The Effect of the Top Marginal Tax Rate on Top Income Inequality

Jihee Kim*
KAIST

November 13, 2015

Abstract

Top income inequality, defined as the income gap *within* the top 1% income group, has been rising in the United States since the 1980s. Coinciding with this rise, the large reductions in the top marginal tax rate exhibit a strong correlation with the increase in both top income inequality and top income shares. This paper identifies endogenous human capital accumulation as the link between changes in top marginal tax rates and increases in top income inequality. We develop an infinite-horizon, heterogeneous agent model, where human capital accumulation is endogenously characterized by a proportional random growth process which depends on the top marginal tax rate. If the top marginal tax rate decreases, the benefit of human capital investment will increase, thereby increasing the growth rate of human capital. Since this growth rate pins down the Pareto inequality measure of the top income distribution, a decrease in the top marginal tax rate will lead to a more unequal Pareto income distribution, while simultaneously increasing every top income. The calibrated model finds that the reduction of the top marginal tax rate from 70% to 40% can account for nearly two-thirds of the increase in top income inequality and 68.4% of the increase in the top 1% income share between 1980 and 2010.

*I am grateful to Chad Jones for his guidance and support throughout this research. I also thank Pete Klenow and John Weyant for many helpful comments and encouragement. I have benefited from discussions with Arthur Chiang, Juho Kim, Jeff Knott, Joo Eun Lee, Christos Makridis, Daniel Murray, Moojoong Ra, Francisco Santos and participants at the Stanford macro bag lunch seminar and KIPF. I gratefully acknowledge financial support from the Samsung Scholarship Foundation. The most recent version of this paper is posted at http://www.jiheekim.net

1. Introduction

While the widening income gap between the top 1% and the bottom 99% has been at the center of recent public debates in the United States, there has been less analysis of the concurrent widening income gap *within* the top 1%. From 1980 to 2010, during the same period that the income share[1] of the top 1% doubled from 8.18% to 17.42%, the income share of the top 0.1% more than tripled, rising dramatically from 2.23% to 7.50% (Alvaredo, Atkinson, Piketty and Saez (2012), see also Figure 1). Ironically, this widening gap within the top 1% income group, referred to hereafter as *top income inequality*, implies that while the top 1% has become better off over time as a group, a majority of the top 1% themselves may not feel much richer than before if they compare themselves to their upper income neighbors.

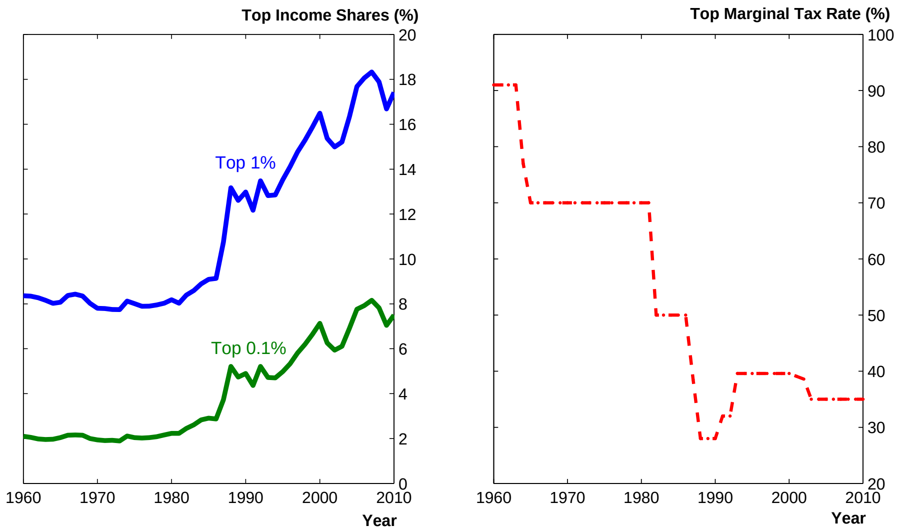
Such increases in top income shares and top income inequality coincided with large reductions in marginal income tax rates. In particular, the top marginal federal income tax rate, the marginal tax rate in the highest tax bracket, was 70% during 1970s until the rate was cut down to 50% in 1982. And after a few further changes, it is now 35% in 2012, only half of the 1980 rate (Figure 1). As most of the top 1% income group is subject to the top marginal tax rate[2], we may wonder whether the large reductions in the rate have had any impact on top incomes. Apparently, Figure 2 shows a strong correlation among the top income shares, top income inequality, and the top marginal tax rates in the United States. As emphasized by Piketty, Saez and Stantcheva (2011), a cross-country observation also tells the same story. This paper aims to help understand the correlation.

Regarding the effect of the changes in tax policy on the top income shares, the

[1]Income here excludes capital gains. It consists of wage and salary income (including bonus and stock-option exercises), entrepreneurial income (profits from S-Corporation, partnership, and sole proprietorship), dividends, and rental income. For details, see Piketty and Saez (2003).

[2]In the 1970s, there was sizable progressivity in the marginal federal tax rates even within the top 1% income group. For example, in 1975, the marginal federal tax rate applied to the top 1% income group ranged from 55% to 70%. However, the level of progressivity gradually decreased until 1982 when everyone in the top 1% income group was subject to the same top marginal federal tax rate of 50%. Since then, the top 1% income group has fallen into at most two tax brackets. Even in the case of the two tax brackets over the top 1% income group, the income threshold for the highest tax bracket was very close to the income threshold for the top 1%, implying the flat tax structure within the top 1% income group. For details, see Tax Foundation (2012).

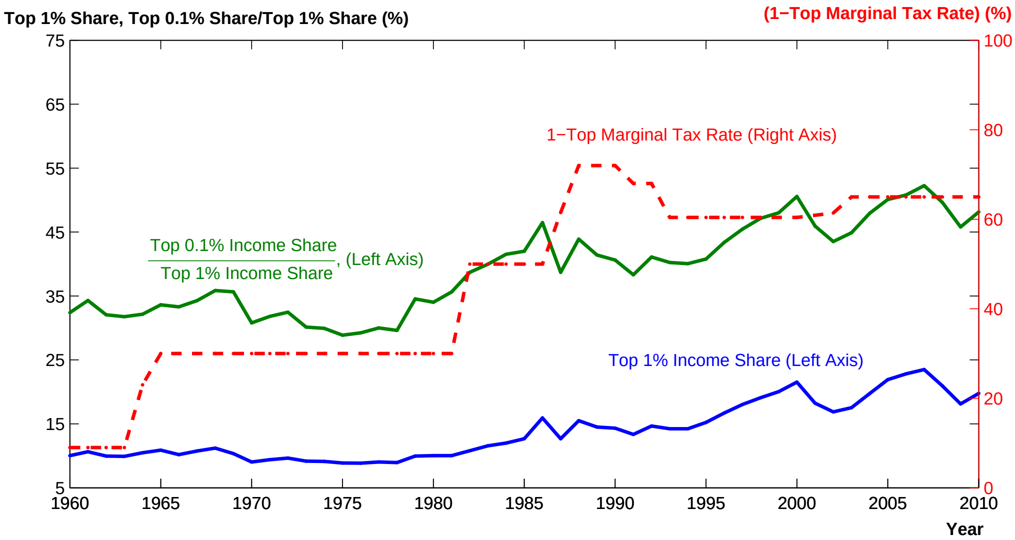
Figure 1: Top Income Shares and Top Marginal Tax Rates (Federal)



Note: The solid lines on the left panel show the income shares of the top 1% and the top 0.1% of taxpayers. For the description of income used in data, see footnote 1. Source: The World Top Incomes Database, The Tax Foundation

empirical literature in public economics has measured the response of top incomes to large cuts in the top marginal tax rate. For example, Lindsey (1987) and Feldstein (1995), using tax return data before and after the Economic Recovery Tax Act of 1981 and the Tax Reform Act of 1986, respectively, suggest substantial short-run responses of taxable income to the changes in marginal tax rates among the high-income taxpayers. Piketty, Saez and Stantcheva (2011) propose three channels of such responses: the standard labor supply response, the tax avoidance, and the compensation bargaining effort. However, because the main interest of this literature is the total tax revenue, these studies have few implications on top income inequality. The standard labor supply models imply that the top income inequality would remain the same while the top income share increases. Then is the top marginal tax rate only relevant to the top income shares, and not to the top income inequality? Or are we missing any channels through which the top marginal tax rate affects top income inequality?

Figure 2: Top Income Shares, Top Income Inequality, and Top Marginal Tax Rates

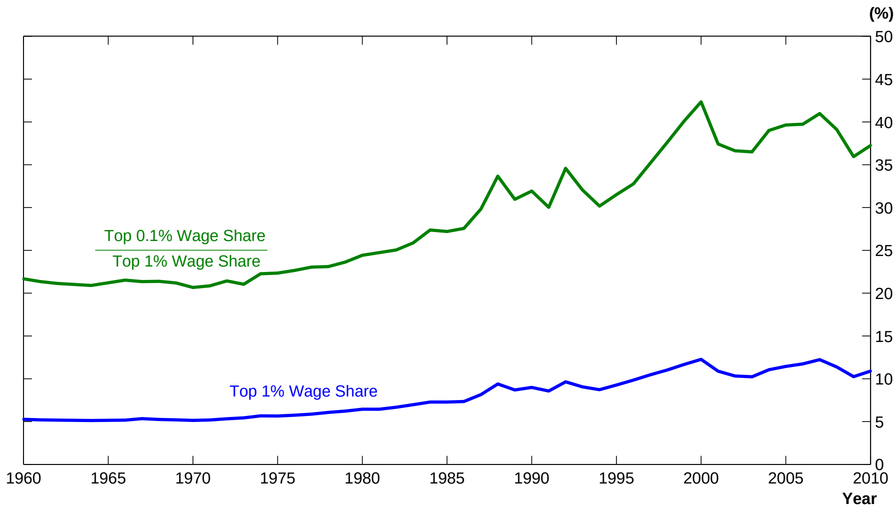


Source: The World Top Incomes Database, The Tax Foundation

This paper suggests that through the channel of endogenous human capital accumulation, a decline in the top marginal tax rate can lead to an increase in top income inequality as well as an increase in top income shares. The main model builds on the standard labor supply model, according to which top income shares increase through the channel of labor supply responses in the short-run. In addition to this, endogenous human capital will serve as a second channel through which the top income share increases in the long-run. More importantly, top income inequality will increase through this endogenous human capital channel. The calibrated model finds that the reduction of the top marginal tax rate from 70% to 35% can account for 46.3% of the increase in top income inequality and 41.2% of the increase in the top 1% income share between 1980 and 2010.

