Abstract

Since the 1970s, income inequality in the U.S. has increased sharply. During the same time span, the U.S. federal income tax has become less progressive. Why? I examine this question in a Ramsey optimal tax policy framework. Within this framework, the tax policy is determined by: (1) a set of Pareto weights representing the government’s preference over different households; and (2) household lifetime utilities summarizing the effects of economic fundamentals. I first study the changes in economic fundamentals using an overlapping generations incomplete-markets life-cycle model with heterogeneous households. The model features both endogenous human capital accumulation and household labor supply and is calibrated to the U.S. economy in the 1970s and 2010s. Then I use this economic model to determine whether the change in income tax is the result of an optimal policy response to changing economic fundamentals or the consequence of a change in Pareto weights. I interpret the latter as changes in the political influences of various income groups. I find that: (1) changes in economic fundamentals alone induce a less progressive optimal income tax and can account for 40% of the reduction in progressivity we observe; and (2) the change in Pareto weights required to explain the remaining part of tax policy change favors high-income households and also implies less valued government services. Finally, using a stylized political economy model, I discuss potential explanations for this change in Pareto weights such as the lower cost of conveying information to swing voters and the rising inequality of voter turnout among different socioeconomic groups.
1 Introduction

Since the 1970s, income inequality in the U.S. has increased sharply.\(^1\) Most of this rising income inequality is due to the more unequal labor income in the upper half of the income distribution (Piketty and Saez (2003)).\(^2\) This is true both at the household and individual level, and for both males and females. The rising income inequality has become a primary concern for people in the U.S., and a popular suggestion in terms of economic policy to counter such rising income inequality is to adopt a more progressive income tax policy to reduce the after-tax income inequality. However, the actual income tax policy in the U.S. moved in the opposite direction. Changes in income tax law since the late 1970s resulted in larger tax cuts for high-income households, and the income tax schedule today is less progressive than it used to be in the 1970s.

There are two potential explanations for this less progressive income tax despite the rising income inequality. The first one is that the policy-making system might have become more favorable to the high-income households, which could be due to changes in the political influences of various income groups. The second explanation is that economic changes since the 1970s might have increased the cost of progressive income taxation and hence require a less progressive optimal income tax to be adopted.

Several economic changes since the 1970s could potentially contribute to a less progressive optimal income tax policy. The first is skill-biased technological change, which has become the most influential explanation for the rising income inequality in the first place (Acemoglu (2002)). On the one hand, the higher income inequality caused by this technological change increased the redistribution benefits of a more progressive income tax. On the other hand, it also increased the value of human capital. If human capital has to be accumulated endogenously with nontrivial costs, the benefits of a less progressive income tax to encourage human capital investment may surpass the benefits of redistribution and therefore require a less progressive income tax to be optimal.

The declining gender income gap due to increased female labor productivity could also help explain this less progressive income tax. In the late 1970s, the female labor income share in the U.S. was only about a quarter of the total labor income, but in the early 2010s, that ratio climbed to about 40\%.\(^3\) The greater female labor income share is caused by the increase in both the female wage rate and the female labor supply relative to the male wage rate and labor supply (Blau and Kahn (2000), O’Neill (2003)).

---

\(^1\)Appendix A provides more information on the empirical facts mentioned in the introduction.
\(^2\)Hence, this paper focuses on the changing labor income structure and labor income tax policy and abstracts away from capital income taxation.
\(^3\)Based on the data from the Current Population Survey Annual Social and Economic Supplement.
ical studies have found that the female labor supply is more elastic compared with the male labor supply.\textsuperscript{4} Hence, as the female labor income takes a larger share of the total household income, the elasticity of household income with respect to marginal income tax increases. Since most married couples file their income taxes jointly, the increased elasticity of household income means the optimal income tax should be less progressive.

A third economic change that might require a less progressive income tax is the aging U.S. population. Life expectancy was 74 in the late 1970s and rose to 79 in the early 2010s. The rising life expectancy increases the age decency ratio\textsuperscript{5} and demands more revenues to be collected through income taxes on the working-age population because: first, it increases the total cost of social security benefits; and second, it increases the demand for government services, as more people are living longer. Previous studies have shown that a more progressive income tax system tends to reduce the government’s ability to extract tax revenues from the economy (Holter, Krueger, and Stepanchuk (2014), Guner, Lopez-Daneri, and Ventura (2015)), and hence, this extra demand for tax revenues could force the policymakers to compromise on redistribution and apply a less progressive income tax policy.

Motivated by the discussions above, the main purpose of this paper is to investigate whether the less progressive income tax policy since the 1970s can be rationalized as an optimal response of tax policy to the changes in economic fundamentals, or whether it is a result of the changing preferences of policymakers over different households. To accomplish this goal, I employ the Ramsey optimal tax policy framework, in which a Ramsey government chooses the tax policy to maximize a weighted sum of household welfare across different competitive equilibria. Within this framework, the tax policy is determined by: (1) a set of Pareto weights representing the government’s preference over different households; and (2) household lifetime utilities summarizing the effects of economic fundamentals. The changes in economic fundamentals since the 1970s are quantified using an economic model with household heterogeneity disciplined by the data in the 1970s and 2010s. The Pareto weights are harder to measure directly but can be inferred from the actual tax policies chosen by inverting the Ramsey problem, i.e., finding the Pareto weights that rationalize the actual tax policy as the solution to the Ramsey problem. By construction, the combination of the change in Pareto weights inferred this way and the economic changes in the model replicates exactly the change of income tax policy in the data. A counterfactual experiment in which economic changes are intro-

\textsuperscript{4}See Blundell, Pistaferri, and Saporta-Eksten (2012) for a recent empirical estimation of the male and female labor supply elasticities using PSID data.

\textsuperscript{5}The age dependency ratio is defined as the ratio of dependents, people younger than 15 or older than 64, to the working-age population, those ages 15 to 64.
duced whereas holding the Pareto weights fixed at the 1970s values identifies the part of income tax change that serves as an optimal response to the underlying economic forces, and the remaining part is then attributed to the change in Pareto weights. I interpret the latter as changes in the political influences of various income groups.

The economic model employed to capture the changes in economic fundamentals since the 1970s and their implications on household behaviors and welfare is a quantitative overlapping generations incomplete-markets life-cycle model with heterogeneous households. To model skill-biased technological change, a Ben-Porath style human capital accumulation technology is introduced as in Guvenen and Kuruscu (2010) and Guvenen, Kuruscu, and Ozkan (2014). The return to human capital investment depends on the heterogeneous learning abilities of earners and increases from the 1970s to 2010s to match the widening gap in the upper half of the labor income distribution. To account explicitly for the changing role of the female labor supply in the economy, each household in the model consists of two earners, a male and a female, and they make joint decisions on household consumption, savings, labor supply, and human capital investment. Female labor productivity increases between the 1970s and 2010s to match the declining gender income gap in the data. Finally, earners face the uninsurable idiosyncratic risk of labor productivity both at labor market entry and over the span of the household life cycle and are subject to tight borrowing constraints. The amount of idiosyncratic risk in the model is calibrated to match the dispersion of labor income among young earners and the earnings dynamics over the life cycle in the data.

The first main finding of my quantitative analysis is that the changes in economic fundamentals since the 1970s alone require a less progressive optimal income tax policy to be adopted and can quantitatively account for 40% of the reduction in progressivity we observe. The progressivity of income tax here is measured by the elasticity of after-tax income with respect to before-tax income. In the late 1970s, this elasticity was about 0.856, and it increased to 0.914 in the early 2010s. The model implies that if the Pareto weights remained the same as they were in the 1970s, the optimal income tax policy in the early 2010s should be with a elasticity of 0.879. Counterfactual experiments show that skill-biased technological change, increased female labor productivity, and the aging U.S. population all contribute to the less progressive optimal income tax and account for 18%, 44%, and 73% of the actual change in progressivity, respectively. However, the effects

---

6With a flat rate income tax, this elasticity is 1. The smaller is this elasticity, the more progressive is the income tax system.

7The contribution of skill-biased technological change may seem small compared with the other two. However, this does not mean that the endogenous human capital accumulation channel is less important. The reason is that skill-biased technological change also causes a significant rise in income inequality at
of these changes are partially offset by the increase of idiosyncratic risk, which increases the insurance benefits of progressive income taxes and accounts for $-88\%$ of the actual change in progressivity.

The second main finding is that the Pareto weights, as implied by the actual income tax policies in the 1970s and 2010s, have changed in two dimensions: (1) the Pareto weights of the high-income households have increased relative to those at the lower end of the income distribution; and (2) the Pareto weights on household private utilities have increased relative to the weight on government services. The first change is most responsible for the remaining $60\%$ of reduction in income tax progressivity, while the second change is most responsible for the significant fall in the overall level of the U.S. income tax since the 1970s.

Finally, since the model ascribes a significant part of observed changes in the income tax policy to changes in political influences, as approximated by the Pareto weights, in the last part of this paper, I provide potential political economy explanations for this phenomenon. Using a stylized probabilistic voting model with political contributions, I show that the lower cost of conveying information to swing voters due to information technology improvements leads to an increased demand for campaign expenditures, as observed in the data. I show that this induces a change in Pareto weights benefiting the high-income households, consistent with the change in income tax policy studied in the first part of this paper. I also show that the rising inequality of voter turnout among different socioeconomic groups may have contributed to such a change in Pareto weights as well.

### 1.1 Related Literature

In terms of the model, this paper is first related to the literature that has studied heterogeneous household models with idiosyncratic risks, as in Huggett (1993) and Aiyagari (1994). This type of models have been widely used in quantitative macroeconomic studies of income, wealth, and consumption inequality and redistributive policies. The model built in this paper is a natural combination of two recent developments of the heterogeneous households life-cycle models. The first development is the adoption of the two-earner household structure, which takes into account not only the role of females in the economy, but also the interaction of behaviors within households. Heathcote, Storesletten, and Violante (2010) and Guner, Kaygusuz, and Ventura (2012) are recent studies with the same time, which would have moved the tax policy in the opposite direction if the endogenous human capital accumulation channel were absent.
such feature of the model. The other development is the introduction of endogenous human capital accumulation in quantitative life-cycle models such as Huggett, Ventura, and Yaron (2011) and Guvenen, Kuruscu, and Ozkan (2014) with the Ben-Porath human capital accumulation technology.

In terms of the topic, this paper is in line with the quantitative Ramsey optimal income tax policy literature using heterogeneous agents incomplete-markets life-cycle models, which departs from the previous static optimal income tax studies such as Mirrlees (1971), Diamond (1998), and Saez (2001). Important works in this direction include Conesa and Krueger (2006), Conesa, Kitao, and Krueger (2009), and most recent Badel and Huggett (2014), Kindermann and Krueger (2014), and Guner, Lopez-Daneri, and Ventura (2015) on the income tax policy for the top 1%. Krueger and Ludwig (2013) ask the optimal labor income tax question with endogenous college education decisions, whereas my paper focuses on the human capital accumulation over the working life cycle. Heathcote, Storesletten, and Violante (2014) develop an analytic framework to link the economic fundamentals to the optimal income tax progressivity, but their focus on closed-form solutions and analytic results force their model to be more stylized in some aspects, such as the zero net wealth, one-shot investment in skills, and the absence of female earners. Kaymak and Poschke (2015) are interested in the tax policy change since 1960s in the U.S., but they focus on the economic consequences of exogenous top income tax cuts, whereas I consider the determination of the tax policy in response to economic and non-economic changes. There is also a literature asking the optimal tax policy questions using the mechanism design approach such as Farhi and Werning (2013) and most recent Stantcheva (2015) with human capital. This approach allows more flexible tax system, but the optimization problem is typically difficult to solve in dynamic or non-standard contexts. Also, whether such flexible and hence more complicated tax system is feasible in practice remains a question.

The numerical method developed in this paper to infer the Pareto weights from the actual income tax policy is related to the so-called “inverse-optimum” research. Bourguignon and Spadaro (2012) first derive the formula to reverse the static optimal income tax problem, and use it to infer the Pareto weights from actual marginal income tax rates in France. Lockwood and Weinzierl (2014) use the same method and apply it to the U.S. data. Their method is more restrictive in the sense that it requires the utility function to be quasi-linear and can only be applied to stylized static model, whereas my numerical method can be applied to more general preferences and quantitative dynamic models. Chang, Chang, and Kim (2015) conduct a cross-country study to uncover the Pareto weights with similar spirit, but they only consider flat income tax and do not have human
capital accumulation and female earners in their model.

The rest of this paper is organized as follows. Section 2 presents the quantitative life-cycle model, the Ramsey optimal tax policy problem, and the numerical method to invert the Ramsey problem. Section 3 describes the calibration strategy and reports the calibration results. Section 4 presents the quantitative analysis and results. Section 5 discusses possible explanations for the change in Pareto weights from a political economy point of view. Section 6 concludes.

