "what if" questions under $\psi_u / \psi_s = 1$.

**What if $q_t = 1$ at all times?** That is, what would be if there were no investment-specific technological change? Figures 12 reports the results of $q_t = 1$ at all times and compare them with those of varying $\{q_t\}$. It is clear from the figure that there is a substitution of labor tax for capital tax with varying $\{q_t\}$, whereas there is no such a substitution as $q_t = 1$ at all times.

**What if $\sigma = \rho$ in the production technology (7)?** That is, what would be if there were no capital-skill complementarity in production? For simplicity, as in the Ramsey approach, we focus on the case where $\sigma = \rho = 0$ so that the production technology (7) reduces to $Y = N^\mu K^{(1-\mu)(1-\lambda)}$, a Cobb-Douglas production function. Since $\frac{\partial \xi}{\partial K} = 0$ if $\sigma = \rho$ were to hold, it follows from (38) that $\tau_K(t+1) = 0$ at the optimum all the time. This result is confirmed in the left-hand panel of Figure 13. The right-hand panel of Figure 13 reports the finding with regard to labor marginal tax rates, which shows that both $\tau_s(t)$ and $\tau_u(t)$ remain constant over time at the optimum. Both results, that capital should go untaxed and that labor tax should be perfectly smoothed over time, are consistent with Proposition 5 in Werning (2007), in which all types of labor are equally complementary to capital.
Figure 13. Mirrleesian taxation as $\sigma = \rho = 0$

production.

7 Extension

This section extends the benchmark model to allow for varying $\{x_t\}$ and $\{z_t\}$, in addition to varying investment-specific technological changes $\{q_t\}$. All of $\{q_t\}$, $\{x_t\}$ and $\{z_t\}$ are exogenous. To simplify exposition and abstract from complication, it is assumed that people are endowed with perfect foresight regarding them.

In the extended model both types of households have the size of unit measure as in the benchmark model. However, the productivity of the skilled’s “raw” labor is no longer equal to that of the unskilled’s as in the benchmark model. Moreover, there are now several members rather than a single member in the household. We assume that all members of the skilled household are skilled workers, while all members of the unskilled household are unskilled workers. In the presence of $\{x_t\}$, labor income earned by the skilled household at time $t$ becomes $w_{st}x_t n_{st}$ rather than $w_{st}n_{st}$ in the benchmark model. Also, the skill premium defined in (9) is modified to be

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

(40)

where $z$ denotes the productivity of the skilled’s “raw” labor relative to that of the unskilled’s.

Under $\psi^u/\psi^s = 1$, figure 14 reports the dynamics of Ramsey taxation and 15 reports the dynamics of Mirrleesian taxation. It is shown that the dynamic patterns of optimal taxation remain qualitatively the same as in the benchmark model.
From (40), we obtain

$$\ln \xi \approx \text{constant} + \lambda \left( \frac{\sigma - \rho}{\rho} \right) \left( \frac{K}{N_s} \right)^\rho + (1 - \sigma) \ln \left( \frac{N_u}{N_s} \right) + \ln z,$$  \hspace{1cm} (41)

where the term $\ln z$ is absent in the benchmark model; see (10). A faster growth in the skilled relative to the unskilled labor input can bring down the rise of the skill premium through the relative quantitative effect. However, this effect can be offset by an increase in the labor productivity of the skill relative to the unskilled, represented by the term $\ln z$ in (41). Figure 4 shows that the relative productivity $z$ did increase significantly over the sample period we study. Overall, the “capital-skill complementarity effect” embedded in the term
$K/N_s$ in (41) still dominates the evolution of the skill premium as in the benchmark model. This explains why the dynamics of optimal taxation remains qualitatively the same as in the benchmark model.

8 Conclusion

The issue of income inequality is nowadays one of the focal points in the world. Former U.S. President Obama, speaking before Congress on Dec 4th, 2013, emphasized that the vastly increased income gap is “the defining challenge of our time.” Atkinson et al. (2011) surveyed recent studies on the time series of top income shares using income tax statistics across countries. They found among other things that a significant portion of the increase in the top income share is due to an increase in top labor income, especially wages and salaries.