The key mechanism here is that human capital accumulation is endogenously characterized by a proportional random growth process, in which the growth rate depends on the top marginal tax rate. If the top marginal tax rate declines, the benefit of human capital accumulation effort will increase, thereby increasing the growth

Figure 3: Top Wage Shares and Top Wage Inequality



Note: The solid lines show the wage income shares of the top 1% and the top 0.1% of taxpayers, sorted by size of wage income. Wage income is wages, salaries, and tips on individual income tax form. It includes bonuses, and profits from exercised stock options. Source:Piketty and Saez (2003)

rate of human capital. Since this growth rate pins down the Pareto inequality measure of human capital distribution, the decline in the top marginal tax rate will lead to a more unequal Pareto distribution of human capital, while simultaneously increasing everyone's human capital.

One might wonder whether the top income share and inequality increases are mainly driven by income earned from accumulated wealth, not by income earned from working. If that is the case, the human capital channel may not provide a convincing answer. However, Figure 3 shows that the top wage shares and top wage inequality have also increased in the last three decades. Piketty and Saez (2003) find that the surge in top wages is the primary cause of the increase in top income shares over this period. Moreover, although the degree of inequality is greater with wealth than with income, wealth inequality has not been rising as much as income inequality (Kopczuk and Saez 2004).

This paper is closely related to Trostel (1993), who studies the effect of taxation on endogenous human capital accumulation. In a representative agent model with endogenous human capital, he shows that an increase in the tax rate has a negative effect on human capital accumulation. Our model can be seen as an extension of his model to the heterogeneous agents framework to study the distributional effect. Erosa and Koreshkova (2007) study the distributional effect of taxation through the channel of generational transfer of human capital. They find that income inequality, measured by the Gini index, is higher under progressive taxation than under linear taxation. This is because progressive taxation creates a disproportionate gap between the benefit and the cost of human capital investment over income distribution. Top income inequality, however, was not in the domain of discussion in their paper.

Apparently, taxation is not the sole factor in determining the top income dynamics. Gabaix and Landier (2008) argue that the large increase in firm sizes accounts for the rise of CEO compensations. Philippon and Reshef (2012) suggest that the expansion of the financial sector has contributed to the increase in the top wage share. However, these studies are silent on the rise in top income inequality. Atkinson (2003) discusses globalization and the superstar effect of Rosen (1981) to understand top income inequality. Jones and Kim (2014) present a model where the rise in entrepreneurial effort or its efficiency (which can be related to the superstar effect, the development of the world wide web, a decline in misallocatiion, etc.) can lead to a rise in top income inequality.

This paper proceeds as follows. The next section discusses some theoretical foundations on which the model in Section 3 is built. Section 3 then develops an infinite-horizon, heterogeneous agent model with endogenous human capital accumulation, which shows that lowering the top marginal tax rate will increase top income inequality as well as the top income share. Section 4 calibrates the model to match the top income distribution in 1980 and identify the quantitative effect of the reductions in the top marginal tax rate on the rises in top income inequality and top income share over the period 1980-2010. Section 5 presents some work-in-progress extensions of the infinite-horizon model, and Section 6 concludes.

2. Theoretical Background

In this section, we discuss some theoretical foundations on which our model in Section 3 is built. First, we introduce a measure of top income inequality, which is derived from the fact that the top income distribution is characterized by a Pareto distribution. Next, we show why the standard labor responses to the marginal tax rate changes cannot be a source of the rise in top income inequality. This leads us to wonder whether the endogenously evolving human capital can be one of the sources. Interestingly, human capital accumulation can be characterized by a Pareto-generating proportional random growth process, in which the growth rate of human capital accumulation will pin down the Pareto measure of income inequality. This implies that if the marginal tax rate affects the growth rate of human capital, then it will also affect income inequality. To see when this is such a case, we study several different human capital technologies in the last subsection.

2.1 Pareto Top Income Distribution

It is well known that the income distribution at the upper tail is well approximated by the Pareto distribution, especially above the top 1% income threshold (Saez 2001). If the top income Y follows the Pareto distribution, then for $y \geq y_{\min}$,

$$\Pr\left[Y > y\right] = \left(\frac{y}{y_{\min}}\right)^{-\xi},$$

where ξ is called the "power law exponent," and $y_{\min}$ is the top 1% income threshold.

An important property of the Pareto distribution is the following fractal nature. Let $y_{x\%}$ denote the top $x\%$ income threshold so that we have $y_{x\%} = x^{-\frac{1}{\xi}} y_{\min}$ for $x \leq 1$. Then we see that the top 0.1% earner makes $10^{\frac{1}{\xi}}$ times more than the top 1% earner, $y_{0.1\%} = 10^{\frac{1}{\xi}} y_{1\%}$. And the same factor of $10^{\frac{1}{\xi}}$ applies between the top 0.01% income threshold and the top 0.1% income threshold. The same pattern continues further along the upper tail of the top income distribution. In other words, wherever you stand in the Pareto income distribution, you always find a person at the top 10% of the people in front of you making $10^{\frac{1}{\xi}}$ times more than you. Top income shares show the same fractal pattern: the top 0.1% income share is $\left(\frac{1}{10}\right)^{1-\frac{1}{\xi}}$ of the top 1% income share, and the same factor applies between the top 0.01% income share and the top 0.1% income share, and so forth.[3]

Since the inequality of the Pareto distirubtion increases in $\frac{1}{\xi}$, it is convenient to define "power law inequality exponent η" as

$$\eta \equiv \frac{1}{\xi},$$

which we will use as the measure of top income inequality. Rephrasing the increase in top income inequality with this measure, η increased from 0.436 in 1980 to 0.613 in 2010[4], which is about a 40% increase. That is, the top 0.1% earner made 2.73 times more in 1980, but 4.10 times more in 2010 than the top 1% earner. Figure 4 shows these changes in the power law inequality exponent η over the last three decades in the United States

Furthermore, assuming that the top 1% income threshold does not change[5], an increase in η will lead to an increase in the average top income since the average of the Pareto distribution is given by

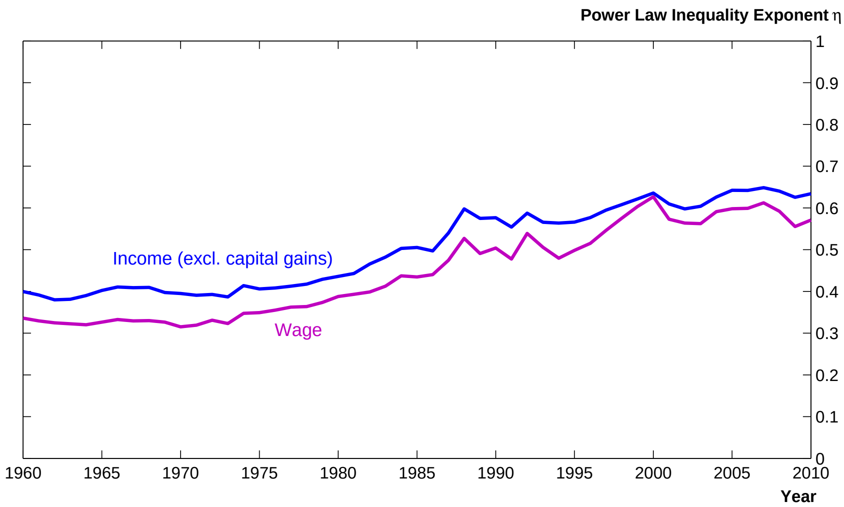
$$\mathbf{E}[Y] = \left(\frac{1}{1-\eta}\right) y_{\min}. \tag{1}$$

This is because an increase in η will increase every top income while the factor of the increase itself increases in income. Hence, what changes η will have a level effect as well as a distributional effect.

Lastly, the following property of the Pareto distribution will be useful in later sections of the paper.[6] If Y follows a Pareto distribution with the power law inequality exponent η_Y, then $Z \equiv Y^\alpha$ $(\alpha > 0)$ also follows a Pareto distribution and its power

[3]See Jones and Kim (2014) for more details.
[4]$\eta = 1 - \log_{10}\left(\frac{\text{Top 1\% Share}}{\text{Top 0.1\% Share}}\right)$
[5]We may think of it as a normalization of top incomes relative to $y_{\min}$.
[6]See Gabaix (2009) for more properties of the Pareto distribution.