2 Model

In this section, I present the quantitative life-cycle model employed to capture the changes in economic fundamentals between the 1970s and 2010s. I first describe the problems of households, the representative firm, and the government and define the stationary competitive equilibrium. Then I formalize the Ramsey optimal tax policy problem. Finally, I describe the method used to infer the Pareto weights from the actual tax policy.

2.1 Households

The economy is populated with overlapping generations households. In each year, a measure one of new households are born at age 1. Each household consists of two members: a male and a female. Both members can work in the first $T$ years of the household’s life cycle, then enter retirement for another $T_R$ years, and then die for sure.\footnote{Unfortunately, I have to abstract away from the marriage and divorce processes because modeling those requires keeping track of equilibrium distributions of single males and females, which is computationally challenging given the number of state variables in the model.} I use subscript $i$ and $j$ to denote the gender and age of an earner, with $i = 1$ and 2 corresponding to the male and the female, and omit the index for different households for simplicity.

Within each cohort, households are heterogeneous at birth in their learning abilities $\{A_i\}_{i=1}^2$. To reduce the number of state variables, I assume that the male and female learning abilities $\{A_i\}_{i=1}^2$ are determined by a common household level ability variable $A$, and $A_i = f_i(A), f_i'(\cdot) > 0, i = 1, 2$. Hence the high ability males are matched with the high ability females. The ability of an earner is constant over the life cycle, and the distribution of household level learning ability $A$, governed by the cdf $F(A)$, is assumed to be the same across cohorts. The initial productivity of an earner $w_{i,1}$ is positively correlated with his or her ability $A_i$ but not perfectly.\footnote{Huggett, Ventura, and Yaron (2011) estimate that the correlation between the log ability and the log initial productivity is about 0.8.}
A Ben-Porath style human capital investment technology is available to all the earners, which allows an earner to increase his or her productivity $w_{i,j+1}$ by spending some time $n_{i,j}$ studying at age $j$. In addition to the study time, the outcome of such investment also depends on the learning ability of the earner $A_j$, the productivity at current age $w_{i,j}$, the rate of return to human capital investment $R^i H$, and the realization of an idiosyncratic human capital shock $z_{i,j+1}$:

$$w_{i,j+1} = e^{z_{i,j+1}} [w_{i,j} + R^i H A_j (w_{i,j} n_{i,j})^a].$$ \hspace{1cm} (2.1)

The main difference from the standard Ben-Porath formula is the additional parameter $R^i H$, which is used to capture the rising return to human capital investment between the 1970s and 2010s as skill-biased technological change.\(^\text{10}\) The idiosyncratic human capital shocks $\{z_{i,j+1}\}_{i=1}^2$ are allowed to be correlated between the two earners within each household, but are i.i.d. over time and across households. The means of these shocks are slightly negative and represent the depreciation of human capital over time.

The working-age households can earn labor income from the labor supply of the male and the female. Let $\bar{w}_{i,j}$ and $l_{i,j}$ denote the wage rate and the time worked of the gender $i$ earner at age $j$, then the before-tax labor income of this earner is $y_{i,j} = \bar{w}_{i,j} l_{i,j}$. The wage rate $\bar{w}_{i,j}$ is determined by three components: $\bar{w}_{i,j} = w_e \theta_i w_{i,j}$, where $w_{i,j}$ is the productivity of the earner, $\theta_i$ is a gender factor of labor productivity, and $w_e$ is the wage rate of effective labor. The gender factor for males $\theta_1$ is normalized to be 1, while $\theta_2$ for females is used to capture the increase of female labor productivity between the 1970s and 2010s. The wage rate of effective labor $w_e$ can also be different between the two time periods to reflect the change of the overall labor productivity.\(^\text{11}\)

Besides labor income, households can earn capital income by saving in a risk-free bond with interest rate $r$, but cannot borrow into negative asset positions. The retired households also receive social security benefits $b$ from the government in each retirement year. There are no insurance markets for the idiosyncratic human capital shocks, so the financial markets are incomplete.

Only labor income is taxed by the government, and the tax policy is summarized by

\(^\text{10}\)The law of motion for $w_{i,j}$ here follows the human capital investment technology in Guvenen and Kuruscu (2010) and Guvenen, Kuruscu, and Ozkan (2014). They assume further that the labor productivity $w_{i,j}$ is determined by two components: raw labor and human capital. In that case, $R^i H$ is the price of human capital, and an increase of $R^i H$ corresponds to skill-biased technological change. More details are available in those two papers.

\(^\text{11}\)The introduction of $w_e$ and $\theta_2$ allows the model to match the levels and the shapes of male and female labor income life cycles at the same time. The same effects can be obtained by allowing the distribution of learning ability to be different over time, which is a less appealing assumption to make.
the function $T(\cdot)$ which gives the tax liability based on the household’s total before-tax labor income. Hence, the after-tax labor income of a household with before-tax labor income $\{y_{ij}\}_{i=1}^2$ is $\sum_{i=1}^2 y_{ij} - T(\sum_{i=1}^2 y_{ij})$.

The state variables of a working-age household are the savings $a$, the male and female labor productivities $\{w_i\}_{i=1}^2$, the age of the household $j$, and the household ability level $A$. In each year, the two earners in each household make joint decisions on current household consumption, savings, labor supply, and study time. They enjoy consumption, but dislike non-leisure time of work and study. Hence, a working-age household’s problem is in the recursive form:

$$V(a, w_1, w_2, j, A) = \max_{\{c, a', l_1, l_2, n_1, n_2\}} u(c, l_1, l_2, n_1, n_2) + \beta \sum_{z'} \pi(z') V(a', w'_1, w'_2, j + 1, A)$$

s.t.

$$c + a' = \left[\sum_{i=1}^2 y_i - T(\sum_{i=1}^2 y_i)\right] + (1 + r)a$$

$$w'_i = e^{z'_i} [w_i + R_i A_i (w_i n_i)^a], i = 1, 2$$

$$z' = (z'_1, z'_2)^T \sim i.i.d. N(\mu_z, \Sigma_z)$$

$$A_i = f_i(A), i = 1, 2$$

$$y_i = w_i \theta_i w_i l_i, i = 1, 2$$

$$a' \geq 0, l_i \geq 0, c \geq 0, n_i \geq 0.$$

where $\beta$ is the discount factor of future utility, and $\pi(\cdot)$ is the joint pdf of the idiosyncratic human capital shocks.

Members of a retired household no longer work or study and suffer no risks. Hence the state variables of a retired household are only the savings $a$ and the age of the household $j$, and its problem in the recursive form is:

$$V^R(a, j) = \max_{\{c, a'\}} u^R(c) + \beta V^R(a, j + 1)$$

s.t.

$$c + a' = b + (1 + r)a$$

$$a(T + T_R) \geq 0, c \geq 0.$$

The instantaneous utility function of working-age households is assumed to be addi-
tive separable between consumption and non-leisure time and takes the following functional form:

\[
u(c, l_1, l_2, n_1, n_2) = \log(c) - \psi_1 \left( \frac{l_1 + n_1}{1 + \frac{1}{\eta_1}} \right) - \psi_2 \left( \frac{l_2 + n_2}{1 + \frac{1}{\eta_2}} \right)\]

where \( \eta_i \) is the earner \( i \)'s Frisch elasticity of labor supply, and \( \psi_i \) captures the level of disutility from earner \( i \)'s non-leisure time. The instantaneous utility function of retired households is simply:

\[
u^R(c) = \log(c).
\]

### 2.2 Representative Firm

The production side of the economy consists of measure one profit-maximizing perfect competitive firms. They rent physical capital at interest rate \( r \) and hire effective labor at wage rate \( w_e \) to produce the final good used for both consumption and investment in physical capital. All the firms have the same production technology, which takes the standard Cobb-Douglas functional form:

\[Y = K^\omega (Z\bar{L})^{1-\omega}\]

where \( Y \) is the final output, \( K \) is the physical capital rented, \( \bar{L} \) is the effective labor hired, \( Z \) is the productivity of effective labor, and \( \omega \) is the parameter governing the capital income share. Because this production function has constant return to scale, all the firms make zero profits at the competitive equilibrium, and the entire production side is equivalent to one representative firm who takes the input and output prices as given and maximizes its profits period by period. The representative firm’s problem is then

\[
\max_{\{K, L\}} K^\omega (Z\bar{L})^{1-\omega} - (1 + r)K - w_e \bar{L} + (1 - \delta)K
\]

where \( \delta \) is the depreciation rate of physical capital. The optimality conditions of the representative firm are then:

\[
r = \omega \left( \frac{K}{Z\bar{L}} \right)^{\omega-1} - \delta,
\]

\[
w_e = (1 - \omega)Z \left( \frac{K}{Z\bar{L}} \right)^\omega.
\]
2.3 Government

The government only levies labor income tax at household level, and the tax liability only depends on the total before-tax labor income of the household, which is given by the function $T(\cdot)$. There are two uses of the tax revenues by the government: (1) paying the total social security benefits to the retired households, $T_Rb$; (2) financing government services, $G$. Hence the labor income tax in the model corresponds to the combination of labor income tax and social security tax in the U.S.

Government can only choose the level of $G$ to balance its budget period by period for a given tax function $T(\cdot)$ and the level of retirement benefits $b$:

$$\sum_{j=1}^{T} \int T(\sum_{i=1}^{2} y_{i,j}(s))d\Phi_{j}(s) = T_Rb + G$$

where $s$ is the vector of household state variables except for age, i.e., $s = (a, \{w_i\}_{i=1}^{2}, A)$ when $j = 1, \ldots, T$, and $s = a$ when $j = T + 1, \ldots, T + T_R$; $y_{i,j}(s)$ is the earner $i$’s labor income in an age $j$ household with state $s$, and $\Phi_{j}(s)$ gives the measure of households with age $j$ and state $s$.

Following Bnabou (2002) and Heathcote, Storesletten, and Violante (2014), the labor income tax function $T(\cdot)$ is assumed to take the form of

$$T(y) = y - (1 - \tau)y^{1-\mu}$$

(2.2)

where $\tau$ and $\mu$ are parameters governing the level and progressivity of the income tax. With this tax function, $1 - \mu$ is the elasticity of after-tax income with respect to before-tax income. If $\mu = 0$, this elasticity is one, and the income tax rate is flat. The larger is the value of $\mu$, the smaller is this elasticity, which means that the income tax is more progressive. The parameter $\tau$ on the other hand affects only the level of income tax and has no impacts on this elasticity.

2.4 Stationary Competitive Equilibrium

The economy is assumed to be open with free capital movement across the border, hence the domestic interest rate $r$ is fixed at the global level $r^*$.\footnote{While the U.S. economy is not the typical small open economy people often have in their minds, it is not a closed economy, either. Studies such as Warnock and Warnock (2009) have shown that the interest rates in the U.S. are significantly affected by the flows of foreign capital, and a large portion of the U.S. government debts are held by foreigners. The key implication of this assumption is that the domestic interest rate will...}
paper, I focus on the stationary competitive equilibrium of the economy, which is defined in the following.

**Definition of Stationary Competitive Equilibrium:** A stationary competitive equilibrium is a collection of household value and policy functions \( \{V, V^R, c, a', (l_i, n_i)_{i=1}^2\} \), the representative firm’s decisions \( \{K, \bar{L}\} \), government expenditure \( G \), the wage of effective labor \( w_e \), the domestic interest rate \( r \), and a sequence of distributions of household state \( \{\Phi_j(s)\}_{j=1}^{T+T_R} \) such that

1. **Households:** given the prices \( \{w_e, r\} \), the tax function \( T(\cdot) \), and the social security benefits \( b \), the collection of household value and policy functions \( \{V, V^R, c, a', (l_i, n_i)_{i=1}^2\} \) solves the household’s problem.

2. **Representative firm:** given the prices \( \{w_e, r\} \), the values of \( \{K, \bar{L}\} \) satisfy the representative firm’s optimality conditions.

3. **Government:** given the tax function \( T(\cdot) \), social security benefits \( b \), and the household policy functions, the value of \( G \) satisfies the government budget constraint.