Based on the premise of capital-skill complementarity in production, Krusell et al. (2000) found that the rising skill premium since the early 1980s can be well explained through the secular cheapening of capital due to investment-specific technological change. This finding holds despite there was a substantial increase in the relative supply of college skills during the sample period they studied. A sensible and interesting question in light of this influential work is: In the simultaneous presence of capital-skill complementarity and secular cheapening of capital, how should taxation be set dynamically in response to the rising skill premium? In this paper we make an attempt to answer the question.

Two main results emerge, no matter whether adopting the Ramsey or the Mirrlees approach. First, capital income should be taxed but, over time, it is optimal to cut capital tax rates in order to accommodate improved technology. Second, the planner substitutes labor tax for capital tax and implements an increasing progressivity of labor tax over time to remedy the worsening redistribution resulting from cutting capital tax rates. The first result is consistent with the finding of Costinot and Werning (2018) that despite distributional concerns create a rationale for positive taxes on new technologies, the magnitude of these taxes may decrease instead of increase as the process of automation deepens and inequality increases.

Our paper focuses on the tax policy response to the rising skill premium between the skilled and the unskilled. The paper could be extended in several directions. For example, allow for more than two types of workers to explore other issues, such as job polarization as exposed by Autor and Dorn (2013). Another direction is to address the limited rather than the full commitment of the government. We plan to pursue these research projects in the future.
References


9 Appendix

Computation method

We compute the transitional dynamics of optimal taxation by using the following procedure:

1. Given the sequence \( \{ q_t \} \) with \( q_0 = 1 \), we compute the steady state of the optimal allocation for the Ramsey problem and that of the constrained efficient allocation for the Mirrless problem at each \( q_t \). We then use these steady-state allocations as an initial guess for computing the transitional dynamics.

2. In the case of the Ramsey problem, we solve for the system of nonlinear equations (19)-(23) plus the implementability conditions (17), the resource constraints (12), the restriction (18), and the FOC with respect to \( \varphi \). In the case of the Mirrlees problem, we solve for the system of nonlinear equations (32)-(35) plus the equality of the IC constraint (30) and the resource constraints (12). We substitute the resulting allocations into the IC constraints (30) and (31) to make sure that the incentive compatibility constraints for skilled and unskilled households are not violated. Finally, we use these allocations to back out the optimal tax rates through equations (24) and (25) for the Ramsey problem and through equations (38) and (39) for the Mirrlees problem.

9.1 Implementation

This section illustrates how the constrained efficient allocation can be implemented by a given tax system. A tax system is said to implement the constrained efficient allocation if an allocation of competitive equilibrium given a tax system is consistent with constrained efficient allocation.

We firstly assume there is no uncertainty in the economy and all agents can perceive the whole process of the exogeneous shocks. In other words, we focus our implementation problem on a perfect foresight model. Agents can save at period \( t \) by buying one period bond \( b_{t+1} \) with interst rate \( R_{bt} \) or by investing in capital \( k_{t+1} \) with price \( \frac{1}{q_t} \) for production. There is a representative firm that rents capital and hire labor to produce the output. The rental rate \( r_t \), wage rates for skilled and unskilled are \( w_{st} \) and \( w_{ut} \) and are taken are given by the firm.

Shocks and Taxes

Since we focus on a perfect foresight dynamic model with steady state, investment shock \( q \equiv \{ q_t \}_{t=0}^{\infty} \) is deterministic and exogeneous sequences converging to some constants after \( T > 0 \), \( q_t = \bar{q} \) for all \( t > T \).
The tax functions we defined are modified from Slavik and Yazici (2014). We define infinite countable income as \( y \equiv \{ y_t \}_{t=0}^{\infty} \) in which \( y_t \) is a real function mapping from \( q \) to \( \mathbb{R}_+ \). Given the definitions of the labor income and asset sequences, we also define a stream of labor income tax rates and \( \{ \tau_{y,t} \}_{t=0}^{\infty} \) and capital tax rates \( \{ \tau_{k,t+1} \}_{t=0}^{\infty} \) such that \( \tau_{y,t} : y_t \rightarrow \mathbb{R} \) and \( \tau_{k,t+1} : q \rightarrow \mathbb{R} \).