Figure 4: Top Income/Wage Inequality Trends in the Power Law Inequality Exponent



Note: The power law inequality exponents are calculated from the top shares data in Piketty and Saez (2003)

law inequality exponent is given by

$$\eta_Z = \alpha \eta_Y . \tag{2}$$

2.2 The Mirrlees Model

This section presents a simple version of the standard labor supply model by Mirrlees (1971) to see why the standard labor responses to a change in the top marginal tax rate does not affect the power law inequality exponent of the top income distribution. In this model, individuals are heterogenous in their skill level n, which measures their marginal productivity of labor effort l_n. Then an individual with the skill level n earns income $y_n = nl_n$. They are also subject to the tax liability $T(y_n)$. For simplicity, suppose there is linear utility from the take-home income $y_n - T(y_n)$ and the isoelastic disutility from the labor effort. We show later how to generalize to concave utility in consumption. We will assume away utility from government transfer programs because high-income individuals are of our main interest. Thus, each individual solves the following problem

$$\max_{l_n} nl_n - T(nl_n) - \frac{l_n^{1+\kappa}}{1+\kappa}.$$

The optimal labor decision of an individual with skill level n is then

$$l_n = (n(1 - T'(nl_n)))^{\frac{1}{\kappa}} = (n(1-\tau))^{\frac{1}{\kappa}},$$

where τ is the top marginal tax rate. The last equality comes from the fact that the tax structure within the top 1% income group is flat. See the discussion on the footnote 2. Therefore we obtain the income y_n as

$$y_n = nl_n = n^{1+\frac{1}{\kappa}}(1-\tau)^{\frac{1}{\kappa}}. \tag{3}$$

A decrease in the top marginal tax rate will raise everyone's income by the same factor through the increased labor effort, which we refer to as the standard labor response channel. These same proportional increases will leave income inequality unchanged. Let's then see how the income inequality is determined in this model. To make y_n Pareto distributed as it is in the top 1% income group, the distribution of the skill n must follow the Pareto distribution. Let η_n denote the power law inequality exponent of the skill distribution. By applying the property (2) to equation (3), the power law inequality exponent of the income distribution is given by

$$\eta_y = \left(1 + \frac{1}{\kappa}\right)\eta_n, \tag{4}$$

which does not depend on τ. This non-effect of the marginal tax rate on inequality holds under more general utility functions. For example, under the standard isoelastic preference $u(c,l) = \frac{c^{1-\sigma}}{1-\sigma} - \frac{l^{1+\kappa}}{1+\kappa}$, the power law inequality exponent of the income distribution is given by

$$\eta_y = \left(\frac{\kappa+1}{\kappa+\sigma}\right)\eta_n,$$

which again does not depend on τ. [7]

This leads us to look at η_n in (4), the inequality of the skill distribution. If the rise in η_y cannot be explained by the standard labor responses to marginal tax rate changes, then it should come from changes in η_n, the inequality of the skill distribution. What skill level n means in this model is the effective wage rate of individual's unit labor effort, which can be well affected by taxation if we let individuals invest in it. In this setting, n can be thought of as human capital, which can endogenously evolve over time. This is where the human capital production functions in Lucas (1988) enter into the picture.

2.3 Lucas (1988) Meets a Pareto Generating Process

In Lucas (1988), the human capital stock $h(t)$ at time t grows according to

$$\dot{h}(t) = \psi f(t)h(t) \tag{5}$$

for some constant $\psi > 0$, where $f(t)$ is the effort the individual expends to accumulate human capital. $f(t)$ is either $1 - l(t)$ (learning by learning) or $l(t)$ (learning by doing), given that $l(t)$ is the fraction of time spent at work. The discrete time version of the human capital production (5) is then

$$h_{t+1} = (1 + \psi f_t)h_t. \tag{6}$$

This form of the law of motion reminds us of the proportional random growth process, one of the commonly used mechanisms to generate a Pareto distribution.[8] That is, if the human capital stock h_t obeys the following process

$$h_{t+1} = \gamma_t h_t \tag{7}$$

[7]Saez (1999) shows that income inequality remain unchanged after the tax rate change under more general preferences using the concepts of uncompensated and compensated elasticities of taxable income with respect to the net-of-tax rate.

[8]The random growth theory, especially the proportional random growth process and the Kesten process have been widely used to explain wealth distribution (Nirei and Souma (2007), Benhabib, Bisin and Zhu (2011)) and the firm size distribution – Zipf's law (Gabaix 2009)).

where $\gamma_t > 0$ is a stochastic process with $\mathbf{E}[\gamma_t] < \infty$ and there exists $\lambda > 0$ such that $\mathbf{E}[\gamma_t^\lambda] = 1$, then the stationary distribution of $h(t)$ (if it exists) is a Pareto distribution with the power law inequality exponent $\frac{1}{\lambda}$. See Levy and Solomon (1996) for details.

This implies that if $(1 + \psi f_t)$ in equation (6) is characterized by some stochastic process, we may endogenously obtain a Pareto distribution for human capital. Furthermore, if f_t is a function of the marginal tax rate, we will later see a change in the marginal tax rate will affect the power law inequality exponent of the endogenous human capital distribution. Since individual income is a function of human capital, a Pareto distribution for human capital will lead to a Pareto income distribution, and the power law inequality exponent will depend on the marginal tax rate.

Our last task is then to find some link from the marginal tax rate to f_t and add some randomness to it to make it stochastic. Before turning to that task, we need to make sure that the stationary distribution of $h(t)$ exists. To achieve this, we assume that the level of human capital cannot go below some $h_{\text{min}} > 0$. Technically this reflective barrier ensures that the variance does not grow over time. See Gabaix (1999), Gabaix (2009) for details. Adding this condition, the law of motion of h_t will be in the following form

$$h_{t+1} = \max \{\gamma_t h_t, h_{\text{min}}\}. \tag{8}$$

This minimum level of human capital not only serves as a technically necessary condition, but it also has an interpretation on the flow of top incomes. We can think of h_{min} as the minimum level of human capital required to enter the top 1% income group. That is, the human capital of a person at the top 1% income threshold. If someone in the top 1% income group is hit by a low shock and moves down to the lower income group, then his or her place in the top income group will be replaced by some other person who moves up the income ladder from the bottom 99%, with the starting human capital h_{min}.

2.4 The Effect of Taxation on Endogenous Human Capital

We now explore three different human capital technologies to find a link between the tax rate and the amount of effort that people put into accumulating human capital. In other words, we will look for some f_t in equation (6) that depends on the marginal tax rate.

Throughout this subsection we will assume that individuals live for two periods and they take the first period human capital h_1 as given. The models we see below differ in the way that the second period's human capital h_2 is produced. An individual with human capital h will earn the income $y = hl$ where l is the labor effort. Individuals face the linear tax liability $T(y) = \tau y$ on income y. For simplicity, we assume that there is neither government transfer nor savings technology, so that the consumption c equals the take-home income $y - T(y)$. An isoelastic utility $u(c) = \dfrac{c^{1-\sigma}}{1-\sigma}$ is assumed. Lastly, β is a discount rate and we use the subscripts 1 and 2 for the two time periods.

2.4.1 Learning-by-Learning

In the first model, the labor effort l is the fraction of the time spent at work, and the rest of the time $1 - l$ is spent in learning to produce the second period's human capital h_2. Thus, h_2 is given by the following learning-by-learning technology:

$$h_2 = \psi h_1 (1 - l), \tag{9}$$

for some $\psi > 0$. Individuals spend all their time at work in the second period. Each individual will then solve

$$\max_l u((1-\tau)h_1 l) + \beta u((1-\tau)h_2)$$

subject to equation (9). The optimal learning decision is then made by

$$[(1-\tau)h_1]^{1-\sigma} l^{-\sigma} = \beta[(1-\tau)h_1\psi]^{1-\sigma}(1-l)^{-\sigma}. \tag{10}$$

The left-hand side of equation (10) is the marginal cost of the time spent in learning – the marginal utility of the foregone first period consumption. And the right-hand side is the marginal benefit of the time spent in learning – the marginal utility from the second period consumption. We can see that $1-\tau$ has the same effect on the cost and on the benefit of the time investment in learning. Therefore, there will be no effect of taxation on the human capital h_2. This learning-by-learning human capital accumulation model assumes that you have to give up today's work to accumulate human capital. However, because on-the-job experience is highly valued in the high-income professions such as doctors, lawyers, or CEOs, learning-by-doing appears to be more relevant than learning-by-learning when studying top incomes. Now we move on to the learning-by-doing human capital production.

2.4.2 Learning-by-Doing

In this second model, the second period's human capital h_2 is produced by l, the fraction of the time spent at work. Then h_2 is given by the following learning-by-doing technology:

$$h_2 = \psi h_1 l, \tag{11}$$

for some $\psi > 0$. This time we introduce utility from leisure $v(1-l)$, which decreases in the labor effort l. Each individual will then solve

$$\max_{l_1,l_2} u((1-\tau)h_1 l_1) + v(1-l_1) + \beta\left[u((1-\tau)h_2 l_2) + v(1-l_2)\right]$$

subject to equation (11). The optimal learning decision is made by

$$[(1-\tau)h_1]^{1-\sigma}\left\{l_1^{-\sigma} + \beta(\psi l_2)^{1-\sigma} l_1^{-\sigma}\right\} = v'(1-l_1). \tag{12}$$

The left-hand side of (12) is the marginal benefit of the time spent in working-and-learning – the marginal utility from the first and the second period consumption. And the right-hand side is the marginal cost of the time spent in working-and-learning – the marginal utility of the foregone first period leisure. We can see that an increase in $1-\tau$ will raise the benefit while the cost remains unaffected. Therefore, the second period human capital h_2 will be a function of $1-\tau$.

2.4.3 Learning-by-Spending

Before turning to the last model, let's discuss what the human capital h represents in the context of top incomes. It surely has a lot broader meaning than just schooling. It would include one's experience both inside and outside of work, established social capital, social status, health, etc. To maintain or raise the level of human capital in this manner, some money investment seems necessary, such as in executive education, hosting or attending social events, country club memberships, high-quality health services, etc. Therefore, the time effort as in learning-by-learning or learning-by-doing is not the only input in the human capital production, and investment in goods can be another input. The last model will capture this point.