4. The labor market clears:

\[
\bar{L} = \sum_{j=1}^{T} \int \sum_{i=1}^{2} \theta_i w_{ij}(s) l_{ij}(s) d\Phi_j(s).
\]

5. The physical capital market clears:

\[
r = r^*.
\]

6. Stationary conditions: given \( \Phi_1(\cdot) \), the law of motion of \( \{\Phi_j(\cdot)\}_{j=1}^{T+T_R} \) induced by the household policy functions, demographics, and idiosyncratic shocks, \( \{H_j(\cdot)\}_{j=1}^{T+T_R-1} \), satisfies

\[
\Phi_{j+1} = H_j(\Phi_j), j = 1, \ldots, T + T_R - 1.
\]

not respond to the changes of income tax policy, which greatly reduces the computation burden.
2.5 Ramsey Optimal Tax Policy Problem

A Ramsey government chooses the labor income tax function \( T(\cdot) \) to solve the following optimization problem at the stationary competitive equilibrium.

\[
\max_{\{T(\cdot), G\}} \int \left[ \int V(a_0, w_{1,1}, w_{2,1}, 1, A) d\Pi(a_0, w_{1,1}, w_{2,1} | A) \right] W(A) dF(A) \\
+ \gamma \left( \sum_{j=1}^{T+T_R} \beta^{j-1} \right) \log \left( \frac{G}{T + T_R} \right)
\]

s.t.

\[
\sum_{j=1}^{T} \int T \left( \sum_{i=1}^{2} y_{ij}(s) \right) d\Phi_j(s) = T_R b + G
\]

The first part of the Ramsey government’s objective function is a weighted sum of expected lifetime utility of a newborn cohort at the stationary competitive equilibrium. In particular, \( V(a_0, w_{1,1}, w_{2,1}, 1, A) \) is the expected lifetime utility of a newborn ability \( A \) household with the initial state \((a_0, w_{1,1}, w_{2,1})\); \( \Pi(a_0, w_{1,1}, w_{2,1} | A) \) is the conditional cdf of initial household state \((a_0, w_{1,1}, w_{2,1})\) given the household ability \( A \); \( W(A) \) is the Pareto weight assigned to the ability \( A \) households;\(^{13}\) and \( F(A) \) is the unconditional cdf of the household ability \( A \). If \( W(A) \) equals to one for all values of \( A \), it becomes utilitarian weight function, and the first part is simply the expected lifetime utility of a newborn household at the stationary competitive equilibrium before any uncertainty is resolved.

The second part of the Ramsey government’s objective function is the lifetime utility of a newborn cohort generated by government services at the stationary competitive equilibrium. \( G \) is the total government services provided in each year, and \( \frac{G}{T + T_R} \) is the government services per household.\(^{14}\) The flow utility of each household generated by government services is assumed to be the log of government services per household,\(^{15}\) and it is discounted by the same discount factor \( \beta \) as the household private flow utility.

---

\(^{13}\)In general, the Pareto weight function \( W(\cdot) \) can be a function of all the household state variables. However, for tractability, I restrict it to be a function of only the household learning ability \( A \). This assumption reduces the identification burden significantly when inferring the Pareto weight function from the actual tax policy. Since \( A \) is directly related to the expected lifetime income of households at equilibrium, this weight function allows us to capture the government’s preference over households with different lifetime income.

\(^{14}\)Recall that there are measure one households in each cohort, so the total measure of households at any given year is \( T + T_R \).

\(^{15}\)The choice of the log function is such that when the model economy is scaled up, the government expenditure share would remain stable relative to the size of the economy.
The parameter $\gamma$ is the Pareto weight on government services. Obviously, the scale of the weight function $W(A)$ and $\gamma$ does not matter, so we can normalize $\gamma = 1$ for identification.

### 2.6 Inverting the Ramsey Problem

For the purpose of this paper, we need to find the Pareto weight function which can rationalize the actual income tax policy we observe, i.e, the Ramsey government would choose the actual income tax policy with such Pareto weight function. One way of doing this is to use the first order conditions of the Ramsey problem.

Suppose the observed income tax function $T(\cdot)$ is parameterized with $M$ parameters $\{\tau_m\}_{m=1}^M$, then the income tax policy is represented by the values of $\{\tau_m\}_{m=1}^M$. We can then parameterize the weight function as $W(A) = \sum_{p=1}^M \xi_p F(A)^{p-1}$, where $F(A)$ is the cdf of household ability $A$, and $\{\xi_p\}_{p=1}^M$ are coefficients to be determined. The first order conditions of the Ramsey problem are then

$$
\int \left[ \int \frac{\partial V(a_0, w_{1,1}, w_{2,1}, A)}{\partial \tau_m} d\Pi(a_0, w_{1,1}, w_{2,1} | A) \right] \left[ \sum_{p=1}^M \xi_p F(A)^{p-1} \right] dF(A) = - \left( \sum_{j=1}^{T+T_R} \beta^{j-1} \right) \frac{1}{G} \frac{\partial G}{\partial \tau_m}, \quad m = 1, ..., M.
$$

where the amount of government services $G$ is treated as an implicit function of the income tax policy $\{\tau_m\}_{m=1}^M$, which is defined by the government budget constraint. To simplify the notations, let $B_{m,p} = \int \int \frac{\partial V(a_0, w_{1,1}, w_{2,1}, A)}{\partial \tau_m} d\Pi(a_0, w_{1,1}, w_{2,1} | A) F(A)^{p-1} dF(A)$, then the above system of equations becomes:

$$
\begin{bmatrix}
B_{1,1} & \cdots & B_{1,M} \\
\vdots & \ddots & \vdots \\
B_{M,1} & \cdots & B_{M,M}
\end{bmatrix}
\begin{bmatrix}
\xi_1 \\
\vdots \\
\xi_M
\end{bmatrix} = - \left( \sum_{j=1}^{T+T_R} \beta^{j-1} \right) \frac{1}{G}
\begin{bmatrix}
\frac{\partial G}{\partial \tau_1} \\
\vdots \\
\frac{\partial G}{\partial \tau_M}
\end{bmatrix}
$$

All the $B_{m,p}$ and $\frac{\partial G}{\partial \tau_m}$ can be computed numerically using the economic model at the observed values of $\{\tau_m\}_{m=1}^M$, so we can simply solve this linear system of equations for the coefficients $\{\xi_p\}_{p=1}^M$ and hence recover the Pareto weight function.

---

16 In the exercises of this paper, the income tax function only has two parameters, $\tau$ and $\mu$, as in Equation (2.2). I describe the method here in the most general form to show that it can be applied to more flexible income tax functions, and hence can allow more flexible Pareto weight functions.

17 Other basis functions can also be used to parameterize the weight function. The advantage of using the cdf $F(A)$ here is such that $W(A)$ is bounded even if the distribution of $A$ is unbounded.
Note that the number of coefficients in the Pareto weight function is intentionally chosen to be exactly the same as the number of parameters in the tax function to guarantee the existence and uniqueness of the inferred Pareto weight function. We also need the objective function to be concave with respect to the choice variables \( \{\tau_m\}_{m=1}^M \) for the sufficiency of the first order conditions and the \( (B_{m,p}) \) matrix to be invertible, both of which can be verified numerically.

3 Calibration

In this section, I describe the calibration strategy for the economic model and report the calibrated parameter values and the empirical targets matched by the model. To choose the values of parameters, an economy at the stationary competitive equilibrium with 20000 households in each cohort is simulated from the model and first calibrated to match the U.S. economy in the years 2010-2012 for the early 2010s. For the late 1970s, part of the parameters are kept the same as in the years 2010-2012 such as those governing the household preference and the distribution of learning ability, whereas the others are recalibrated to match the 1978-1980 empirical targets such as the return to human capital investment and the female factor of labor productivity, etc. The data used to compute the empirical targets are from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) and the core sample of the Panel Study of Income Dynamics (PSID). All the nominal variables are deflated using the Bureau of Labor Statistics’ Consumer Price Index Research Series (CPI-U-RS).

3.1 Calibration Strategy

In the model, one unit of the final good represents $58026 in 2012 dollars, which is the mean labor income of a working male with age 23 to 65 in the years 2010-2012. The empirical life cycle profiles used are cross-sectional life cycles, i.e, they are computed from the CPS cross-sectional data for households at different ages. Since the upper half of the income distribution is more important for the income tax policy questions, the calibration of the model focuses on matching the life cycle profiles of the 50th, 90th and 99th percentiles of the male and female labor income in addition to a selected group of empirical moments. Many parameters are jointly selected to match these empirical targets at the same time, hence there are no exact one-to-one mappings between the parameters and the empirical targets. However, we can still link different parameters to the empirical targets which offer most of the identification powers for them.
Demographics. The starting age of households is set to be 23 and the retirement age is 65. Hence $T$ is 43 in the model. For the years 2010-2012, the length of retirement, $T_R$, is set to be 14 years based on the life expectancy of 79 in the U.S.

Preference and Interest Rate. The Frisch elasticities of the male and female labor supply $\eta_1$ and $\eta_2$ are set at 0.4 and 0.8, respectively. These values are consistent with recent studies on family labor supply such as Blundell, Pistaferri, and Saporta-Eksten (2012). The parameter capturing the disutility of the male non-leisure time $\psi_1$ is normalized to be 1 because the unit of time can be freely adjusted in the model. The female counterpart $\psi_2$ is mainly identified by the female-male labor income ratio and hours worked ratio. The global risk-free real interest rate $r^*$ is set at 1%. The discount factor of flow utility $\beta$ is set at 0.99.

Production Technology. Because the economy is open with free capital movement across the border, the domestic interest rate $r$ is fixed at the global level $r^*$ across equilibria. Also, the ratio $\frac{K}{ZL}$ is pinned down by $r^*$, $\delta$, and $\omega$ from the representative firm’s optimality conditions, and it is not affected by changes of income tax policy. Therefore, $w_e$ is also a constant across equilibria. So instead of calibrating the values of $Z$, $\omega$, and $\delta$, we only need the values of $r$ and $w_e$ to solve the model. The value of $w_e$ is identified by the mean labor income of working males.

Gender Factors of Labor Productivity. For males, the gender factor of productivity $\theta_1$ is normalized to be 1. For females, $\theta_2$ is identified from the female-male labor income and hours worked ratios together with $\psi_2$.

Human Capital Accumulation Technology. The parameter $\alpha$ in the Ben-Porath human capital accumulation formula governs the curvature of the return to human capital investment. Its value is set at 0.7 such that the life cycle of the median male labor income reaches its peak around age 55. The level of the return to human capital investment $R^1_H$ is not separately identified from the scale of the learning ability, and hence it can be normalized to be any positive number for the years 2010-2012. I choose $R^1_H = R^2_H = 0.05$ such that the scales of other variables are convenient.

Distribution of Learning Ability. The distribution of the household learning ability $A$ is assumed to be a shifted Pareto-log-normal distribution, i.e., $A \sim PLN(\mu_A, \sigma_A^2, \lambda_A) + e^{\mu_A Const_A}$. Within each household, the ability of the male is determined by $A_1 = f_1(A) = A$. Since the scales of $A_1$, $w_{1,1}$, $w_e$, and $\theta_2$ are not separately identified, the median of $A_1$ is normalized to be 1 by choosing a proper value of $\mu_A$. The standard de-

\[ 18\text{If } x_1 \sim LN(\mu, \sigma^2) \text{ and } x_2 \sim Pareto(\lambda), x_3 = x_1x_2 \sim PLN(\mu, \sigma^2, \lambda). \]

\[ 19\text{If we multiply } A_1 \text{ by } x^{1-a}, w_{1,1} \text{ by } x, w_e \text{ by } 1/x, \theta_2 \text{ by } x, \text{ and keep all the other parameters the same, the economy will be exactly the same as before.} \]
violation of the log-normal part $\sigma_A$ and the parameter $\text{Const}_A$ are chosen to match the life cycle profiles of the 90th and 99th percentiles of the male labor income in the years 2010-2012. The Pareto parameter $\lambda_A$ is used to target the Pareto ratio of male labor income within the top 1%.

$$\log(A_2) = \{1[A \geq \text{median}(A)]\beta_1^{A_2} + 1[A < \text{median}(A)]\} \log\left(\frac{A}{\text{median}(A)}\right) + \beta_0^{A_2}.$$ 

The scales of $A_2$, $w_{i,1}$, and $\theta_2$ are not separately identified, hence the median of $A_2$ is also normalized to be 1 by setting $\beta_0^{A_2} = 0$. The value of $\beta_1^{A_2}$ is chosen to match the life cycles of the 50th, 90th and 99th percentiles of the female labor income in the years 2010-2012.

**Initial Asset and Labor Productivity.** The initial asset at birth $a_0$ is assumed to be zero for all households. The initial labor productivity of a gender $i$ earner, $w_{i,1}$, is positively correlated with his or her learning ability, and it is determined by

$$\log(w_{i,1}) = \beta_{w_1}^{i} \log(A_i) + \beta_{w_0}^{i} + \epsilon_i$$

where $\epsilon_i \sim N(0, \sigma_{\epsilon_i}^2)$ is the idiosyncratic shock to the gender $i$ earner’s initial productivity. The values of $\beta_{w_1}^{i}$, $\beta_{w_0}^{i}$ and $\sigma_{\epsilon_i}$ are calibrated to best match the life cycle profiles of the 50th, 90th and 99th percentiles of the male and the female labor income and the correlation between log initial productivity and log learning ability estimated by Huggett, Ventura, and Yaron (2011), which is about 0.8.

**Human Capital Shocks, Transitory Shocks and Measurement Errors.** The idiosyncratic shocks associated with the human capital accumulation $z_j = (z_{1,j}, z_{2,j})^T$ are assumed to be i.i.d. over time and across households, but they are joint-normal distributed within each household between the male and the female.