**Agent’s problem** Given the price \( \{ r_t, w_{st}, w_{ut}, R_{bt} \}_{t=0}^{\infty} \) and tax rates \( \{ \tau_{y,t}, \tau_{k,t+1} \}_{t=0}^{\infty} \), agent \( i \) solve the following problem:

\[
\max \sum_{t=0}^{\infty} \beta^t U \left( c_{it}, 1 - \frac{y_{it}}{w_{it}} \right)
\]

subject to

\[
c_{it} + b_{it+1} + I_{it} = y_{it} - \tau_{y,t}(y_{it}) + R_{bt} b_{it} + (1 - \tau_{k,t}) \left( r_{kt} - \frac{\delta q_{it}}{q_{it}} \right) k_{it}
\]

and

\[
k_{it+1} = (1 - \delta) k_{it} + q_{it} I_{it}
\]

**Firm’s problem** Taking price \( \{ r_t, w_{st}, w_{ut} \}_{t=0}^{\infty} \) as given, the representative firm solves

\[
\max F (N_{ut}, K_t, N_{st}) - w_{st} N_{st} - r_t K_t - w_{ut} N_{ut}
\]

**Equilibrium under tax program**

Given investment specific shocks \( q \), tax system \( \{ \tau_{y,t}, \tau_{k,t+1} \}_{t=0}^{\infty} \), bond policy \( \{ B_{t+1} \}_{t=0}^{\infty} \) and initial asset holdings \( \{ k_{i0}, b_{i0} \}_{i \in \{s,u\}} \), an equilibrium is an allocation \( \{ y_{it}, k_{it+1}, b_{it+1}, c_{it} \}_{t=0}^{\infty} \) for agent \( i \in \{ s, u \} \), \( \{ N_{ut}, K_t, N_{st} \}_{t=0}^{\infty} \) for the representative firm and prices \( \{ r_t, w_{st}, w_{ut}, R_{bt} \} \) satisfying the following conditions:

1. \( \{ y_{it}, b_{it+1}, I_{it}, c_{it} \}_{t=0}^{\infty} \) solves decision problem of agent \( i \in \{ s, u \} \).
2. \( \{ N_{ut}, K_t, N_{st} \}_{t=0}^{\infty} \) solves the profit maximization problem of the firm.
3. Market clearing conditions

\[
\begin{align*}
\frac{y_{st}}{w_{st}} &= N_{st} \\
\frac{y_{ut}}{w_{ut}} &= N_{ut} \\
\sum_{i \in \{s, u\}} b_{it} &= B_t \\
\sum_{i \in \{s, u\}} k_{it} &= K_t
\end{align*}
\]

4. Aggregate variables are consistent with

\[
\begin{align*}
\sum_{i \in \{s, u\}} c_{it} &= C_t \\
\sum_{i \in \{s, u\}} k_{it} &= K_t \\
\sum_{i \in \{s, u\}} b_{it} &= B_t
\end{align*}
\]

5. Resource constraint holds in each period:

\[
C_t + K_{t+1} + G_t \leq F(N_{ut}, K_t, N_{st}) + (1 - \delta) \frac{K_t}{q_t}
\]

Implementation

Let constrained allocation \( A \equiv \{c_{st}^*, c_{ut}^*, K_t^*, N_{st}^*, N_{ut}^*\} \) that solve Mirrleesian problem we posed in Section 4 under exogeneous shocks \( q \). A tax system tax system \( \{\tau_{y,t}, \tau_{k,t+1}\}_{t=0}^{\infty} \) implements constrained allocation if \( \{c_{st}^*, c_{ut}^*, K_t^*, N_{st}^*, N_{ut}^*\} \) corresponds with the competitive equilibrium \( \{c_{it}, y_{it}\} \) for agent \( i \in \{s, u\} \) and \( \{K_t^*, N_{st}^*, N_{ut}^*\} \) for firm’s inputs. We will specify the optimal tax system and prove that it indeed implement the constrained efficient alloctaion and characterize its properties.