The human capital accumulation in the last model takes the investment in goods as an input. h_2 is given by the following learning-by-spending technology

$$h_2 = \psi h_1 e, \tag{13}$$

for some $\psi > 0$ where e is the investment spending in the human capital. Thus, the consumption c in this model equals $y - T(y) - e$. It is a Ben-Porath (1967) type human capital production, but without the time component. We again have the utility from leisure $v(1-l)$. Each individual will then solve

$$\max_{e,l_1,l_2} u((1-\tau)h_1 l_1 - e) + v(1-l_1) + \beta\left[u((1-\tau)h_2 l_2) + v(1-l_2)\right]$$

subject to equation (13). The optimal decision in the human capital investment e is made by

$$[(1-\tau)h_1 l_1 - e]^{-\sigma} = \beta[(1-\tau)\psi h_1 l_2]^{1-\sigma} e^{-\sigma}. \tag{14}$$

The left-hand side of equation (14) is the marginal cost of the investment in human capital goods – the marginal utility of the foregone consumption. And the right-hand side is the marginal benefit of the investment in human capital goods – the marginal utility from the second period consumption. We see that a increase in $1-\tau$ reduces the cost and raises the benefit, thereby giving us the higher incentive to invest in human capital. When we earn more thanks to a reduction in the marginal tax rate, we have more money to spare in the human capital investment while the reduction makes the return on the human capital investment more attractive.

2.5 Discussion

We now summarize what we have discussed in this section. We first defined the "power law inequality exponent", a measure of top income inequality derived from the fact that top 1% income distribution is a Pareto distribution. Then we learned a lesson from the standard labor supply model that labor responses cannot explain the rise in top income inequality. This led us to ponder over the endogenous human capital. Surprisingly, human capital growth in Lucas (1988) has the same structure as the Pareto-generating proportional random growth process, just without some randomness in it. If human capital evolves as a proportional random growth process, the growth rate of human capital pins down the power law inequality exponent of the human capital distribution. Therefore, if the top marginal tax rate affects the growth of human capital accumulation, then it will also affect the inequality of the human capital distribution. However, not all human capital accumulation decision is affected by the top marginal tax rate. Among the three different human capital technologies, learning-by-doing and learning-by-spending decisions were affected by the top marginal tax rate. Since human capital accumulation will require all three technologies in the real world, having at least two technologies influenced by the top marginal tax rate implies that the top marginal tax rate can actually be an important factor in human capital accumulation, therefore affecting the top income inequality.

3. Infinite-horizon with Endogenous Human Capital

In this section, we develop an infinite-horizon heterogeneous agents model to see how the learning-by-spending human capital accumulation in Section 2.4.3 leads to a Pareto income distribution, of which the power law inequality exponent depends on the marginal tax rate. We choose the learning-by-spending human capital accumulation since it will enable us to see the relationship between the marginal tax rate and the top income inequality in a closed-form. Either the learning-by-doing or the learning-by-spending human capital accumulation, or some combination of the two should also give us the qualitatively similar result.

This model extends Section 2.4.3 to the infinite-horizon. Infinitely-lived individuals are heterogenous in their human capital level h. An individual with human capital h will earn income $y = hl$ where l is labor effort. We now treat l as effective work effort rather than work hours because the work hours would be a less important factor in top incomes. Individuals face the linear tax liability function $T(y) = \tau y$, $0 < \tau < 1$ on income y. For simplicity, we again assume that there is neither government transfer nor savings technology. Lastly, β is a discount rate and we use subscripts for time periods.

We assume that the flow utility takes the following form:

$$u(c_t, l_t) = c_t - \frac{1}{\rho}\frac{l_t^{1+\kappa}}{1+\kappa},$$

where ρ and κ are positive. The first term is the linear utility from consumption c_t. This linearity assumption is to obtain the closed-form solution. The standard iso-elastic utility function will deliver the same qualitative result. The second term is the disutility from the labor effort l_t. Note that under the static environment of the standard labor supply model, $\frac{1}{\kappa}$ is interpreted as the uncompensated elasticity of labor effort with respect to the marginal net-of-tax rate $(1-\tau)$.

The law of motion for human capital is given by

$$h_{t+1} = \epsilon_t h_t^{\alpha} e_t^{\gamma}, \tag{15}$$

where α and γ are positive and e_t is the goods investment in human capital, which is measured in the consumption unit. We assume that h_t cannot go below some $h_{\min} > 0$. We introduce a multiplicative idiosyncratic i.i.d. shock $\epsilon_t > 0$ such that $\mathbf{E}[\epsilon_t] < \infty$ in the human capital accumulation. This captures uncertainty in the returns to human capital accumulation, making human capital risky under no insurance channel. Technically it will enable us to apply the result from the proportional random growth process to generate a Pareto distribution. Lastly, we restrict the parameters to satisfy

$$\alpha + \gamma\left(1 + \frac{1}{\kappa}\right) = 1 \tag{16}$$

to ensure the stability of the human capital distribution as in Lucas (1988). Individuals then solve

$$\max_{\{c_t, l_t, e_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t, l_t)$$

subject to equation (15), the budget constraint

$$c_t + e_t = (1-\tau)h_t l_t,$$

and the constraint on the positivity of consumption, $c_t > 0$ for $\forall t \in \{1, 2, 3, ..., \infty\}$.

It is convenient to express this problem in the form of a Bellman equation. Let $V(h)$ denote the value function of an individual with the current level of human capital h. Then the Bellman equation is

$$V(h) = \max_{c,l,e} u(c,l) + \beta E[V(h')] \tag{17}$$

subject to

$$\begin{aligned} &c + e = (1-\tau)hl, \\ &h' = \epsilon h^{\alpha} e^{\gamma}, \\ &c > 0, \end{aligned}$$

where h' denotes the level of the next period's human capital.

This dynamic programming problem has the following closed-form solution for the policy functions $l(h)$ and $e(h)$ and the value function $V(h)$:

$$l(h) = (\rho(1-\tau)h)^{\frac{1}{\kappa}}, \tag{18}$$
$$e(h) = \left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]X\right)^{\frac{1}{\alpha}} h^{1+\frac{1}{\kappa}}, \tag{19}$$
$$V(h) = Xh^{1+\frac{1}{\kappa}}, \tag{20}$$

where X is a solution of

$$\begin{aligned} X &= \frac{\alpha}{1-\alpha}\left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]\right)^{\frac{1}{\alpha}} X^{\frac{1}{\alpha}} + \frac{\kappa}{1+\kappa}\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}, \\ 0 < X &< \left(\frac{1-\alpha}{\alpha}\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}\right)^{\alpha} / \left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]\right). \end{aligned} \tag{21}$$

We can see that X is a function of τ. However, without any closed-form solution, it is not easy to see whether X has an increasing or decreasing relationship with τ. The following lemma will tell us when X is increasing in $(1-\tau)$ (proofs for this lemma and the remaining propositions are given in the appendix).

Lemma 1 (Value Function in the Infinite-Horizon Model): *If*

$$\beta\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}][\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}]^{1-\alpha} < \frac{\kappa+1}{\kappa+\alpha}, \tag{22}$$

then the value function $V(h)$ is uniquely determined and increasing in the take-home rate $(1-\tau)$. In other words, there exists a unique X with $\frac{\partial X}{\partial(1-\tau)} > 0$.

This lemma implies that if (22) is satisfied, the value function $V(h)$ will increase in the marginal take-home rate $(1-\tau)$.

Now we look at the growth of individual human capital and income. Plugging the policy function (19) into the human capital accumulation (15), we obtain the following individual human capital growth:

$$h' = \max\left\{\epsilon\left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]X\right)^{\frac{\gamma}{\alpha}} h, h_{\min}\right\}. \tag{23}$$

Finally, from (18) and (23) we can derive the following growth process for income:

$$y' = \begin{cases} \epsilon^{1+\frac{1}{\kappa}} \left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]X\right)^{\frac{\gamma}{\alpha}(1+\frac{1}{\kappa})} y & \text{if } h' > h_{\text{min}} \\ (\rho(1-\tau))^{\frac{1}{\kappa}} h_{\text{min}}^{1+\frac{1}{\kappa}} & \text{if } h' = h_{\text{min}} \end{cases} \tag{24}$$

Both the human capital and income growth processes are exactly in the same form as the proportional random growth process that we have seen in Section 2.3. Therefore, both human capital and income will have a stationary Pareto distribution as we show in the next proposition.

Proposition 1 (Power Law Inequality in the Infinite-Horizon Model): *If there exists $\eta_y > 0$ such that*

$$\mathbf{E}\left[\left\{\epsilon^{1+\frac{1}{\kappa}} \left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]X\right)^{\frac{\gamma}{\alpha}(1+\frac{1}{\kappa})}\right\}^{\frac{1}{\eta_y}}\right] = 1, \tag{25}$$

then top income y has a stationary distribution whose tail follows a Pareto distribution, of which the power law inequality exponent is η_y. The power law inequality exponent of human capital is then given by $\eta_h = \eta_y / \left(1 + \frac{1}{\kappa}\right)$.

Furthermore, if

$$\beta\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}][\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}]^{1-\alpha} < \frac{\kappa+1}{\kappa+\alpha},$$

then an increase in the take-home rate $(1-\tau)$ will raise η_y and η_h as long as η_y and η_h exist on $(0,1)$.

This proposition summarizes what we expected from the previous discussion in Section 2. A reduction in the marginal tax rate will increase human capital investment, then it will push human capital risk upward, and the final effect of this will show up as a more unequal Pareto distribution. If we further assume that the human capital shock is log-normal, this positive relationship between the marginal take-home rate $(1-\tau)$ and the power law inequality exponent η_y is available to see in a closed-form equation. We show this in the following proposition.