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{z_1} \\ \mu_{z_2} \end{bmatrix}, \begin{bmatrix} \sigma_{z_1}^2 & \rho_{z_1z_2}\sigma_{z_1}\sigma_{z_2} \\ \rho_{z_1z_2}\sigma_{z_1}\sigma_{z_2} & \sigma_{z_2}^2 \end{bmatrix}\right)$$

The mean of the joint-normal distribution of human capital shocks $(\mu_{z_1}, \mu_{z_2})^T$ is chosen to match the decline of the life cycle income profiles near the retirement. The human capital shocks are permanent income shocks given their ways of entering the human capital accumulation formula. To match the earnings dynamics in the data, I multiply the labor

---

20The Pareto ratio at a cutoff income level $y_{\text{cutoff}}$ is defined as $\frac{E(y|y \geq y_{\text{cutoff}})}{y_{\text{cutoff}}}$. This is a measure of the shape of the income distribution at the upper tail.
income by another i.i.d. transitory income shocks, $\varepsilon_{ts,i} \sim LN(0, \sigma_{ts,i}^2)$, when calibrating the model. These transitory income shocks are assumed to be fully insurable through risk sharing among households, and therefore they have no effects on household behaviors.\(^{21}\) Also, the micro level income survey data usually bear nontrivial measurement errors. Hence, whenever compared with the empirical targets from CPS or PSID, the income data simulated from the model are multiplied by an i.i.d. measurement error component $\varepsilon_{me} \sim LN(0, \sigma_{me}^2)$. The standard deviation $\sigma_{me}$ is set at 0.15.\(^{22}\)

The covariance matrix of the human capital shocks and the variances of the transitory income shocks are calibrated jointly to match the earnings dynamics in the PSID data. The earnings dynamics are captured by the variances and first-order autocovariances of the male and female residual labor income growth, and the correlation between the male and female residual labor income growth within households. To compute these empirical moments from the PSID data, I first regress the log labor income on a group of age and year dummies to estimate the “life cycle” components of labor income and the time effects for each gender. The residual log labor income $\hat{y}_{i,j}$ is computed by subtracting the “life cycle” components and the time effects from the actual log labor income, $\log(y_{i,j})$. Because the PSID data are biennial since the year 1997, the income growth is calculated over a two-year span, i.e., $\Delta\hat{y}_{i,j} = \hat{y}_{i,j} - \hat{y}_{i,j-2}$, and the empirical targets are $\text{var}(\Delta\hat{y}_{i,j})$, $\text{cov}(\Delta\hat{y}_{i,j}, \Delta\hat{y}_{i,j-2})$, and $\text{corr}(\Delta\hat{y}_{1,j}, \Delta\hat{y}_{2,j})$.\(^{23}\)

**Taxable Income.** Because a half of the Federal Insurance Contributions Act (FICA) tax for Social Security and Medicare is paid by the employers and is not counted as a part of the taxable income of employees, the before-tax income in the data is different from the total income of employees. To account for this difference, all the income data simulated from the model are transformed into comparable before-tax income based on the FICA tax rate schedule before any statistics are calculated.

**Income Tax Function.** The labor income tax function takes the form of

$$T(y) = y - (1 - \tau)y^{1-\mu}$$

where $\tau$ and $\mu$ are measures of the level and progressivity of the income tax schedule. To

\(^{21}\)The main benefit of this assumption is to avoid adding two additional state variables to the household’s problem. It should not affect the results much as previous empirical estimates and model simulation results all suggest that households can attain almost perfect insurance (> 90%) against transitory shocks (Blundell, Pistaferri, and Preston (2008), Kaplan and Violante (2010)).

\(^{22}\)The typical value of $\sigma_{me}$ assumed in the literature ranges from 0.15 to 0.20.

\(^{23}\)For the years 2010-2012 and 1978-1980, I use the PSID data from 1998-2012 and 1971-1980, respectively. The sample is restricted to the married male and female with the head age between 30 and 55. The observations with a residual income growth larger than 400% or smaller than −80% are excluded. The same process for the PSID data is also applied to the model-simulated data.
pin down the values of \( \tau \) and \( \mu \) for the U.S., I use the NBER’s TAXSIM program to create a mapping between a household’s total income, \( y \), and its total liability of the federal income tax and FICA tax, \( T(y) \), based on the actual U.S. tax policy. The employer’s share of the FICA tax is included in the total income and the total tax liability. Then \( \tau \) and \( \mu \) can be derived from the coefficients of the following equation, which can be estimated using the OLS method:

\[
\log(y - T(y)) = \log(1 - \tau) + (1 - \mu) \log(y).
\]

**Social Security Benefits.** The social security benefits \( b \) in the model are chosen to be the sum of the average male and female social security benefits in the U.S.

**The Years 1978-1980.** When calibrating the model to match the U.S. economy in the years 1978-1980, four sets of parameters are recalibrated to reflect the changes in economic fundamentals: (1) the length of retirement \( T_R \) to reflect the change of life expectancy; (2) the return to human capital investment \( R^H_i \), the female factor of labor productivity \( \theta_2 \), and the wage of effective labor \( w_e \) to reflect the changes of the production and human capital accumulation technologies; (3) the parameters governing the dispersion of initial productivity, \( \beta_1^{\omega_i} \) and \( \sigma^2_{\epsilon_i} \), the covariance matrix of human capital shocks \( \Sigma_z \), and the variances of transitory income shocks \( \sigma^2_{\epsilon_{i,s}} \) to capture the changes of idiosyncratic risk; (4) the income tax function parameters \( \tau \) and \( \mu \), the FICA tax rates, and the social security benefits \( b \) to account for the changes of the income tax and social security policies. Other parameters are kept the same as those for the years 2010-2012. Because less parameters need to be calibrated, some empirical targets are not used in the calibration for the years 1978-1980, for example, the female-male hours worked ratio and the Pareto ratio at the top 1% of the male labor income distribution.

### 3.2 Calibration Results

**3.2.1 Parameter Values**

Table 1 provides the calibrated parameter values for the years 2010-2012.\(^{25}\) The female learning ability is less dispersed than the male in the upper half of the distribution as the value of \( \beta_1^{\omega_i} \) is less than one. Compared with the male earners, the female earners face smaller initial labor productivity risk but larger human capital risk over the life cycle.

\(^{24}\)\( R^H_i \) can now be separated from \( A_i \) because the learning ability is assumed to be the same between the years 1978-1980 and 2010-2012.

\(^{25}\)Although the female factor of labor productivity \( \theta_2 \) is greater than one here, it does not mean females are more productive than males because the total productivity is determined by \( \theta_i w_{i,j} \). The value of \( \theta_2 \) could be different depending on the normalization choice on the level of learning ability.
The human capital shocks between the two earners of each household are positively correlated with a correlation coefficient of 0.335. The depreciation rate of human capital is about 2% per year. The magnitudes of transitory income risk are similar for both genders. The income tax policy in the years 2010-2012 implies that the elasticity of after-tax income with respect to before-tax income is about 0.914 as \( \eta \) equals 0.86. The social security benefits for the years 2010-2012 are 0.416 per retired household per year, which is about $24139 in 2012 dollars.

The changes of parameter values between the years 1978-1980 and 2010-2012 are reported in Table 2. From the 1970s to 2010s, the length of retirement \( T_R \) increases from 9 to 14 as a result of the extended life expectancy in the U.S. from 74 to 79. The rise of return to human capital investment \( R^H \) and the increase of female factor of labor productivity \( \theta_2 \) reflect skill-biased technological change and increased female labor productivity. Combined with the declining wage of effective labor \( w_e \) between the 1970s and 2010s, these changes are more favorable to the high ability, the experienced, and the female earners...
relative to the low ability, the young, and the male earners. In the 1970s, the dispersion of initial productivity is much smaller. For the human capital shocks, the variance increases for males but decreases for females between the 1970s and 2010s. The correlation between these shocks within households has become more positive, which increases the risk at household level. If the variance of measurement errors is the same in the 1970s and 2010s, the data imply that the variances of transitory shocks have decreased. The income tax policy in the 1970s is more progressive because the elasticity of after-tax income with respect to before-tax income is 0.856 compared with the value of 0.914 in the 2010s. The income tax level is also about 3.8% higher in the 1970s as measured by the value of $\tau$.

### 3.2.2 Empirical Targets Matched

Table 3 reports the empirical moments targeted in calibration, which can be matched almost perfectly by the model. The empirical moments about (residual) income growth are directly linked to the amount of idiosyncratic risk over the life cycle, which is important for the study of optimal income tax policy. Previous literature in labor economics has developed methods to estimate the amount of idiosyncratic risk in the data using only statistic models of income process with both a permanent component and a transitory component such as Moffitt and Gottschalk (1995), Blundell, Pistaferri, and Preston (2008). Following that approach, I also estimate the amounts of idiosyncratic risk using both the PSID data and the model-simulated data and show that they are consistent with each

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>$j = 1, \ldots, T, \ldots, (1 + t_T)$</td>
<td>$t_T$</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Wage and Gender Factor</td>
<td>$y_{it} = w_t\theta_i w_{it \gamma}$</td>
<td>$(\omega, \theta_2)$</td>
<td>(2.27, 0.77)</td>
<td>(1.93, 1.07)</td>
</tr>
<tr>
<td>Human Capital Technology</td>
<td>$w'<em>i = e^{z_i} [w_i + R_i A_i (w</em>{it \gamma})^{2}]$</td>
<td>$(R_H^1, R_H^2)$</td>
<td>(0.035, 0.020)</td>
<td>(0.05, 0.05)</td>
</tr>
<tr>
<td>Initial Labor Productivity</td>
<td>$\log(w_{it}) = \beta_{1i} \log(A_i) + \beta_{0i} + \epsilon_i$</td>
<td>$(\beta_{1i}, \sigma_1)$</td>
<td>(1.25, 0.15)</td>
<td>(2.0, 0.24)</td>
</tr>
<tr>
<td>Human Capital Shocks</td>
<td>$\left[ z_1 \right] \sim N\left(\left[\mu_{z1}, \mu_{z2}\right], \left[\begin{array}{cc} \sigma_{z1}^2 &amp; \rho_{z1z2} \sigma_{z1} \sigma_{z2} \ \rho_{z1z2} \sigma_{z1} \sigma_{z2} &amp; \sigma_{z2}^2 \end{array}\right]\right)$</td>
<td>$(\mu_{z1}, \mu_{z2})$</td>
<td>(0.104, 0.153, 0.225)</td>
<td>(0.109, 0.133, 0.335)</td>
</tr>
<tr>
<td>Transitory Shocks</td>
<td>$\epsilon_{it} \sim LN(0, \sigma_{\epsilon_{it}}^2)$</td>
<td>$(\sigma_{\epsilon_{i1}}, \sigma_{\epsilon_{i2}})$</td>
<td>(0.130, 0.125)</td>
<td>(0.110, 0.115)</td>
</tr>
<tr>
<td>Income Tax Function</td>
<td>$T(y) = y - (1 - \tau) y^{1-\tau}$</td>
<td>$(\tau, \mu)$</td>
<td>(0.258, 0.144)</td>
<td>(0.220, 0.086)</td>
</tr>
<tr>
<td>Social Security Benefits</td>
<td>$b$</td>
<td>$b$</td>
<td>0.274</td>
<td>0.416</td>
</tr>
</tbody>
</table>
Table 3: Empirical Moments Matched

<table>
<thead>
<tr>
<th>Empirical Target</th>
<th>2010-2012 Data</th>
<th>Model</th>
<th>1978-1980 Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Male Income</td>
<td>1</td>
<td>1.007</td>
<td>0.865</td>
<td>0.867</td>
</tr>
<tr>
<td>Female-Male Income Ratio</td>
<td>0.647</td>
<td>0.639</td>
<td>0.377</td>
<td>0.379</td>
</tr>
<tr>
<td>Female-Male Hours Worked Ratio</td>
<td>0.830</td>
<td>0.833</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pareto Ratio of Male Income at Top 1%</td>
<td>1.839</td>
<td>1.882</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Variance of Male Income Growth</td>
<td>0.129</td>
<td>0.133</td>
<td>0.123</td>
<td>0.126</td>
</tr>
<tr>
<td>Variance of Female Income Growth</td>
<td>0.153</td>
<td>0.152</td>
<td>0.205</td>
<td>0.203</td>
</tr>
<tr>
<td>Autocovariance of Male Income Growth</td>
<td>-0.036</td>
<td>-0.037</td>
<td>-0.040</td>
<td>-0.041</td>
</tr>
<tr>
<td>Autocovariance of Female Income Growth</td>
<td>-0.034</td>
<td>-0.035</td>
<td>-0.037</td>
<td>-0.038</td>
</tr>
<tr>
<td>Correlation of Male and Female Income Growth</td>
<td>0.033</td>
<td>0.035</td>
<td>-0.026</td>
<td>-0.024</td>
</tr>
</tbody>
</table>

Note: Only labor income is included. “Female-Male Hours Worked Ratio” and “Pareto Ratio of Male Income at Top 1%” are not used in the calibration for the years 1978-1980. The moments about income growth are calculated using the residual income as defined in Section 3.1.