Given the exogeneous process \( q \), we first define \( \{r_t^*, w_{st}^*, w_{ut}^*, p_{bt}^*\} \) and \( Y_t^* \equiv \{y_{st}^*, y_{ut}^*\}^{\infty}_{t=0} \) are factor prices labor income derived from \( A \), we set tax system \( J \equiv \{\tau_{y,t}, \tau_{k,t+1}\}_{t=0}^{\infty} \) and bond policy \( \{B_{t+1}\}_{t=0}^{\infty} \) as follows where \( Y_t^* \) is the set of labor income histories observed at the constrained efficient allocation and \( R_{at+1}^* \). As for income tax \( \{\tau_{i}(y_{it})\} \), we pose the following settings:
1. For capital income tax rates:

\[ \tau_{k,t+1} = 1 + \frac{1 - MRS_{t,t+1}^* q_{t+1}/q_t}{(q_{t+1}r_{t+1} - \delta)} \text{ if } y_t \in \mathcal{Y}_t^* \]

\[ = \infty \text{ otherwise} \]

where \( MRS_{t,t+1}^* \equiv \frac{U^* (c_{it}^*, y_{it}^*)}{\beta U^* (c_{it+1}^*, y_{it+1}^*)} \)

2. For bond policy and asset distributive law:

\[ B_{t+1}^* = \sum_{i \in \{s,u\}} \tau(y_{it}^*)y_{it}^* + \tau_{kt}^* \left( r_{t}^* - \frac{\delta}{q_t} \right) K_{it}^* + B_{t}^* - G_t \]

given initial asset \( a_{it0} \) distribution law:

\[ a_{it0} = \sum_{t=0}^{\infty} Q^t \left[ c_{it}^* - y_{it}^* - \tau(y_{it}) \right] \]

where

\[ Q^t = \prod_{m=0}^{t} \frac{1}{(1 + R_{b,m})} \]

3. For marginal tax rates:

\[ \tau'(y_{it}) = 1 + \frac{U_{y}(c_{it}^*, y_{it}^*)}{U_{c}(c_{it}^*, y_{it}^*)} \text{ if } y_t \in \mathcal{Y}_t^* \]

\[ = 2y_t \text{ otherwise} \]

**Proposition** The optimal tax system \( \mathcal{J} \) together bond policy \( \{B_{t+1}^*\} \) with described above implements the constrained efficient allocation.

**Proof.**

First, define factor prices of constrained efficient allocation as

\[ r_{t}^* = F_K (N_{ut}^*, K_{it}^*, N_{st}^*) \]

\[ w_{st}^* = F_N (N_{ut}^*, K_{it}^*, N_{st}^*) \]

\[ w_{ut}^* = F_N (N_{ut}^*, K_{it}^*, N_{st}^*) \]

\[ R_{bt}^* = MRS_{t,t+1}^* - 1 \]

and we will investigate what allocations will be under the tax system \( \mathcal{J} \) and prices \( (r_{t}^*, w_{st}^*, w_{ut}^*, R_{bt}^*) \).
Second, observe that the only budget feasible income strategy for the household is the one that corresponds to some agent’s income process at the constrained efficient allocation in \( \{ y_{jt}^* \}_{t=0}^\infty \). In other word, the agent is either choose \( \{ y_{st}^* \}_{t=0}^\infty \) or choosing \( \{ y_{ut}^* \}_{t=0}^\infty \).

Third, we claim that if an agent \( i \) chooses one of the income path \( \{ y_{jt}^* \}_{t=0}^\infty \), \( j \in \{ s, u \} \), then facing the prices defined above, the agent \( i \) will also also chooses.