Proposition 2 (Power Law Inequality under the Log-Normal Shock): *If $\log\epsilon$ follows a normal distribution with the mean $-\sigma^2/2$ and the variance σ^2, then the power law inequality exponents η_y and η_h are explicitly given by*

$$\frac{1}{\eta_y} = \frac{\kappa}{1+\kappa}\left(1 - \frac{\gamma}{\alpha}\frac{\log\left(\beta(1-\alpha)X\right) + \left(1+\frac{1}{\kappa}\right)\sigma^2/(2\kappa)}{\sigma^2/2}\right), \tag{26}$$

$$\frac{1}{\eta_h} = 1 - \frac{\gamma}{\alpha}\frac{\log\left(\beta(1-\alpha)X\right) + \left(1+\frac{1}{\kappa}\right)\sigma^2/(2\kappa)}{\sigma^2/2}.$$

Furthermore, if $\beta\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}][\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}]^{1-\alpha} < \frac{\kappa+1}{\kappa+\alpha}$, then an increase in the take-home rate $(1-\tau)$ will raise η_y and η_h since there exists a unique X with $\frac{\partial X}{\partial(1-\tau)} > 0$.

It is easy to see now that an increase in X will lead to an increase in the power law inequality exponent η_y. As we will see in the next section, η_y and η_h are less than 1 under the reasonable set of parameter values.[9]

As we briefly discussed before, an increase in η has two effects on the income distribution: the distributional effect and the level effect. This implies that an increase in $(1-\tau)$ will have the same two effects. The distributional effect comes directly from the increase in η. To see the level effect , let's compare the two steady states, one with the low marginal take-home rate $(1-\tau)_L$ and therefore low η_L, and the other state with high $(1-\tau)_H$ and high η_H. And suppose that the two states have the same level of human capital of the top 1% income threshold, which we denote h_{min}. From the model above, the top $x\%$ income is given by $y_{x\%} = (\rho(1-\tau))^{\frac{1}{\kappa}} h_{x\%}^{1+\frac{1}{\kappa}}$, where $h_{x\%} = x^{-\eta_h} h_{\text{min}}$ for $x \leq 1$. Then we can see that $y_{x\%}$ increases in both $(1-\tau)$ and η_h. That is, an increase in $(1-\tau)$ will raise everyone's income through the two channels: the standard labor response channel in $(1-\tau)^{\frac{1}{\kappa}}$ and the human capital channel in $x^{-\eta_h}$.

Lastly, we make a quick digression to discuss the implications of this model on the outflow of top incomes to the lower income level. We define $\gamma \equiv \epsilon\left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]X\right)^{\frac{\gamma}{\alpha}}$ so that we can write $h' = \gamma h$. Then the fraction of the top incomes that will fall down to the lower income group is

[9] If η is greater than 1, the Pareto distribution does not have a finite mean.

outflow from top incomes

$$
\begin{aligned}
&= \Pr[h' < h_{\min}] \\
&= 1 - \Pr[h' \geq h_{\min}] \\
&= 1 - \Pr[h > \frac{h_{\min}}{\gamma}] \\
&= 1 - \mathbf{E}[\min\{1, \gamma^{\frac{1}{\eta_h}}\}].
\end{aligned}
$$

Only the low shocks such as $\gamma < 1$ will determine this outflow. If $(1 - \tau)$ increases, then it will move the shock γ upward, raising some of the low shocks to be greater than 1, resulting in a decrease in the outflow probability. In addition to this, the increased inequality will reduce this fraction further down due to the factor $\frac{1}{\eta_h}$. Thus, a reduction in the top marginal tax rate will make the top incomes less likely to fall out of the top income group.

4. Quantitative Analysis

4.1 Calibration

To illustrate the quantitative significance of the endogenous human capital channel, we calibrate the infinite-horizon model in the context of Proposition 2, assuming that ϵ is log-normal with parameters $(-\sigma^2/2, \sigma^2)$. We suppose that the top income distribution in 1980 was at a steady state under the top marginal tax rate of 70%. This is a reasonable assumption since both the top marginal tax rate and the power law inequality exponent were relatively constant during the 1970s. Then we calibrate our model parameters to match the power law inequality exponent $\eta_{1980} = 0.4359$ when $\tau = 0.7$.

We first calibrate κ, the parameter governing the disutility from labor effort, to match the empirical estimate of the elasticity of taxable income with respect to the marginal net-of-tax rate in Lindsey (1987). Lindsey (1987) estimated the elasticity using the changes in the baseline income distribution for 1980-1981 upon the marginal tax rate change in 1982. Although his estimation includes responses from the two channels in our model – the standard labor responses and the increase in human capital, we will assume that the increase in the top 1% income threshold from 1980 to 1982 comes only from the labor response. That is, the human capital of the top 1% income threshold $h_{\min}$ remains unchanged so that the increase in the top 1% income threshold $y_{1\%}$, given by $(\rho(1-\tau))^{\frac{1}{\kappa}} h_{\min}^{1+\frac{1}{\kappa}}$, comes from the change in $(1-\tau)^{\frac{1}{\kappa}}$, but not from the change in $h_{\min}^{1+\frac{1}{\kappa}}$. It implies that our analysis under this assumption will give the lower bound for the effect of the human capital channel on the income increase by overestimating the standard labor response channel. Furthermore, the possible underestimation in the effect of the human capital channel will be modest considering the fact that the extent of an increase in human capital is lowest for the top 1% income threshold within the top 1% income distribution. This is because that the factor of the increase in human capital rises with income. Thus, we match the elasticity of taxable income estimate 0.6522 for the top 1% income threshold [10] to $1/\kappa$. The elasticity is about the half of the estimate for the top incomes in the highest tax bracket in Feldstein (1995). We will later show that the other half of the elasticity estimate will come from the human capital channel.

For the variance of the log shock, we use the estimate of the variance of 1-year log earning differentials from Huggett, Ventura and Yaron (2011). Their estimate is 0.0876 across the population.[11] We assume that the standard deviation of the top incomes' log earning differential is twice the size of that of the general population. This assumption is based on Carnevale, Rose and Cheah (2011), who document that the lifetime earning variation is about 2-4 times higher for bachelor's or higher degrees holders than for high-school graduates. From $Var(\log(y'/y)) = (1+\frac{1}{\kappa})^2 Var(\log\epsilon)$, we take $\sigma^2 = Var(\log\epsilon) = 0.1539$.

ρ, the parameter governing the weight on the disutility of labor effort, is chosen to match the top 1% income threshold in 1980, which is \$ $192,253$ in 2010 dollars[12]. We first normalize the level of human capital so that $h_{\min} = 1$, then we use $y_{1\%} =$

[10]We calculate the elasticity for the top 1% income threshold from the equation for the federal tax rate in Table 6 of Lindsey (1987). Table 6 is constructed assuming the elasticity rises with natural log of income.

[11]We take the estimate for PSID sample between age 23 and age 60 in the period 1969-2004 from Huggett, Ventura and Yaron (2011).

[12]The source is the 2010 updated data from Piketty and Saez (2003)

Table 1: Calibrated Parameter Values

Parameter Value	Assumptions/Comments
$\kappa = 1.5327$	to match est. of elasticity of top 1% income threshold in Lindsey (1987)
$\alpha = 0.93$	to match η in 1980
$\gamma = 0.0424$	from the parameter restriction (16)
$\beta = 0.9957,$	$1/(1+r)$, where r is the real effective federal funds rate in 1971-1980
$\sigma^2 = 0.1539$	std(1-yr Δ(log earning)) $\approx$ twice the estimate in the population
$\rho = 0.266$	to match the top 1% income threshold in 1980

$(\rho(1-\tau))^{\frac{1}{\kappa}} h_{\text{min}}^{1+\frac{1}{\kappa}} = 0.192253$ to obtain $\rho = 0.266$.

Next, we calibrate α to match the power law inequality exponent $\eta_{1980} = 0.4359$ in 1980. Applying calibrated parameter values above to Proposition 2, the model estimate $\tilde{\eta}_{1980}$ becomes 0.4359 when $\alpha = 0.93$. This implies that 93% of the current level of log human capital will be directly preserved to the next period's level of log human capital $(\log h' = \alpha \log h + \gamma \log e + \log \epsilon)$. Then κ and α will pin down γ, the parameter governing the weight of e in human capital accumulation, to be 0.0424 from the parameter restriction (16).

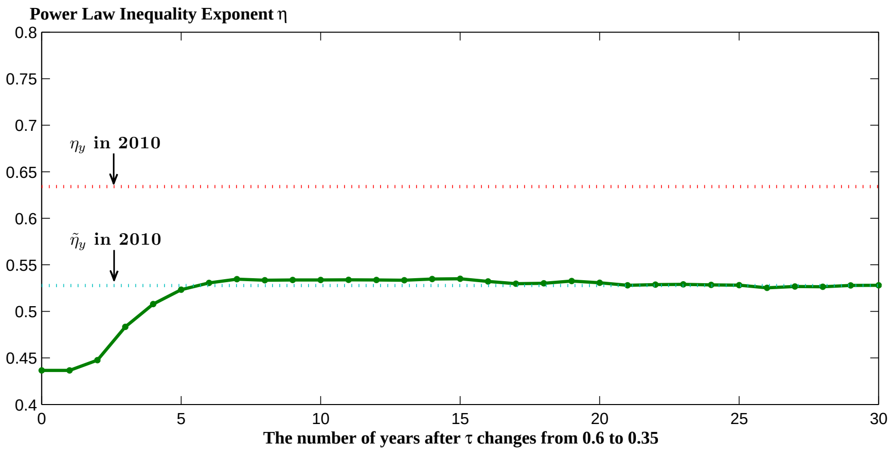
Lastly, we calibrate the discount factor β to 0.9957 so that β is equal to $1/(1+r)$, where $r = 0.00434$ is the average real federal funds rate (annual) from the period 1971-1980. The calibration of the model parameters is summarized in Table 1.