other. Suppose the income process is determined by

\[
\log(y_{i,j,t}) = F(i, j, t) + P_{i,j} + u_{i,j}
\]

\[
P_{i,j+1} = P_{i,j} + v_{i,j+1}
\]

where \(F(i, j, t)\) is the income trend determined by gender \(i\), age \(j\), and year \(t\); \(P_{i,j}\) is the permanent component of residual labor income; \(v_{i,j}\) is the permanent shock; and \(u_{i,j}\) is the transitory component. Suppose the permanent shocks are i.i.d. over time and across households but are joint-normal distributed within households,

\[
\begin{bmatrix}
v_{1,i,j} \\
v_{2,i,j}
\end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{v_1} & \rho v_1 v_2 \sigma_{v_1} \sigma_{v_2} \\ \rho v_1 v_2 \sigma_{v_1} \sigma_{v_2} & \sigma^2_{v_2} \end{bmatrix}\right).
\]

And the transitory components are just i.i.d white noises: \(u_{i,j} \sim N(0, \sigma^2_{u_i}), i = 1, 2\). Let \(\tilde{y}_{i,j} = \log(y_{i,j,t}) - F(i, j, t)\) and \(\Delta \tilde{y}_{i,j} = \tilde{y}_{i,j} - \tilde{y}_{i,j-1}\), then it is easy to derive that

\[
\text{var}(\Delta \tilde{y}_{i,j}) = \sigma^2_{v_i} + 2\sigma^2_{u_i};
\]

\[
\text{cov}(\Delta \tilde{y}_{i,j+1}, \Delta \tilde{y}_{i,j}) = -\sigma^2_{u_i};
\]

\[
\text{cov}(\Delta \tilde{y}_{1,j}, \Delta \tilde{y}_{2,j}) = \rho \sigma_{v_1} \sigma_{v_2} \sigma_{v_1} \sigma_{v_2}.
\]
Table 4: Permanent and Transitory Shocks (Data vs. Model)

<table>
<thead>
<tr>
<th></th>
<th>1970s Data</th>
<th>1970s Model</th>
<th>2000s Data</th>
<th>2000s Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Permanent Shocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{v_1}$</td>
<td>0.0211</td>
<td>0.0225</td>
<td>0.0287</td>
<td>0.0299</td>
</tr>
<tr>
<td>$\sigma^2_{v_2}$</td>
<td>0.0662</td>
<td>0.0635</td>
<td>0.0419</td>
<td>0.0404</td>
</tr>
<tr>
<td>$\rho_{v_1v_2}$</td>
<td>-0.0551</td>
<td>-0.0498</td>
<td>0.0663</td>
<td>0.0718</td>
</tr>
<tr>
<td><strong>B. Transitory Shocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{u_1}$</td>
<td>0.0404</td>
<td>0.0406</td>
<td>0.0357</td>
<td>0.0366</td>
</tr>
<tr>
<td>$\sigma^2_{u_2}$</td>
<td>0.0365</td>
<td>0.0378</td>
<td>0.0345</td>
<td>0.0354</td>
</tr>
</tbody>
</table>

Note: The results reported have been converted to the values corresponding to the one-year interval.

Therefore, we can estimate the covariance matrix of the permanent shocks and transitory components from these moments. The results are reported in Table 4. It is not a surprise that the model fits the data well in this aspect because the structural shocks in the model are calibrated to match the empirical moments used in this estimation.

Figure 1 plots the life cycle profiles of the 50th, 90th, and 99th percentiles of the male and female labor income for the years 2010-2012 and 1978-1980 in the data and in the model. The model matches well the life cycle profiles of the male labor income in both the 2010s and 1970s, which is a success of the model. For the female life cycle profiles, the differences between the model and the data are larger. This is partly because the assumption that the male and female learning abilities in each household are linked by the household level ability variable $A$, which limits the degrees of freedom in the distributions of the male and female learning abilities and therefore reduces the model’s ability to match all the patterns in the data. However for the 1970s, the more significant differences of the female life cycle profiles between the data and the model are mainly driven by the differences between the CPS and PSID data. The PSID data imply a large variance of permanent income shocks for females in the 1970s, and hence a large variance of human capital shocks for females in the model. As a result, even if there was no human capital accumulation at all, the cross-sectional dispersion of female labor income should rise rapidly with age as shown by the life cycle profiles in the model. But this is not the case in the life cycle profiles from the CPS data in the 1970s because CPS data are cross-sectional.

---

26Note the PSID data are biennial after 1997, so the formula should be adjusted accordingly. The sample is restricted to the married male and female with the head age between 30 and 55 in the years 1971-1980 and 1998-2012.
My own interpretation of this fact is that the income processes of young females and old females were already very different in the 1970s. For those females who were near retirement in the late 1970s, their income when they entered the labor market was probably much lower (if not zero) than the income of young females in the late 1970s.\footnote{One may then question whether we should use the cross-sectional life cycles to calibrate the model, or the life cycles from the same cohorts. Because the purpose of this paper is to study the income tax policy which mainly redistributes income cross-sectionally, I think the cross-sectional life cycles are the more appropriate calibration targets to use.}

Figure 2 plots the actual average income tax rates at different income levels together with the tax rates calculated from the fitted income tax functions for the years 2010-2012 and 1978-1980. As we can see, the income tax function can fit the actual tax schedule very well with the calibrated parameter values.

4 Quantitative Analysis and Results

In this section, I present the quantitative analysis and results based on the calibrated economic model. I first report the household life cycles in the model and show how household behaviors differ according to their learning abilities and between the 1970s
and 2010s. Then I report the inferred Pareto weights from the actual income tax policies in the 1970s and 2010s. With the inferred Pareto weights in the 1970s, I compute the optimal response of the income tax policy with respect to only the changes in economic fundamentals since the 1970s and ascribe the remaining part of income tax policy change to the change in Pareto weights.\(^{28}\) I also conduct a detailed decomposition of the income tax policy change with respect to each economic change in the model. Then I examine the sources of rising income inequality and labor income growth through the lens of my model and quantify the roles of income tax policy in those changes. Finally, I report the optimal income tax policy with utilitarian Pareto weights. All the results are based on simulated economies at the stationary competitive equilibrium with 20000 households in each cohort.

### 4.1 Household Life Cycles in the Model

Figure 3 plots the life cycles of three household groups in the model for the years 2010-2012. The three household groups are determined by their percentiles in the distribution of learning ability. In particular, those within the 10-percent intervals centered around the 10th, 50th, and 90th percentiles are selected, and the life cycle profiles plotted are the cross-sectional means within each group across ages.

Over the household life cycle, household consumption rises significantly at first due to increasing household income and binding borrowing constraints, but becomes relatively flat in the rest of the life cycle as a result of consumption smoothing. In the model,

\(^{28}\)Appendix B provides sensitivity analysis for these results.
there are two motives for households to save: the precautionary savings against future idiosyncratic risk and the savings for consumption after retirement. On average, households have almost no savings until their middle 30s and reach the highest wealth level at their retirement, which is common in life-cycle models. As households earn more income and accumulate savings over the life cycle, the share of borrowing-constrained households declines with age to almost zero after age 40.

The male labor income grows with age and reaches its peak around age 55. The male labor supply also grows with age, but the magnitude of the rise is smaller. The difference is the result of increasing labor productivity due to human capital accumulation, which is evident in the life cycle profiles of study time. Over half of the non-leisure time is devoted to human capital investment for the young male earners, and this effort declines over the life cycle as the benefit of human capital investment decreases relative to its cost, which is measured by the current earnings lost due to study. Near retirement, the benefit of additional human capital investment is very low since little time is left to collect it, whereas the cost is higher because the current wage is high due to previous human capital accumulation. Consequently, human capital investment is almost zero near retirement, and wage declines due to the negative mean of human capital shocks capturing the deprecia-
tion of human capital. This is the reason for the declining labor supply and labor income shortly before retirement. Compared with males, the rise of female labor income and labor supply are less significant over the life cycle, and a smaller share of the non-leisure time is devoted to human capital investment. This is because females are disadvantaged in production based on the calibrated technology parameters,29 and therefore the benefit of human capital investment is lower for females.

Depending on the learning ability, household life cycle profiles are quite different. For males, the high ability earners study more and work less than the low ability earners when they are young because the return to human capital investment is higher for them. Therefore, they have a much steeper rising labor income profile over the life cycle. Also, because the initial labor productivity is positively correlated with an earner’s learning ability, and the existence of tight borrowing constraints dictates households to finance their early consumption with contemporaneous income, the labor income of high ability males is still higher than that of the low ability ones in spite of less time worked.30 As human capital accumulation almost completes after age 50, the wages of high ability males are much higher. Hence, they work longer hours, and the male labor income inequality within the cohort reaches its peak over the life cycle.

For females, the life cycle profiles of study time are no longer monotonic in learning ability. In particular, among the young female earners, both the high and low ability females study less than the middle ability ones. This is partly due to the perfect assortative marriage assumption. In the model, a high ability female is also married with a high ability male who has the same rank in the corresponding ability distribution. Because the expected lifetime income is positively correlated with an earner’s ability at equilibrium, this reduces the incentive of the high ability female to increase her future labor income through human capital investment. On the other hand, the low ability females want to increase their labor income, but their return to human capital investment is too low due to their low abilities. While the same argument also works for males, as mentioned earlier, females are disadvantaged in production compared to males, and hence households optimally rely more on the male labor income. This is why the same effect is much weaker for males. For similar reasons, the labor supply of the high ability females is uniformly lower than that of the low ability ones.31

29This is mainly governed by the combination of $\theta_2$, $w_{2,1}$, and the distribution of $A_2$.
30If there were no borrowing constraints, the high ability earners would borrow to finance their early consumption, devote more time to study, and have lower income than the low ability earners when they are young.
31The trough of labor supply for the high ability females around age 40 is because: before age 40, the high ability households are borrowing constrained, and they need the female labor income to increase their
Because the high ability households have higher and steeper labor income profile over the life cycle, they have higher consumption, higher peak savings, and are more likely to be borrowing constrained when they are young.

Figure 4 plots the same household life cycles for the years 1978-1980. Because the overall productivity of technology is lower than that in the 2010s, the levels of household consumption and savings are both lower. Also, because the return to human capital investment is lower in the 1970s, the male earners study less and work more when they are young. As a result, their labor income life cycle profiles are flatter than those in the 2010s. For the female earners, they are even more disadvantaged in the 1970s than in the 2010s, so their labor income and labor supply decline over the life cycle, and there is almost no human capital accumulation for females. With flatter income profiles over the life cycle, a much smaller share of young households are borrowing constrained.

early consumption and to allow the male earners to study more; around age 40, most of the high ability households are away from the borrowing constraints, so the high ability females allocate more non-leisure time in study; after age 40, as the wage of high ability females increases, so is their labor supply.
4.2 Inferred Pareto Weights in the 1970s and 2010s

In Section 2.5 and 2.6, I have described the Ramsey optimal tax policy problem and the method to recover the Pareto weights implied by the actual income tax policy in that framework. The Pareto weights inferred this way capture the effects of non-economic forces in the determination of income tax policy as the economic forces are already taken into account by the household lifetime utilities in the objective function of the Ramsey problem.

One non-economic force which is widely recognized as a critical determinant for the actual income tax policy is the political influences of various income groups. However, it is extremely hard to identify those directly. Hence I consider the Pareto weights inferred this way as an indirect measure of the political influences and interpret the change in Pareto weights between the 1970s and 2010s as evidence of changes in political influences. I provide more details including empirical evidences and a political economy model in support of this interpretation in Section 5. It is certainly possible that other non-economic forces might also have affected the Pareto weights and the change of them, so my interpretation in this paper is not definite, but rather a good starting point for thinking the non-economic causes of income tax policy change.

Because the calibrated income tax function has only two parameters \((\tau, \mu)\), we can identify the Pareto weight function \(W(A)\) up to the first order approximation, i.e., the linear form. I choose \(F(A)\) as the basis function, where \(F(\cdot)\) is the cdf of household learning ability \(A\). So the functional form assumption of \(W(A)\) is \(W(A) = \xi_0 + \xi_1 F(A)\), and \(\{\xi_p\}_{p=0}^1\) are coefficients to be inferred from the actual income tax policy. Figure 5 reports the Pareto weights inferred from the U.S. income tax policies in the 1970s and 2010s.

For the years 1978-1980, the slope of the Pareto weight function is negative, which means that policymakers value the lifetime utilities of the low ability/income households more than those of the high ability/income households.\(^{32}\) The change of the Pareto weight function from the 1970s to 2010s can be decomposed into two steps: a change in level and a change in slope. The change in level is a scale-up of the Pareto weight function while keeping the relative importance of any two households unchanged, i.e., keeping the ratios between the weights at any two ability levels \(\frac{W(A')}{W(A)}\) unchanged. Recall that in the objective function of the Ramsey problem, the weight on government services \(\gamma\) is normalized to be 1 for identification purposes. Therefore, the scale-up of Pareto weights on household private utilities implies a relative decline in the importance of government ser-

---

\(^{32}\)Because the expected lifetime income increases with ability, the weights on high/low ability households can also be roughly interpreted as weights on high/low income households.
vices, at least as perceived by policymakers. The change in slope is a counterclockwise rotation of the Pareto weight function while keeping the relative importance of household private utilities with respect to government services unchanged, i.e., keeping the area under the weight function \( \int W(A)dF(A) \) unchanged. Due to this change in slope, the Pareto weights assigned to the high ability households are much larger than those at the lower end of the ability distribution in the years 2010-2012. Hence the overall change in Pareto weights between the 1970s and 2010s implies less valued government services and benefits the high ability/income households.