If these claims can be proved, then we can verify the tax system we proposed can indeed implement constrained efficient allocations since the constrained efficient allocation is incentive compatible.

To verify this claim, take an agent \( i \) that follows income strategy \( \{ y_{jt}^* \}_{t=0}^\infty \), the decision problem now becomes

\[
\max_{c_{it}, k_{it+1}, b_{it+1}} \sum_{t=0}^\infty \beta^t U \left( c_{it}, 1 - \frac{y_{jt}^*}{w_{it}^*} \right)
\]

subject to

\[
c_{it} + \frac{k_{it+1}}{q_t} + b_{it+1} = y_{jt}^* - \tau_{yt}(y_{jt}^*) - \tau_{k,t+1} \left( r_{t+1}^* - \frac{\delta_t}{q_t} \right) + (1 + R_{bt}^*) b_{it}
\]

The first-order conditions to this problem are the budget constraint with equality and

\[
\frac{U_c \left( c_{it}, \frac{y_{jt}^*}{w_{it}^*} \right)}{\beta U_c \left( c_{it}, \frac{y_{jt+1}^*}{w_{it+1}^*} \right)} = \left( 1 + R_{bt+1}^* \right) = q_t \left( \frac{1}{q_{t+1}} + (1 - \tau_{k,t+1}^*) \left( r_{t+1}^* - \frac{\delta}{q_{t+1}} \right) \right)
\]

Since the agent’s problem is concave, these first-order conditions are necessary and sufficient for optimality provided that a relevant transversality condition holds. By applying guess and verify approach, we can check that \( \{ c_{jt}^*, b_{jt}^*, k_{jt}^* \}_{t=0}^\infty \) satisfy the Euler equations and flow constraint every period given the tax system \( J \) and prices \( (r_t^*, w_{st}^*, w_{ut}^*, R_{bt}^*) \).

Finally, one also needs to make sure that the also firm chooses the correct allocation. The firm’s optimality conditions for labor are satisfied at the constrained efficient allocation by construction of wages. The firm’s optimality conditions are

\[
r_t^* = F_K (N_{ut}^*, K_t^*, N_{st}^*)
\]

\[
w_{st}^* = F_N (N_{ut}^*, K_t^*, N_{st}^*)
\]

\[
w_{ut}^* = F_N (N_{ut}^*, K_t^*, N_{st}^*)
\]

### 9.2 Robust check

In this section we document how sensitive our results are by varying the elasticity of substitution between capital and unskilled workers, letting the parameter \( \sigma \) change to 0.5\( \sigma \) and
1.5σ respectively. We report the results are as follows.

Firstly, we find both capital and labor income tax rates increase with parameter σ in Ramsey taxation case. During 1965-2015, capital tax rates decrease on average 18.5% and labor tax rates decrease 16.9% when σ change to 0.5σ. On the other hand, both capital and labor tax rates increase on average 23% and 21.6% respectively as σ change to 1.5σ. This experiment suggests that with larger elasticity of substitution difference, government would impose higher capital and labor income tax rates to improve income inequality problem. In addition, the tendency of capital and labor tax rates are robust and unaffected by changing σ, implying qt is the main driving force of the shape of capital and labor tax rates in Ramsey taxation.

![Figure 16. Robust check in Ramsey tax rates](image)

Similar argument can also be applied in Mirrleesian taxation.

- When σ reduce 50%
  2. Skilled labor tax rates 29.6%
  3. Unskilled labor tax rates decrease 15.1% during 1965-2015

- When σ increase 150%
  1. Capital tax rates increase 25.1% on average during 1965-2015
  2. Skilled labor tax rates 67.9% on average during 1965-2015
3. Unskilled labor tax rates increase 14.3% on average during 1965-2015

Figure 17. Robust check in Mirrleesian tax rates

9.3 Quasilinear case

Another example to verify the robustness of the tax rates in our model is to apply quasilinear production in
Figure 18. Ramsey tax and multipliers $\Upsilon_t$ in quasi-linear case

Figure 19. Mirrleesian tax in quasi-linear case