4.2 Tax Regime Change

Going back to Figure 1 or Figure 2, we notice a regime change in the top marginal tax rate around 1982. We were in a high-tax regime with the top marginal tax rate of 70% until the rate was cut down to 50% in 1982. And a few further changes were made since 1982, entering the era of a low-tax regime with the top marginal tax rate in the range of 28 40%. We estimate the effect of this regime change in this subsection.

From the set of parameters calibrated to the 1980 level of top income inequality, we will see how top incomes change when the top marginal tax rate is cut down

Figure 5: Transition to the New Steady State After the Tax Cut



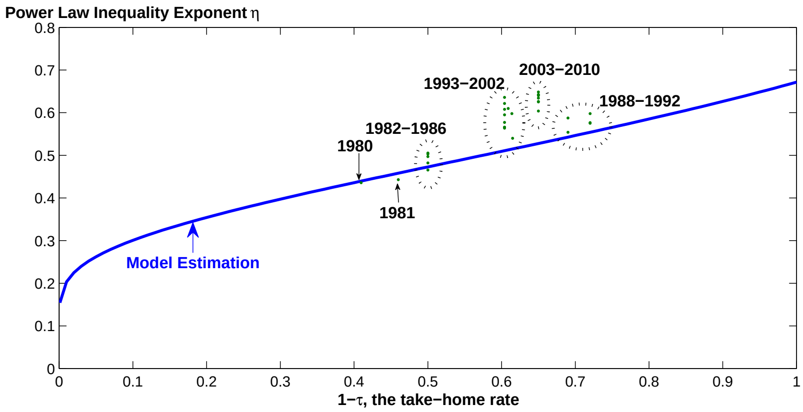
Note: The left panel shows the changes in the top income distribution in response to the reduction in the top marginal tax rate from 70% to 40% (The slope indicates the reciprocal of the power law inequality exponent). The right panel shows the power law inequality exponent trajectory after this change.

to 40% from 70%. We suppose that people in the steady state under the high-tax regime in 1980 will reoptimize assuming that this tax cut is permanent, and that they will be at the new steady state under the low-tax regime in 2010. Proposition 2 suggests that the power law inequality exponent rises to $\tilde{\eta}_{2010} = 0.5665$ in the new steady state under the low-tax regime with the top marginal tax rate of 40%. It is a 30% increase from $\eta_y^{1980} = 0.4359$. The power law inequality exponent of the U.S in 2010 was $\eta_y^{2010} = 0.6341$, a 45.5% increase from 1980. Thus, our model generates about 65.9% of the increase in the top income inequality.

Next, we discuss the speed of convergence to the new low-tax steady state from the high-tax steady state. We first simulate one million top incomes assumed to be in a steady state in 1980 to respond to the top marginal tax rate cut to 40%. Figure 5 shows the transition dynamics of the top income distribution upon this change. After about 7 years[13] the top income distribution converges to a new steady state income distribution, of which the power law inequality exponent is about $\tilde{\eta}_{2010} = 0.5665$, the theoretical estimate. Figure 6 shows the theoretical estimates

[13]We suppose that 1 year is one period in the model.

Figure 6: Top Marginal Take-home Rates and Inequality



Note: Green dots show the correlation between the historical marginal take-home rates and the power law inequality exponents.

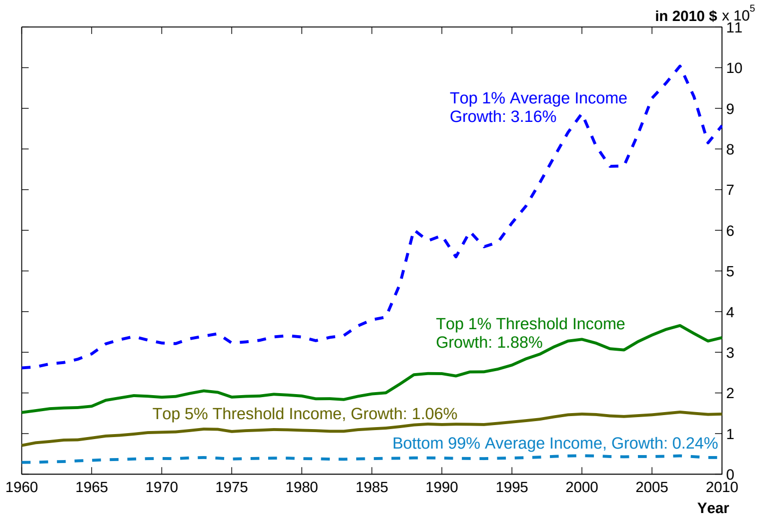
of the power law inequality exponent over the range of the top marginal take-home rates, along with the historical correlation between the marginal take-home rates and the power law inequality exponents.

The transition from the high-tax regime to low-tax regime will also have the level effect as well as the distributional effect as we discussed earlier. From equation (1), we know that the average top 1% income is given by $\frac{1}{1-\eta}y_{\text{min}}$. Thus, the change of the average can be decomposed into as follows.

$$\begin{aligned}
\Delta \log(\text{Average Top 1\% Income}) &= \Delta \log\left((\rho(1-\tau))^{\frac{1}{\kappa}}\left(\frac{1}{1-\eta_h}h_{\text{min}}\right)^{1+\frac{1}{\kappa}}\right) \\
&= \underbrace{\frac{1}{\kappa}\Delta\log(1-\tau)}_{\substack{\text{labor response} \\ \text{short-run effect} \\ =0.452,\ 51\%}} + \underbrace{\left(1+\frac{1}{\kappa}\right)\Delta\log\left(\frac{1}{1-\eta_h}\right)}_{\substack{\text{human capital increase} \\ \text{long-run effect} \\ =0.435,\ 49\%}}.
\end{aligned}$$

We refer to the effect through the labor response channel, the first term in the above equation, as the short-run effect since the effect is immediate. The second term, the change through the rise in human capital, is the long-run effect since the human

Figure 7: Income Growth Rates in the U.S.



Note: The growth rate on this figure is the annual growth rate from 1980 to 2010, and the calculations are based on the 2010 updated data from Piketty and Saez (2003)

capital increases over the transition period. Each channel accounts for about the half of the increase. Considering that we took a conservative approach regarding the human capital channel when we calibrated κ, this implies that the long-run effect of the human capital channel can be greater than the short-run effect of the labor response channel. The elasticity of taxable income with respect to the marginal net-of-tax rate implied from this calculation is 1.43, which is close to the lower bound estimate 1.48 for the highest tax bracket in Feldstein (1995).

Lastly, we discuss the implication of the tax regime change on the top 1% income share. Because our model does not have the full income dynamics of the total population, we cannot directly measure the top income share from the model. Thus, we will use the average real income growth rate of the bottom 99% from 1980 to 2010 to estimate the top 1% income share. As we can see in Figure 7, the bottom 99% incomes have grown at the relatively low rate of 0.24% from 1980 to 2010. We also assume that the minimum human capital level to enter the top 1% income group,

h_{min} has not changed. This is consistent with the assumption that we made to calibrate κ. From equation (24), we then express the top 1% income threshold y_{min} in 1980 and in 2010 as

$$\begin{aligned} y_{\text{min}}^{1980} &= (\rho(1-\tau)_H)^{\frac{1}{\kappa}} h_{\text{min}}^{1+\frac{1}{\kappa}}, \\ y_{\text{min}}^{2010} &= (\rho(1-\tau)_L)^{\frac{1}{\kappa}} h_{\text{min}}^{1+\frac{1}{\kappa}}, \end{aligned}$$

where $(1-\tau)_H = 0.3$ and $(1-\tau)_H = 0.6$. From this we obtain

$$y_{\text{min}}^{2010} = \left(\frac{(1-\tau)_L}{(1-\tau)_H}\right)^{\frac{1}{\kappa}} y_{\text{min}}^{1980} = 2^{\frac{1}{\kappa}} y_{\text{min}}^{1980}.$$

We know that from the equation (1) in Section 2.1, the average top 1% income m^{2010} in 2010 is given by

$$m^{2010} = \frac{1}{1-\eta_y^{2010}} y_{\text{min}}^{2010} = \frac{1}{1-\eta_y^{2010}} 2^{\frac{1}{\kappa}} y_{\text{min}}^{1980}.$$

Finally, we can calculate s^{2010}, the top 1% income share in the new steady state in the following way.

$$s^{2010} = \frac{m^{2010}}{m^{2010} + m^{1980} \frac{1-s_{1980}}{s_{1980}} (1.0024)^{30}} \tag{27}$$

$$= \frac{2^{\frac{1}{\kappa}}/(1-\eta_y^{2010})}{2^{\frac{1}{\kappa}}/(1-\eta_y^{2010}) + \frac{1}{1-\eta_y^{1980}} \frac{1-s^{1980}}{s^{1980}} (1.0024)^{30}} = 14.5\%,$$

where $s^{1980} = 0.0818$ is the top 1% income share in 1980. Thus, our model implies that the top 1% income share increases about a 77.2% from 8.18% in 1980 to 14.5% in 2010. Compared to the real top 1% income share of 17.42% in 2010, our model generates about 68.4% of the real increase.