4.3 Optimal Response of Income Tax Policy to Economic Changes

A key question to answer in this paper is how much of the U.S. income tax policy change since the 1970s can be rationalized as an optimal response of income tax policy to changes in economic fundamentals. To address this question, we need to separate the effects of economic forces and non-economic forces in shaping the income tax policy change we observe. In particular, we can use the Pareto weights inferred from the 1970s income tax policy as a measure of the non-economic forces in the 1970s and then combine them with the economic model calibrated to the 2010s U.S. economy in a Ramsey optimal tax policy problem. The solution to such counterfactual Ramsey problem is the income tax policy which would be chosen in the 2010s if there were only changes in economic fundamentals since the 1970s. Hence the difference between the solution to this
Ramsey problem and the actual income tax policy in the 1970s gives the optimal response of income tax policy to only the economic changes. The remaining change of income tax policy since 1970s is then attributed to the change in Pareto weights by the structure of the Ramsey framework.

The optimal response of income tax policy to economic changes computed using the above method is plotted in Figure 6, together with the actual income tax policies in the 1970s and 2010s. In terms of the progressivity of income tax, the elasticity of after-tax income with respect to before-tax income implied by the income tax progressivity parameter $\mu$ in the tax function is about 0.856 in the 1970s and 0.914 in the 2010s. The optimal response of income tax policy to economic changes implies an elasticity of 0.879. That means the optimal response tax policy is less progressive than the 1970s tax policy but more progressive than the 2010s tax policy and quantitatively account for about 40% of the reduction in progressivity between the 1970s and 2010s. This result is apparent in the right graph of Figure 6 where I normalize the tax rates under different tax policies at income level one to eliminate the differences in the level of income tax from the graph.

![Figure 6: Optimal Income Tax Response to Economic Changes](image)

Note: The bounds of household income in the graphs correspond to the cutoff income levels of the first and last tax bracket in 2012. The graph on the right plots the average tax rates under different policies subtracted by the corresponding average tax rates at income level 1, which reflect the levels of income tax $\tau$ but are not affected by the progressivity parameter $\mu$.

In terms of the level of income tax, the optimal response of income tax policy is almost the same as the income tax in the 1970s as shown in the left graph of Figure 6. The value of the income tax level parameter $\tau$ in the tax function is about 0.258 in the 1970s, 0.220 in
the 2010s, and 0.259 for the optimal response of income tax policy. Therefore, most of the reduction in income tax level since the 1970s is due to the change in Pareto weights, more specifically, the lower weight on government services.

### 4.4 Decomposition of Income Tax Policy Change

Several economic changes have occurred since the 1970s as demonstrated by the calibration results of the economic model in Section 4, including skill-biased technological change, increased female labor productivity, change of idiosyncratic risk, and aging of the U.S. population, etc. To further understand how each economic change contributes to the income tax policy change separately, I conduct a detailed decomposition of the income tax policy change since the 1970s by solving a sequence of counterfactual Ramsey problems.

The exercise starts from the Ramsey problem with the 1970 Pareto weights and the economic model calibrated to the 1970s U.S. economy. By construction, the actual income tax policy in the 1970s is the solution to this problem. Then I introduce economic changes sequentially into the economic model. After all the economic changes are included, I introduce the change in Pareto weights in two steps: first add the change in level and then the change in slope as defined in Section 4.2. Whenever the economic model or the Pareto weights are updated, the optimal income tax policy is solved for the corresponding Ramsey problem. The change of optimal income tax policy between two consecutive steps is ascribed to the change introduced between them. The results of this decomposition are reported in Table 5.

For the progressivity of income tax, economic changes overall account for 39.8% of reduction in the data, but each economic change contributes differently to this result, both quantitatively and qualitatively. The first economic change introduced is the change of idiosyncratic risk which includes both the initial labor productivity risk and the human capital shocks over the life cycle. The increase of idiosyncratic risk causes a higher income inequality and raises the redistribution/insurance benefit of progressive taxation. As a result, it contributes negatively to the reduction in progressivity and accounts for −88.5% of the change in the data. The female-biased technological change increases female labor productivity, which is captured in the model by the increase of $\theta_2$. Because female labor supply is more elastic than male labor supply, this change requires the optimal income tax to be less progressive and can account for 44.8% of the reduction in progressivity. The skill-biased technological change increases the return to human capital accumulation $R^i_{H}$ in the model and benefits the high ability earners more than the low ability ones.
Table 5: Decomposition of Income Tax Policy Change

<table>
<thead>
<tr>
<th>Due to</th>
<th>% in Total Change of Income Tax</th>
<th>% in Total Change of Income Tax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Progressivity ($\mu$)</td>
<td>Level ($\tau$)</td>
</tr>
<tr>
<td>A. Economic Changes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change of Idiosyncratic Risks</td>
<td>$-88.5%$</td>
<td>$67.7%$</td>
</tr>
<tr>
<td>Female-biased Technological Change</td>
<td>$44.8%$</td>
<td>$28.7%$</td>
</tr>
<tr>
<td>Skill-biased Technological Change</td>
<td>$18.0%$</td>
<td>$122.0%$</td>
</tr>
<tr>
<td>Universal Technological Change</td>
<td>$-8.1%$</td>
<td>$-60.6%$</td>
</tr>
<tr>
<td>Aging of Population</td>
<td>$73.5%$</td>
<td>$-158.5%$</td>
</tr>
<tr>
<td>Subtotal</td>
<td>$39.8%$</td>
<td>$-0.7%$</td>
</tr>
<tr>
<td>B. Change in Pareto Weights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta$Level</td>
<td>$-58.8%$</td>
<td>$179.1%$</td>
</tr>
<tr>
<td>$\Delta$Slope</td>
<td>$119.0%$</td>
<td>$-78.4%$</td>
</tr>
<tr>
<td>Subtotal</td>
<td>$60.2%$</td>
<td>$100.7%$</td>
</tr>
<tr>
<td>Total Change</td>
<td>$-0.058$</td>
<td>$-0.038$</td>
</tr>
</tbody>
</table>

This induces a higher income inequality. However, it also increases the efficiency cost of progressive taxation, which discourages human capital investment. The higher income inequality requires a more progressive income tax for redistribution, whereas the larger efficiency cost demands the tax policy to move in the opposite direction. In the end, my quantitative result suggests that the efficiency cost channel dominates the redistribution channel, and overall the skill-biased technological change accounts for 18.0% of the reduction in progressivity. The universal technological change represents the decline of overall labor productivity, i.e., the declining wage of effective labor $w_e$ in the model. This change has a relatively small effect on progressivity because it affects all earners in a similar way. Finally, the aging of the U.S. population, i.e., larger values of $T_R$ and social security benefits $b$ in the model, increases the age dependency ratio, which means more tax revenues need to be collected from the working age population to finance the rising demand for social security benefits and government services. This change results in a less progressive optimal income tax to boost tax revenues and explains 73.5% of the reduction in progressivity.

The change in Pareto weights accounts for the rest 60.2% of reduction in progressivity, more specifically, the change in slope of the Pareto weight function $W(A)$. The change in level of the Pareto weight function reduces the importance of government services. Therefore, it lowers the demand for tax revenues and actually requires a more progres-
sive income tax to be adopted. However, this effect of change in level on progressivity is completely offset and reversed by the change in slope of the Pareto weight function, which benefits the high ability/income households and produces a much less progressive income tax policy in the end.

For the level of income tax, the economic changes which raise the total income lead to lower level of income tax, such as the change of idiosyncratic risks and female-biased and skill-biased technological changes, whereas the economic changes which reduce the total income or raise the demand for tax revenues increase the level of income tax, such as the universal technological change and aging of population. In spite of the significant impacts of each economic change, their effects counteract each other, and hence the comprehensive effect of economic changes on the income tax level is quite small. On the other hand, the change in level of Pareto weights is responsible for most of the reduction in level of income tax since the 1970s.

4.5 Income Inequality, Growth, and Income Tax Policy

Through the lens of the economic model calibrated to match the economic changes between the 1970s and 2010s, we can ask how each economic change contributes to the rising income inequality and income growth since the 1970s, and what are the roles of the income tax policy in those changes. To answer this question, I conduct a decomposition for the change of labor income inequality and mean labor income similar to the one for the change of income tax policy in Section 4.4. Starting with the economic model calibrated to the 1970s U.S. economy, economic changes are introduced sequentially into the model, and the income tax change is added in the last step of this exercise. The decomposition results are reported in Table 6.

The income inequality is measured by the ratio between the 90th percentile and 50th percentile of the cross-sectional labor income distribution in the model. I report the decomposition results with this measure of inequality for both the male labor income and household labor income.33 The decomposition shows that the increase of initial productivity risk and skill-biased technological change are the two main causes of the rising inequality of both the male labor income and household labor income since the 1970s. Notably, female-biased technological change increases the male labor income inequality but reduces the household labor income inequality. This is a result of behavior interaction between the male and female earners within households. As female labor produc-

33The change of this ratio from the years 1978-1980 to 2010-2012 in the CPS data is 0.684 for working males and 0.312 for married working couples. Although the model is not directly calibrated to match these ratios, we can see it matches these rising inequality patterns in the data reasonably well.
Table 6: Sources of Rising Income Inequality and Income Growth

<table>
<thead>
<tr>
<th>Due to Change of</th>
<th>% in Total Change of p90-p50 Income Ratio</th>
<th>% in Total Change of Mean Labor Income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Household</td>
</tr>
<tr>
<td>Initial Productivity Risk</td>
<td>36.4%</td>
<td>65.5%</td>
</tr>
<tr>
<td>Human Capital Shocks</td>
<td>5.9%</td>
<td>−0.5%</td>
</tr>
<tr>
<td>Female-biased Technological Change</td>
<td>6.3%</td>
<td>−18.6%</td>
</tr>
<tr>
<td>Skill-biased Technological Change</td>
<td>53.7%</td>
<td>44.0%</td>
</tr>
<tr>
<td>Universal Technological Change</td>
<td>−3.0%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Other Economic Factors</td>
<td>−3.7%</td>
<td>5.3%</td>
</tr>
<tr>
<td>Income Tax</td>
<td>4.4%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Total Change</td>
<td>0.670</td>
<td>0.279</td>
</tr>
</tbody>
</table>

Activity increases, the idiosyncratic shocks to females become more important to household income. And hence, the male earners have to adjust their behaviors more in response to these shocks, such as labor supply and human capital investment, to smooth household income. That is why the male labor income inequality increases. However, at the same time, the higher female labor productivity increases the ability of female earners to provide insurance against the idiosyncratic shocks to males. As a result, it reduces the dispersion of income at household level. The change of income tax policy since the 1970s increases the income inequality due to the less progressive tax schedule, but the contribution is much smaller relative to those of the two main economic causes.

The income growth is measured by the change of mean labor income of males and females. For males, the most important source of income growth is skill-biased technological change which increases the return to human capital investment, but a significant part of its effect is offset by the universal technological change which reduces the labor productivity for all earners. Female-biased technological change which increases female labor productivity is an additional contributing factor to the growth of female labor income, but it reduces the male labor income at the same time because households optimally shift a part of the burden of earning income to the female earners. The change of income tax policy since the 1970s plays a more significant role in the income growth than in the rising income inequality, and it accounts for 19.4% and 14.2% of the growth of mean labor income of males and females. In terms of 2012 dollars, the growth of mean labor income due to the change of income tax policy is about $1587 and $2596 per year for males and females, respectively.

Overall, economic changes are the main causes of both the rising income inequality.
and income growth since the 1970s. The change of income tax policy since the 1970s contributes to both of them positively but has a larger effect on income growth. This indicates the classic trade-off between equity and efficiency for the income tax policy.

4.6 Optimal Income Tax with Utilitarian Weights

Following the convention of optimal tax policy literature, I also compute the optimal income tax policy for the early 2010s U.S. economy with utilitarian Pareto weights as the normative criterion in the Ramsey problem. One important choice to make in this exercise is how much the government services should be valued relative to household private utilities. Since the weight on government services is normalized to be one in the objective function of the Ramsey problem, this is equivalent to choosing the constant value of utilitarian weights on household lifetime utilities. I consider two alternative values for this choice, and for each of them, the relative importance of government services corresponds to that implied by the income tax policy in the 1970s and 2010s, respectively.\footnote{More specifically, the two levels of utilitarian weights are: $W_U^{2010s} = \int W^{2010s}(A)dF(A)$ and $W_U^{1970s} = \int W^{1970s}(A)dF(A)$. The 1970s income tax policy implies a more valued government services than the 2010s income tax policy.}

Figure 7 plots the optimal and the actual income tax policies for the early 2010s. As the graphs show, both the progressivity and level of the optimal income tax depend on the relative importance of government services. If government services are less important as implied by the actual income tax policy in the 2010s, the optimal income tax should be much more progressive and lower in level. The elasticity of after-tax income with respect to before-tax income should be 0.854 relative to the value of 0.914 for the actual income tax in the 2010s, and the level of income tax as measured by $\tau$ should be 0.192 relative to the value of 0.220 in reality. However, if government services are more important as implied by the actual income tax policy in the 1970s, the progressivity of the optimal income tax would reduce and imply an elasticity of 0.890, closer to the actual income tax in the 2010s. And the level of the optimal income tax would rise significantly to $\tau = 0.263$.

One interesting takeaway from this exercise is that if government services become more important or the government wants to collect more tax revenues, the economically optimal response of income tax policy is to lower the progressivity and raise the overall level of income tax, assuming the preferences of policymakers over different households remain the same. From this point of view, the recent increase of marginal income tax rate for the top income bracket in the year 2013 is more likely to be a result of “rebalancing” the Pareto weights rather than rebalancing the government budget and collecting more
Figure 7: Optimal Income Tax with Utilitarian Weights

Note: “Utilitarian 2010s” and “Utilitarian 1970s” denote the optimal income tax policy for the 2010s U.S. economy with utilitarian weights and relative importance of government services implied by the 2010s and 1970s income tax policies. The bounds of household income in the graphs correspond to the cutoff income levels of the first and last tax bracket in 2012.

tax revenues.

5 Explanations for the Change in Pareto Weights

The quantitative results in Section 4 show that the change of income tax policy in the U.S. since the late 1970s implies a change in Pareto weights benefiting the high ability/income households, which contributes significantly to the reduction in progressivity of income tax we observe. In this section, I provide potential explanations for this change in Pareto weights from a political economy point of view. In particular, I discuss two possible causes of this change: (1) the lower cost of conveying information to swing voters due to information technology improvement; (2) the rising inequality of voter turnout among different socioeconomic groups. I first present empirical evidences in support for these explanations and discuss the intuitions. Then in a stylized probabilistic voting model with political contributions, I derive a closed-form expression for the Pareto weight function in the Ramsey framework and show that the two causes proposed can indeed induce a change in Pareto weights benefiting the high ability/income households, consistent with what the change of income tax policy implies in Section 4.
5.1 Empirical Evidences

5.1.1 Money in Political Campaigns

It is not a secret that money plays an important role in political elections. In the 2012 presidential campaign, Barack Obama and Mitt Romney each spent around 1 billion dollars, which is arguably enough to send a person to the moon. Most of the money was spent on media and other forms of political persuasion to potential voters. Previous studies have shown that the political information, whether biased or not, delivered by media has real effects on voting behaviors. For example, DellaVigna and Kaplan (2007) find that the introduction of Fox News between October 1996 and November 2000 convinced 3 to 28 percent of its viewers to vote Republican, depending on the audience measure; Ladd and Lenz (2009) estimate that the endorsement switch to the Labour Party by several prominent British newspapers before the 1997 United Kingdom general election persuaded 10 to 25 percent of their readers to vote for Labour, depending on the statistical approach.

There is evidence that the importance of money in the U.S. elections has increased. Figure 8 plots the normalized real average campaign expenditures per candidate for the U.S. House of Representatives and Senate since 1974. It is apparent that campaign expenditures have grown a lot and faster than GDP. Since politicians can only use their campaign funds in elections to improve their chances of winning, this sharp rise of campaign expenditures relative to GDP implies that money might have become more effective for the purpose of gaining votes.

Why money has become more important in political campaigns? One possible reason is the lower cost of conveying information to swing voters due to information technology improvement. It is obvious that the cost of passing through information to voters is much lower today than in the 1970s due to the expansions of television and telephone networks and most recently the internet. The lower transportation costs today also make it easier for politicians and voters to gather in person more often. Political campaigns are all about conveying information to voters and persuading them to vote accordingly. The advantage of spending one more dollar than the opponents depends on how many additional information flows to voters can be generated with this amount of money. When the cost of information flows is high, the incentive of politicians to collect and spend more money in such activity is low. On the other hand, when that cost is lower, politicians may devote more efforts in fund-raising activities and spend larger amount of money in all kinds of media to improve their chances of winning as we observe nowadays.

\[35\text{Source: Campaign Finance Institute analysis of Federal Election Commission data.}\]
The increased demand for campaign funds may induce politicians to propose policies more favorable to the high-income households because the high-income households are typically more willing and able to donate more to their preferred politicians. This could potentially explain the change in Pareto weights benefiting the high-income households inferred from the change of income tax policy in Section 4.

5.1.2 Rising Inequality of Voter Turnout

A basic fact about political elections is that not all people participate in voting. For example, only 56.5% of voting age population voted in the year 2012. And the participation rate, i.e., voter turnout, varies a lot among populations of different socioeconomic groups. Data suggest that voter turnout increases with income, educational attainment and age. Since the 1970s, there are evidences that such inequality of voter turnout among socioeconomic groups has increased. Figure 9 plots voter turnout for population groups with different educational attainments (hence different learning ability) since the mid-1970s.\textsuperscript{36} Voter turnout has declined over time for all the groups, but the reduction is more significant for the groups with lower educational attainments. Consequently, the shares of

\textsuperscript{36}Source: U.S. Census Bureau, Current Population Survey, Voting and Registration data.
votes from the low education groups in elections have declined, whereas the opposite is true for the high education groups. Since politicians care about winning elections, they certainly should have responded to this change in the composition of voters by adjusting their policies towards the high education/income households. Similar patterns exist for population groups with different income levels, and I refer the reader to Freeman (2003) for more detailed information and discussions.

5.2 A Probabilistic Voting Model with Political Contributions

Motivated by the empirical evidences and intuitions presented in Section 5.1, I build a stylized probabilistic voting model with political contributions to formalize the argument and derive a closed-from expression for the Pareto weight function in the Ramsey framework, which shows explicitly how different factors affect the Pareto weights. Unlike the quantitative life-cycle model in Section 2, which is used to match data and deliver reliable numerical results, the probabilistic voting model here is mainly used to illustrate ideas and help us understand the determination of Pareto weights from a political economy point of view. Hence the economic side of the model is simplified to avoid unnecessary complications. The notations in this section are independent from other parts of this paper.

Consider an economy populated by a measure one of infinitely-lived households with heterogeneous ability \( A \). The distribution of ability is governed by the cdf \( F(A) \). The before-tax income of a household \( y_A \) is exogenous and increases with its ability level.
Households are not allowed to save or borrow for simplicity. Let $\tau$ denote the income tax policy, then household consumption is simply equal to the after-tax income, i.e., $c_A(\tau) = y_A - T(y_A, \tau)$, where $T(\cdot, \tau)$ is the income tax function governed by $\tau$. The instantaneous utility function is $u(\cdot)$.

Suppose for time $t \leq T$, the policy is $\tau^0$, and an election is organized at the end of time $T$. In this election, two candidates (or parties), $a$ and $b$, propose future policy $\tau^a$ and $\tau^b$ for time $t \geq T + 1$, and have full commitment power. The objective of each candidate is to maximize its own number of votes received. Each household has the right of exactly one vote, and voter turnout is $\pi_v(A)$ for the ability $A$ households.

Conditional on voting, an ability $A$ household can either be a determined voter with probability $\pi_d(A)$ or a swing voter with probability $\pi_s(A) = 1 - \pi_d(A)$. The determined voters vote according to their discounted utilities from time $t \geq T + 1$ under the policies proposed by candidates, $V_A(\tau^a)$ and $V_A(\tau^b)$. There is also an idiosyncratic component $\varepsilon_A$ reflecting the heterogeneous tastes among the ability $A$ voters with respect to other characteristics of candidates. $\varepsilon_A$ is i.i.d across voters and follows a uniform distribution governed by $\phi_A$, i.e., $\varepsilon_A \sim \text{Unif}(-\frac{1}{2\phi_A}, \frac{1}{2\phi_A})$. The probability of an ability $A$ determined voter to vote for candidate $a$ is then

$$\Pr\{\text{vote for } a | (d, A)\} = \Pr\{V_A(\tau^a) - V_A(\tau^b) > \varepsilon_A\} = \phi_A[V_A(\tau^a) - V_A(\tau^b)] + \frac{1}{2} \equiv P_A(\tau^a, \tau^b)$$

where $d$ represents the type “determined”.

The voting behaviors of swing voters are similar to those of determined voters, but their preferences over the two candidates can be affected by the amounts of information they receive from each candidate. The information flows from a specific candidate to a swing voter increase the probability of the swing voter to vote for this candidate and reduces the same amount of that for the other candidate.

To finance the money costs of information flows to swing voters, candidates need to

---

37In general, $\tau$ can be a vector, but for notation’s sake, I write it as a scalar when deriving the formulas. It is straightforward to extend them to the vector case.

38This case is like two parties fighting over the number of seats in congress. If we assume the candidates maximize their probabilities to win, i.e., probabilities of receiving more than a half of the total votes, instead of maximizing the number of votes received, the qualitative results would be similar in the end, but the notations would become more complicated.

39Voter turnout is defined as the percentage of voting age population who actually vote in the election.

40The value of $V_A(\tau^a) - V_A(\tau^b)$ is assumed to be within the support of the uniform distribution when calculating the probability. This is not an issue because I focus on the symmetric Nash equilibrium later at which $V_A(\tau^a) - V_A(\tau^b) = 0$. 

41
raise campaign funds from the group of determined voters who support them. In particular, a candidate can devote effort $e$ in fund-raising activities, and encourages its own determined voters to donate a part of their current consumption as political contributions to the candidate. On the other side, determined voters enjoy providing support to the candidate they would vote for, and their utility function is assumed to be

$$u(c) = \max_{(c_p, c_d)} u\left((\frac{c_p}{1 - \chi(e)})^{1 - \chi(e)}\left(\frac{c_d}{\chi(e)}\right)^{\chi(e)}\right)$$

s.t.

$$c_p + c_d = c, c_p \geq 0, c_d \geq 0$$

where $c$ is the total consumption, $c_p$ is the private consumption, and $c_d$ is the political contribution; $\chi(e) \in (0, 1)$ controls the weight of political contribution in the determined voter’s utility, and it satisfies $\chi'(0) = +\infty$, $\chi'(1) = 0$, and $\chi''(\cdot) < 0$. With this specification of utility function, a determined voter always donates $\chi(e)$ share of its total consumption $c$, but its total utility level is not affected by the effort $e$.\footnote{This setting of utility function is to make sure that the voting decisions of determined voters are not affected by the fund-raising efforts of candidates.}

Each candidate is endowed with total one unit of effort, and hence only $1 - e$ unit of effort can be devoted to non-fund-raising activities, which increase the probability for a swing voter to vote for this candidate by $B(1 - e)$ and reduce the same amount for the other candidate. The function $B(\cdot)$ satisfies $B'(0) = +\infty$, $B'(1) = 0$, and $B''(\cdot) < 0$.

Therefore, the probability of an ability $A$ swing voter to vote for candidate $a$ is

$$\Pr\{\text{vote for } a | (s, A)\} = P^a_A(\tau^a, \tau^b) + z(n^a - n^b) + B(1 - e^a) - B(1 - e^b)$$

where $s$ represents the type “swing”; $z$ is the effectiveness of information flows in persuading the swing voter; $n^a$ and $n^b$ are the amounts of information flows delivered to this swing voter by candidate $a$ and $b$; $e^a$ and $e^b$ are the levels of effort devoted to fund-raising activities by candidate $a$ and $b$.

Given candidate $b$’s choice of $(\tau^b, e^b)$, the optimal campaign strategy problem of candidate $a$ is

$$\max_{(\tau^a, e^a, n^a, n^b)} \int P^a_A(\tau^a, \tau^b) \pi_d(A) \pi_v(A) dF(A)$$

$$\quad + \int [P^a_A(\tau^a, \tau^b) + z(n^a - n^b) + B(1 - e^a) - B(1 - e^b)] \pi_s(A) \pi_v(A) dF(A)$$
s.t.

\[ pn^a \int \pi_s(A)\pi_v(A)dF(A) = \int \chi(e^a)c_A(\tau^0)P_A(\tau^a, \tau^b)\pi_d(A)\pi_v(A)dF(A) \]
\[ pn^b \int \pi_s(A)\pi_v(A)dF(A) = \int \chi(e^b)c_A(\tau^0)[1 - P_A(\tau^a, \tau^b)]\pi_d(A)\pi_v(A)dF(A) \]
\[ 0 \leq e^a \leq 1, \tau^a \in \Gamma, n^a \geq 0, n^b \geq 0 \]

where \( p \) is the price of information flows, and \( \Gamma \) is the feasible set of policy \( \tau \). The first row of the objective function corresponds to the votes for candidate \( a \) from determined voters; the second row corresponds to the votes for candidate \( a \) from swing voters. The choice variables \( n^a \) and \( n^b \) in the objective function can be substituted out using the budget constraints, and the first order conditions of candidate \( a \)'s problem are then:

\[ \tau^a : \int \phi_A\{1 + \frac{z}{p}[\chi(e^a) + \chi(e^b)]c_A(\tau^0)\pi_d(A)\}V'_A(\tau^a)\pi_v(A)dF(A) = 0; \quad (5.1) \]
\[ e^a : \chi'(e^a)\frac{z}{p}\int c_A(\tau^0)P_A(\tau^a, \tau^b)\pi_d(A)\pi_v(A)dF(A) = B'(1 - e^a) \int \pi_s(A)\pi_v(A)dF(A). \]

(5.2)

Since candidate \( a \) and \( b \) are identical except for their names, I focus on the symmetric pure strategy Nash equilibrium at which both candidates chose the same optimal strategy, i.e., \( \tau^{a*} = \tau^{b*} = \tau^* \) and \( e^{a*} = e^{b*} = e^* \).