4.3 Myopic Optimization with Real Historical Changes

Next we see how top incomes would evolve if we take the real historical changes in the top marginal tax rate to the model. We assume that people are myopic. That is, people reoptimize to the new top marginal tax rate assuming that the change is permanent every time the top marginal tax rate changes. We again start from year 1980 and let people reoptimize every year to the real historical change in the top marginal tax rate. Figure 8 and Figure 9, respectively shows the trajectory of the change in income and wage distribution.[14]

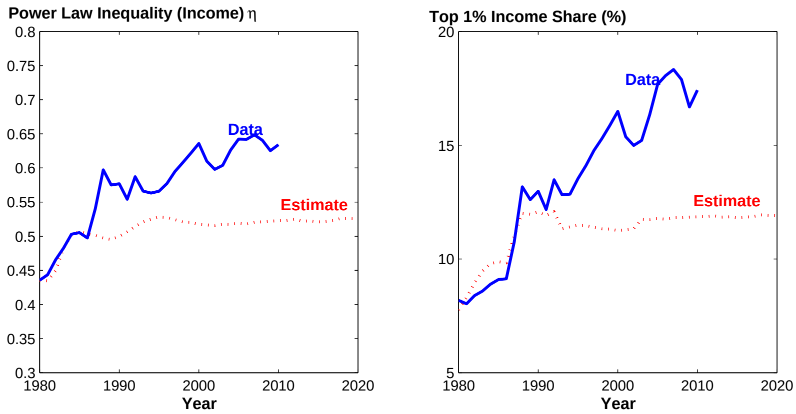
As we can see from the left panel of the figures, our model explains the changes in top income inequality fairly well until the mid-'90s, but it fails to match the subsequent increases in top income inequality. This is because we settled down in the low-tax regime around the mid-90s and there has not been much change in the top marginal tax rate since then. In other words, there was no room for the top marginal tax rate to have any sizable effect on the top income inequality since the mid-'90s. We discuss two possibilities to explain the changes in top income inequality after the mid-'90s under no sizable tax cuts. First, while the statutory top marginal tax rate we impose here has not changed much, the *effective* top marginal tax rate may have lowered further. Second, there may be other forces driving up the inequality after the mid-'90s. The possible forces include the rise in the return to experience[15], changes in the entry rate of new college graduates, talent-biased technical change, and better matching between talented workers and firms.

On the other hand, the top 1% income and wage shares on the right panel of Figure 8 and Figure 9 imply that the effect of the decrease in the top marginal tax rate in 1982 on the top income growth may have been overestimated in our model. This can be seen from the black line in Figure 8, which shows the top 1% income share estimates calculated from substituting the true values of η in (27). The real data trends in the top income distribution in Figure 8 tells us that the tax cut in 1982 did not have much level effect while it had some distributional effect on the

[14] For the wage distribution, the model is recalibrated to match the wage distribution in 1980.

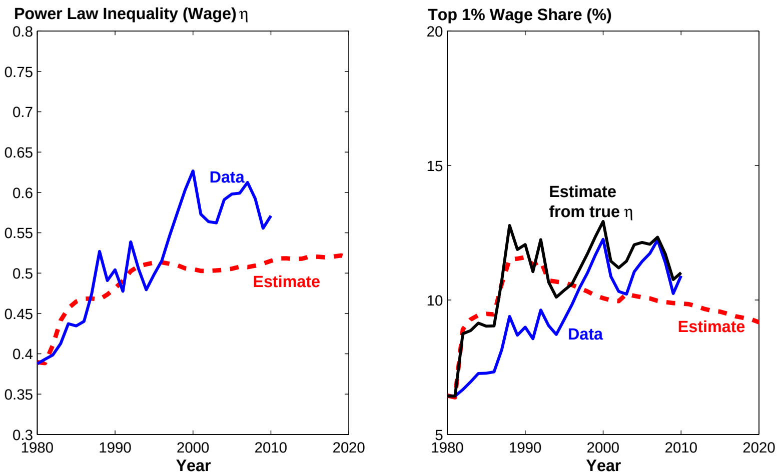
[15] Or changes in the growth rate of individual human capital, which can work the same way as in this paper

Figure 8: Top Income Inequality and Top Income Share Trajectory under Myopic Optimization



Note: The top marginal tax rate of 35% is assumed after 2010.

Figure 9: Top Wage Inequality and Top Wage Share Trajectory under Myopic Optimization



Note: The top marginal tax rate of 35% is assumed after 2010.

top incomes. Since our model estimation matches the real top income growth in the late 1980s, our calibration of κ is not out of range. One possible explanation for this discrepancy is that overall human capital including h_{min} decreased due to the early 1980s recession. Thus, our calculation of the average income growth under the assumption of unchanging h_{min} leads to the overestimation in the early 1980s, while the estimation of top income inequality remains valid. Lastly, the decline in the top 1% wage share in Figure 9 comes from the fact that the average wage of the bottom 99% has grown at the relatively large rate of 0.91%.

4.4 Sensitivity Analysis

4.4.1 Sensitivity to κ

As κ determines the elasticity of labor effort, the result of our model can be very sensitive to the choice of κ. For example, if we calibrate κ to match $1/2 \times$ elasticity estimate in Lindsey (1987), then the tax cut from 70% to 40% in the model generates 44.24% of the real inequality increase and 30.1% of the real income share increase during 1980-2010. If we calibrate κ to match $2 \times$ elasticity estimate in Lindsey (1987), then the same tax cut in the model generates 127% of the real inequality increase and 204% of the real income share increase during 1980-2010. Figure 10 shows this model sensitivity to κ in a wide range of the marginal tax rate.

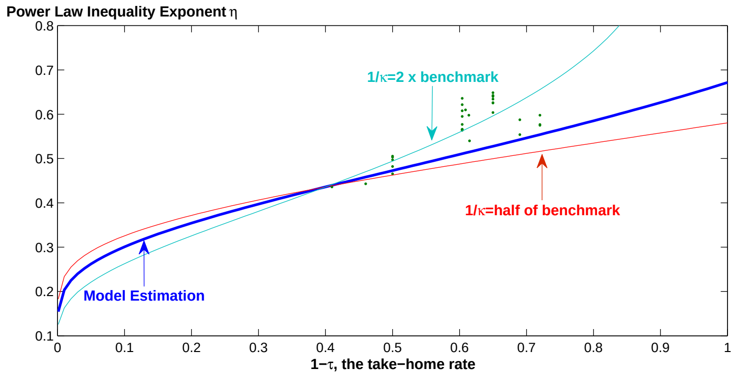
4.4.2 Infinite-Horizon with Non-Linear Taxation

First, we now relax the linear tax liability assumption in the previous section to see if our result is robust. We will estimate the non-linear tax liability function $T(\cdot)$ from TAXSIM, which is based on the real tax return data. The bellman equation of this problem will then be given by

$$V(h) = \max_{c,l,e} u(c,l) + \beta E[V(h')] \tag{28}$$

subject to

Figure 10: Sensitivity to κ



Note: The purple line indicates the case when κ is calibrated to match $2 \times$ elasticity estimate in Lindsey (1987) and the light blue line indicates the case when κ is calibrated to match $1/2 \times$ elasticity estimate in Lindsey (1987). Green dots show the correlation between the historical marginal take-home rates and the power law inequality exponents

$$c + e = hl - T(hl),$$
$$h' = \epsilon h^{\alpha} e^{\gamma},$$
$$c > 0,$$

where h' denotes the level of the next period's human capital. The numerical results of this model will be reported in the later version of this paper.

5. Conclusion

Why have top income inequality and the top 1% income share risen in the United States over the last three decades? And does the top marginal tax rate play any role in this changing dynamics of top incomes? This paper finds the relationship among the top marginal tax rate, top income inequality, and top 1% income share in an infinitely-lived heterogeneous agents model with endogenous human capital. This model estimates that the effect of the reductions in the top marginal tax rate from 1980 to 2010 can explain 65.9% of the increase in top income inequality, and 68.4%

of the increase in the top 1% income share. While the rise in top income inequality comes from the changes in the endogenous human capital accumulation process upon the changes in the top marginal tax rate, it is estimated that the standard labor supply channel and the endogenous human capital channel each contributes to one half of the rise in the top 1% income share.

While this paper delivers some insights on the dynamics of top incomes, more research is necessary to better understand it. For example, we only considered the learning-by-spending channel in human capital accumulation, which may have underestimated the effect of the top marginal tax rate on top income inequality. Moreover, our model abstracted from government transfer programs or any utility from government, which makes it not applicable to the total population. Further research that integrates the top income dynamics with the bottom 99% population is crucial for the study of top incomes to have truly meaningful policy implications.

This paper points to several interesting directions for future research. First, empirical study of the human capital technology within the top income group will be useful to better understand the top income dynamics. Second, introducing risk on the rate of return from physical capital and studying how it interacts with the risk on the rate of return from human capital will give us a better picture of income and wealth inequality. Lastly, one of the side implications of our model is that an increase in top income inequality lowers the mobility in and out of top incomes. It will be interesting to see some empirical analysis on the top income mobility especially during the last three decades in the United States.

A. Appendix: Derivations and Proofs

This appendix contains outlines of the proofs of the lemma and propositions reported in the paper.