Define the function \( W(A) \) as

\[ W(A) \equiv \phi_A\{1 + 2\frac{z}{p}[\chi(e^*)c_A(\tau^0)\pi_d(A)]\pi_v(A)\}. \]

(5.3)

Then Equation (5.1) at the symmetric Nash equilibrium becomes

\[ \int W(A)V'_A(\tau^*)dF(A) = 0, \]

which is the same as the first order condition of a Ramsey optimal policy problem with the Pareto weight function \( W(A) \). Hence, the equilibrium policy \( \tau^* \) from the political economy model is also the solution to the Ramsey problem with Pareto weight function

\footnote{The reason why \( n^b \) is also a choice variable and the budget constraint of candidate \( b \) is also a constraint for candidate \( a \) is because the choice of \( \tau^a \) affects the budget constraint of candidate \( b \) and hence \( n^b \), which should be taken into consideration by candidate \( a \). In other words, the strategy of each candidate is only \((\tau, e)\), and \((n^a, n^b)\) are endogenously determined by the budget constraints.}
\( W(A),^{43} \) i.e.,

\[
\tau^* = \arg \max_{\tau \in \Gamma} \int W(A) V_A(\tau) dF(A).
\]

This justifies the use of the Ramsey framework as a parsimonious way to model the actual policy-making process in the quantitative study of this paper.

Equation (5.3) offers a closed-form expression for the Pareto weights in the Ramsey framework, and hence it allows us to link changes of different factors into changes of Pareto weights. Among all the factors, those corresponding to the previous discussion in Section 5.1 are the price of information flows \( p \) and the ratios \( \pi_\tau(A_1) / \pi_\tau(A_2) \) between different ability levels. Proposition 1 states how the Pareto weight function is related to these two factors.\(^{44}\)

**Proposition 1** Let \( A_1 \) and \( A_2, A_1 > A_2, \) denote two ability levels, then under the assumptions made in the model, the ratio of Pareto weights between these two ability levels, \( \frac{W(A_1)}{W(A_2)} \), is

1. decreasing with \( p \) if \( c_{A_1}(\tau^0) \pi_d(A_1) > c_{A_2}(\tau^0) \pi_d(A_2) \);

2. increasing with the ratio \( \frac{\pi_\tau(A_1)}{\pi_\tau(A_2)} \) if \( e^* \) is fixed.

Mapping the results in Proposition 1 to the discussions in Section 5.1: (1) A decrease of \( p \) lowers the cost of conveying information to swing voters, and hence it increases the effectiveness of money in gaining votes for candidates. The model predicts that this change increases the Pareto weights of the high ability households as long as the high ability households have higher \( c_A(\tau^0) \pi_d(A) \) than the low ability households, which is likely to be the case in reality.\(^{45}\) The model also implies that more efforts are devoted by candidates to the fund-raising activities when \( p \) is lower, and therefore the total political contributions rise relative to the size of the economy, which is consistent with the empirical evidence on campaign expenditures in Figure 8. (2) In the model, the ratio \( \frac{\pi_\tau(A_1)}{\pi_\tau(A_2)} \) measures the inequality in voter turnout between the high and low ability households. Figure 9 shows that this ratio has increased in the data. In response to such change, the model predicts an increase of \( \frac{W(A_1)}{W(A_2)} \), i.e., the Pareto weights on the high ability households would increase relative to those on the low ability households, if the fund-raising effort re-

\(^{43}\)Suppose the first order condition is both necessary and sufficient.

\(^{44}\)The proof of Proposition 1 is straightforward given the expression for Pareto weights in Equation (5.3) and the functional form assumptions on \( \chi(\cdot) \) and \( B(\cdot) \).

\(^{45}\)The high ability (income) households clearly should have higher consumption. The share of determined voters \( \pi_d(A) \) is harder to measure directly, but the high ability households typically participate more in political activities and receive better education, so they should be less affected by the media, i.e., be more determined and believe in their own knowledge and judgement.
Therefore, the political economy model confirms that the two proposed explanations for the change in Pareto weights, the lower cost of conveying information to swing voters and the rising inequality of voter turnout, would indeed induce a change in Pareto weights benefiting the high ability/income households, consistent with what is inferred from the change of income tax policy in the quantitative study.

6 Conclusions

In this paper, I examined the causes of the less progressive income tax in the U.S. since the 1970s in the Ramsey optimal tax policy framework. Using a quantitative life-cycle model with heterogeneous households calibrated to match the U.S. economy in the late 1970s and early 2010s, I showed that changes in economic fundamentals alone require a less progressive optimal income tax to be adopted and can account for 40% of the reduction in progressivity we observe. In particular, skill-biased technological change, increased female labor productivity, and the aging U.S. population contribute to this reduction in progressivity, but their effects are partially offset by the increase of idiosyncratic risk. The remaining change of income tax implies a change in Pareto weights in the Ramsey framework. The change in Pareto weights benefits the high ability/income households and lowers the importance of government services. Finally, I proposed two potential explanations for this change in Pareto weights from a political economy point of view: the lower cost of conveying information to swing voters due to information technology improvement and the rising inequality of voter turnout among different socioeconomic groups.

The quantitative life-cycle model in this paper has already shown its success in matching several aspects of the data, but there is still room for improvement such as the inclusion of female extensive margin and an explicit modeling of the marriage and divorce processes. These features are not in the model of this paper due to the limit of computing power. With more computing power available or some simplifying assumptions on other aspects of the model, it might be possible to take those features into account in future work and examine how they affect female and household behaviors such as labor supply and human capital accumulation decisions.

The Ben-Porath style human capital accumulation in the model of this paper represents “on-the-job training”. There is another way of modeling human capital accumula-

---

46If the fund-raising effort is allowed to respond to the changing inequality of voter turnout, the final result depends on the quantitative properties of the model such as the exact distributions of voter turnout, household consumption and share of determined voters, etc.
tion in the literature, which is “learning-by-doing”. It would be helpful for future work to build a quantitative life-cycle model with “learning-by-doing” human capital accumulation, and examine how different the implications on income tax policy are.

Finally, my quantitative study ascribes a significant part of observed change in income tax policy to a change in Pareto weights benefiting the high-income households. I interpret it as evidence of changing political influences of various income groups and provide my own explanations. My political economy model only considers a one-shot income tax policy change, and hence it does not allow the possible effects of the tax policy change on the distribution of political influence, which could again affect future tax policy. Future work allowing this channel from income tax policy to the distribution of political influence could provide more insight into this issue and help us understand the evolution of income tax policy and the distribution of political influence jointly.

References


A Income Inequality, Gender Gap, and Tax Policy

This section presents empirical facts mentioned in the introduction of this paper. Figure 10 reports the time series of the 10th, 50th, and 90th percentiles of the labor income distribution for males and females since the late 1970s. Most of the rising income inequality is caused by the widening gap in the upper half of the distributions. Figure 11 plots the 50th and 90th percentiles of the labor income distribution over the life cycle for males and females in the years 1978-1980 and 2010-2012. The income inequality did not change much for the young earners but increased sharply for the middle-age and old earners. The female life cycle profiles of income have become higher in level and steeper in slope. Figure 12 plots the time series of the labor income ratio between females and males in the U.S. economy since the late 1970s. The female labor income used to be only one third of male labor income in the late 1970s, but has risen to about two thirds of male labor income in the early 2010s.

Figure 13 reports the federal income tax policy since the year 1979 when Jimmy Carter was the president until the year 2013 under the Barack Obama administration. Overall, federal income tax has become lower in level for almost all income levels, but the tax cuts were larger for high-income households. Even though the marginal tax rate at the very top was increased in 2013, most of the tax schedule remained the same as after the Bush tax cut in 2003, and it is still much less progressive than the tax policy in the 1970s.
Figure 10: Rising Income Inequality since the 1970s

Note: Only labor income is included. The income data are from CPS and have been converted to 2012 dollars.

Figure 11: Rising Income Inequality over Life Cycle

Note: Only labor income is included. The income data are from CPS and have been converted to 2012 dollars.
Figure 12: Rise of Female-Male Income Ratio
Note: Only labor income is included. The income data are from CPS. The time series plotted is the ratio between total labor income of females and males in each year.

Figure 13: The U.S. Income Tax Policy since the 1970s
Note: Income tax policy data are from NBER’s TAXSIM program. The tax rates plotted are for married couples filing jointly.
B Sensitivity Analysis

In this section, I provide sensitivity analysis for the main quantitative results in Section 4 and show that the main conclusions of this paper are robust to these variations.

B.1 Alternative Basis Function for Pareto Weight Function

In Section 4.2, I reported the inferred Pareto weight function with $F(A)$ as the basis function, i.e., $W(A) = \xi_0 + \xi_1 F(A)$. An alternative basis function is simply $A$, i.e., $W(A) = \xi_0 + \xi_1 A$. The inferred Pareto weight function with this specification is reported in Figure 14. We can see the change in Pareto weights between the 1970s and 2010s with this specification is similar to the benchmark case in the main text. The Pareto weights in the 2010s are higher in level, i.e., the government services are less important, and the change in Pareto weights from the 1970s benefits the high ability/income households.

![Figure 14: Inferred Pareto Weights $W(A) = \xi_0 + \xi_1 A$](image)

Note: The lower and upper bounds of ability in the graph correspond to the 1st and 99th percentiles of the ability distribution. The vertical line represents the 90th percentile of the ability distribution.
Figure 15: Quantitative Results with Alternative Government Budget Constraint

### B.2 Alternative Specifications of the Ramsey Problem

#### B.2.1 Alternative Government Budget Constraint

In the benchmark case, the government is assumed to balance its budget period-by-period. An alternative specification is to assume the government to balance its budget cohort-by-cohort:

\[
\sum_{j=1}^{T} \frac{1}{(1+r)^{j-1}} \int T \left( \sum_{i=1}^{2} y_{ij}(s) \right) d\Phi_j(s) = \sum_{j=T+1}^{T+T_R} \frac{1}{(1+r)^{j-1}} b + \sum_{j=1}^{T+T_R} \frac{1}{(1+r)^{j-1}} \left( \frac{G}{T+T_R} \right).
\]

Because the objective function of the Ramsey problem is equivalent to a weighted sum of expected lifetime utilities of a newborn cohort, the advantage of this specification of government budget constraint is that there is not transition dynamics of the Ramsey problem.\(^{47}\) Figure 15 reports the inferred Pareto weights and the optimal response of income tax policy with respect to only economic changes with this specification of the government budget constraint. The results are quite close to the benchmark case, and the economic changes account for about 41% of the reduction in progressivity in the data.

#### B.2.2 Alternative Objective Function of the Ramsey Problem

In the benchmark case, the objective function of the Ramsey problem is a weighted sum of lifetime utilities of each cohort at the stationary competitive equilibrium. An alternative way of measuring the welfare level at the stationary competitive equilibrium is to maximize the weighted sum of household flow utilities. With that measure, the objec-

\(^{47}\)On the other hand, the disadvantage of this specification is that it rules out the possibility of a pay-as-you-go social security system.
The objective function of the Ramsey problem is now

\[
\left\{ \sum_{j=1}^{T} \int u(c_j(s), \{l_{ij}(s), n_{ij}(s)\}_{i=1}^{2})W(A)d\Phi_j(s) \\
+ \sum_{j=T+1}^{T_R} \int u^R(c_j(s))W(A)d\Phi_j(s) \right\} + \gamma (T + T_R) \log \left( \frac{G}{T + T_R} \right).
\]

Figure 16 reports the main quantitative results with this objective function of the Ramsey problem. The results are overall similar to the benchmark case, and economic changes now account for about 72% of reduction in progressivity we observe. This larger reduction in progressivity in response to economic changes is because the flow utilities of old households are no longer discounted by the $\beta$’s, and they are more sensitive to the amount of human capital accumulation. So as the return to human capital investment increases, the benefit of a less progressive income tax to encourage human capital accumulation is larger with this specification of the objective function than in the benchmark case, and the optimal tax policy reflects this difference with a lower progressivity in the end.