Proof of Lemma 1 Value Function in the Infinite-Horizon Model The dynamic programming problem of (17) is simplified to

$$V(h) = \max_{l,e} u(hl - T(hl) - e, l) + \beta E[V(h')] \tag{29}$$

subject to

$$h' = \epsilon h^{\alpha} e^{\gamma}$$
$$hl - T(hl) - e > 0$$

We start from the guess $V(h) = Xh^{\frac{1}{\kappa}+1}$.
From the first order conditions, we obtain

$$l = \{Ah(1-\tau)\}^{\frac{1}{\kappa}},$$
$$e = \left\{\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]\right\}^{\frac{1}{\alpha}} X^{\frac{1}{\alpha}} h^{1+\frac{1}{\kappa}}.$$

Plugging these into $V(h)$ we obtain,

$$V(h) = \left(\frac{\kappa}{1+\kappa}\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}} + \frac{\alpha}{1-\alpha}\left\{\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]\right\}^{\frac{1}{\alpha}} X^{\frac{1}{\alpha}}\right) h^{1+\frac{1}{\kappa}}.$$

Thus, our guess is verified for X such that

$$X = \frac{\kappa}{1+\kappa}\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}} + \frac{\alpha}{1-\alpha}\left\{\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]\right\}^{\frac{1}{\alpha}} X^{\frac{1}{\alpha}}.$$

Finally we check the positivity of the consumption condition to get the following condition:

$$X < \left(\frac{1-\alpha}{\alpha}\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}\right)^{\alpha} / \left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]\right). \tag{30}$$

To simplify the notation, let

$$C \equiv \frac{\kappa}{1+\kappa}\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}, B \equiv \left\{\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]\right\}^{\frac{1}{\alpha}}. \tag{31}$$

Then we can write that X is a solution of

$$X - \frac{\alpha}{1-\alpha}BX^{\frac{1}{\alpha}} = C \tag{32}$$

$$\frac{\kappa+1}{\kappa}C \geq BX^{\frac{1}{\alpha}} \Longleftrightarrow X < \left(\frac{\kappa+1}{\kappa}\frac{C}{B}\right)^{\frac{\alpha}{1-\alpha}} \equiv X_{con}. \tag{33}$$

The left hand side of equation (32) is a hump-shaped function which crosses the X-axis at $X = 0$ and reaches its maximum $(1-\alpha)\left(\frac{1-\alpha}{B}\right)^{\frac{\alpha}{1-\alpha}}$ at $X = \left(\frac{1-\alpha}{B}\right)^{\frac{\alpha}{1-\alpha}}$. Let $f(X)$ denote the function on the left-hand side of equation (32). If $C > 0$ is less than this maximum, $f(X)$ meets the constant function C at two points, one on the left side of the hump and the other on the right side of the hump. Also, the left point will increase in C and the right point will decrease in C. Since C is increasing in $(1-\tau)$, we want the left point to be the only solution. This can be achieved from the condition (33). If $f(X_{con})$ is bigger than C, X_{con} must fall between the two solutions of the equation (32), making the right point infeasible. Simplifying $f(X_{con}) > C$, we get the condition

$$\beta\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}][\rho^{\frac{1}{\kappa}}(1-\tau)^{1+\frac{1}{\kappa}}]^{1-\alpha} < \frac{\kappa+1}{\kappa+\alpha}. \tag{34}$$

This completes the proof. QED.

Proof of Proposition 1. Power Law Inequality in the Infinite-Horizon Model

This proposition is proved by directly applying Levy and Solomon (1996) to the process of human capital (23) and the process of income (24). Moreover, the positive relationship between $(1-\tau)$ and η_y, η_h comes from the fact that $\partial\mathbf{E}[x^\lambda]/\partial\lambda = \mathbf{E}[x^\lambda \ln\lambda] > 0$ for $x > 0$ and $\lambda > 1$. QED

Proof of Proposition 2. Power Law Inequality under the Log-Normal Shock

In this proof, we use the following properties of the log normal distribution. If ϵ follows a log-normal distribution with the parameters μ and σ^2, denoted by $\epsilon \sim \ln\mathcal{N}(\mu, \sigma^2)$, then for some $\alpha, \beta > 0$,

$$\mathbf{E}[\epsilon^\alpha] = e^{\alpha\mu + \alpha^2\sigma^2/2} \tag{35}$$

$$\epsilon^\alpha \sim \ln\mathcal{N}(\alpha\mu, \alpha^2\sigma^2) \tag{36}$$

$$\beta\epsilon \sim \ln\mathcal{N}(\mu + \ln\beta, \sigma^2) \tag{37}$$

To simplify the notation, define $Z \equiv \left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]X\right)^{\frac{\gamma}{\alpha}(1+\frac{1}{\kappa})}$. Then the condition (25) in proposition 1 is rewritten as

$$\mathbf{E}\left[\left\{\epsilon^{1+\frac{1}{\kappa}}\left(\beta(1-\alpha)\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}]X\right)^{\frac{\gamma}{\alpha}(1+\frac{1}{\kappa})}\right\}^{\frac{1}{\eta y}}\right] = \mathbf{E}[\left(Z\epsilon^{1+\frac{1}{\kappa}}\right)^{\frac{1}{\eta y}}] = 1. \tag{38}$$

To determine the left hand side of equation (38), we apply the properties of the log-normal distribution to get

$$\begin{aligned}
& \epsilon^{1+\frac{1}{\kappa}} \sim \ln\mathcal{N}(-\sigma^2/2\left(1+\frac{1}{\kappa}\right), \sigma^2\left(1+\frac{1}{\kappa}\right)^2) && \text{by (36)} \\
\Rightarrow \quad & Z\epsilon^{1+\frac{1}{\kappa}} \sim \ln\mathcal{N}(-\sigma^2/2\left(1+\frac{1}{\kappa}\right) + \ln Z, \sigma^2\left(1+\frac{1}{\kappa}\right)^2) && \text{by (37)} \\
\Rightarrow \quad & \mathbf{E}[\left(Z\epsilon^{1+\frac{1}{\kappa}}\right)^{\frac{1}{\eta y}}] = e^{-\frac{1}{\eta y}\sigma^2/2(1+\frac{1}{\kappa}) + \ln Z + \frac{1}{\eta y}^2\sigma^2(1+\frac{1}{\kappa})^2/2} && \text{by (35)}
\end{aligned}$$

Equating the last expression to 1, we finally obtain

$$-\frac{1}{\eta_y}\sigma^2/2\left(1+\frac{1}{\kappa}\right) + \ln Z + \frac{1}{\eta_y}^2\sigma^2\left(1+\frac{1}{\kappa}\right)^2/2 = 0$$

Simplifying it and plugging $\mathbf{E}[\epsilon^{1+\frac{1}{\kappa}}] = e^{\sigma^2(1+\frac{1}{\kappa})\frac{1}{2\kappa}}$ into Z will give us (26). This completes the proof. QED.

References

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez, "The World Top Incomes Database," http://g-mond.parisschoolofeconomics.eu/topincomes November 2012.

Atkinson, A. B., "Income inequality in OECD countries: Data and explanations," *CESifo Economic Studies*, 2003, *49* (4), 479–513.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, August 1967, *75* (4), 352–365.

Benhabib, Jess, Alverto Bisin, and Shenghao Zhu, "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents," *Econometrica*, January 2011, *79* (1), 123–157.

Carnevale, Anthony P., Stephen J. Rose, and Ban Cheah, "The College Payoff: Education, Occupations, Lifetime Earnings," 2011.

Erosa, Andress and Tatyana Koreshkova, "Progressive Taxation in a Dynastic Model of Human Capital," *Journal of Monetary Economics*, 2007, *54*, 667–685.

Feldstein, Martin, "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy*, 1995, *103* (3), 551–572.

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, August 1999, *114* (3), 739–767.

_ , "Power Laws in Economics and Finance," *Annual Review of Economics*, 2009, *1* (1), 255–294.

_ and Augustin Landier, "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 2008, *123* (1), 49–100.

Huggett, Mark, Gustavo Ventura, and Amir Yaron, "Sources of Lifetime Inequality," *American Economic Review*, December 2011, *101* (7), 2923–54.

Jones, Charles I. and Jihee Kim, "A Schumpeterian Model of Top Income Inequality," October 2014. NBER Working Paper No. 20637.

Kopczuk, Wojciech and Emmanuel Saez, "Top Wealth Shares in the United States, 1916-2000: Evidence from Estate Tax Returns," *National Tax Journal*, June 2004, *57*, 445–487.

Levy, Moshe and Sorin Solomon, "Power laws are logarithmic Boltzmann laws," *International Journal of Mondern Physics C*, August 1996, *7*, 595–601.

Lindsey, Lawrence B., "Individual Taxpayer Response to Tax Cuts: 1982-1984 With Implications for the Revenue Maximizing Tax Rate," *Journal of Public Economics*, 1987, *33*, 173–206.

Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 1988, *22* (1), 3–42.

Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38*, 175–208.

Nirei, Makoto and Wataru Souma, "A Two Factor Model of Income Distribution Dynamics," *Review of Income and Wealth*, 2007, *53*, 440–459.

Philippon, Thomas and Ariell Reshef, "Wages and Human Capital in the U.S. Financial Industry: 1909-2006," *Quarterly Journal of Economics*, 2012. forthcoming.

Piketty, Thomas and Emmanuel Saez, "Income Inequality In The United States, 1913–1998," *Quarterly Journal of Economics*, February 2003, *118* (1), 1–39.

—, —, and Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," Technical Report November 2011. NBER Working Paper 17616.

Rosen, Sherwin, "The Economics of Superstars," *American Economic Review*, December 1981, *71* (5), 845–58.

Saez, Emmanuel, "Essays on the Economics of Income Taxation," May 1999. Ph.D. Dissertation.

—, "Using Elasticities to Derive Optimal Tax Rates," *Review of Economic Studies*, 2001, *68*, 205–229.

Tax Foundation, "U.S. Federal Individual Income Tax Rates History, 1913-2011," November 2012. http://taxfoundation.org/article/us-federal-individual-income-tax-rates-history-1913-2011-nominal-and-inflation-adjusted-brackets.

Trostel, Philip A., "The Effect of Taxation on Human Capital," *Journal of Political Economy*, 1993, *101*, 327–335.