High Marginal Tax Rates on the Top 1%?

Lessons from a Life Cycle Model with Idiosyncratic Income Risk

July 14, 2020

Fabian Kindermann
University of Regensburg, CEPR, HCEO and Netspar

Dirk Krueger*
University of Pennsylvania, CEPR, CFS, NBER and Netspar

Abstract

This paper argues that high marginal labor income tax rates on top earners are an effective tool for social insurance even when households have a high labor supply elasticity, make dynamic savings decisions, and policies have general equilibrium effects. We construct a large scale overlapping generations model with uninsurable labor productivity risk, show that it has a realistic wealth distribution and then numerically characterize the optimal top marginal rate. We find that marginal tax rates for top 1% earners of 79% are optimal as long as the model earnings and wealth distributions display a degree of concentration as observed in US data.

Keywords: Progressive Taxation, Top 1%, Social Insurance, Income Inequality

JEL Classifications: E62, H21, H24

We thank numerous seminar and conference participants, as well as Juan Carlos Conesa, William Peterman and Chad Jones for many useful comments. Krueger thanks the National Science Foundation for support under grants SES 1123547 and SES 1757084.

**Corresponding author:*

Address:	Department of Economics, University of Pennsylvania
Email:	dkrueger@upenn.edu

1 Introduction

In the last 40 years labor earnings, market income and wealth inequality have increased substantially in the U.S. at the top end of the distribution. For example, Alvaredo et al. (2013) report that the share of total household income accruing to the top 1% income earners was about 10% in the early 1970ies but now exceeds 20% in the U.S.[1] At the same time the highest marginal tax rate declined from levels consistently above 60% to below 40%. This triggered academic and popular calls to very substantially raise marginal income tax rates at the top of the distribution, with the explicit objective of reversing the trend of increasing economic inequality, see e.g. Diamond and Saez (2011), Piketty (2014), Reich (2010), but also the political Occupy Wall Street movement. However, reducing inequality is not necessarily an objective in and of itself for a benevolent government.

In this paper we characterize the welfare-maximizing labor income tax rate on the top 1% earners.[2] To do so, we first develop a simple analytically tractable model in which we can theoretically characterize the revenue-maximizing and the welfare-maximizing tax rate on top income earners. We show that these rates depend on the elasticity of earnings with respect to the top marginal tax rate, the shape of the top income distribution, the use of the extra tax revenue, as well as the value of social insurance. To quantify the optimal marginal tax rate on the top 1% of earners we then extend the simple model to a dynamic overlapping generations economy with ex-ante skill heterogeneity, idiosyncratic wage risk and endogenous labor supply and savings choices. The calibration of the model (as well as the sensitivity analysis) focuses in detail on the parameters the simple model has identified as the key determinants of the revenue- and welfare-maximizing rates. Specifically, to ensure that the model delivers an empirically plausible earnings and wealth distribution (relative to the evidence from the 2007 Survey of Consumer Finances), including at the very top end of the distribution, we follow Castaneda et al. (2003) and calibrate a labor productivity process with "superstar" states directly to empirically observed top income and wealth shares. As a consequence, in the model the top 1% look exactly as in the data, at least with respect to their key economic characteristics.

We then use the calibrated version of the model to compute, within a restricted class of income tax functions, the optimal one-time tax reform, which in turn induces an economic transition from the current status quo[3] towards a new steady state. The key policy choice variable is the marginal tax rate applying to the top 1%. We find that this optimal top marginal tax rate along the transition path is indeed very high at 79%, consistent with the empirically observed levels after World War II. If we had ignored the transition path

[1] This increase was not nearly as severe in other countries such as France and Japan. Jones and Kim (2018) explore a Schumpeterian growth model with creative destruction to rationalize the cross-country differences in these trends.

[2] Welfare is measured as the weighted sum of expected lifetime utility of households currently alive and born in the future. This welfare measure includes households currently in the top 1% as well as the chance of being a top 1% household in the future. Excluding the top 1% from the social welfare function increases optimal top rates, but the difference in results is quantitatively small.

[3] The status quo is a stylized version of the current U.S. personal income tax code.

and the implied dynamics of the wealth distribution and had instead maximized steady state welfare, the optimal marginal tax rate would be even higher at 82%.

We then show that this result is primarily driven by the social insurance benefits that these high taxes imply. To match the very high concentration of labor earnings and wealth in the data, our model requires that households, with low probability, have the opportunity to work for very high wages. These high wages stand in for attractive entrepreneurial, entertainment or sports opportunities in the real world. In the model, the labor supply of these households is not too strongly affected by very high marginal tax rates even with a utility function with high Frisch labor supply elasticity, as long as these households have not yet accumulated massive amounts of wealth, i.e. as long as they have not been "superstars" for too long. A strong negative income effect on leisure makes these households maintain their labor effort as marginal tax rates rise. From the perspective of implementing social insurance against idiosyncratic labor productivity risk via the income tax code it is then optimal to tax these incomes at a very high rate. To be clear, we do not wish to argue that very high marginal tax rate on top earners are optimal in all models that generate empirically plausible earnings and wealth distributions. Rather, our objective is to show that when these distributions emerge from one prominent mechanism in the literature (large, persistent but mean-reverting labor productivity shocks) that enjoy some empirical support (see Section 7.1.3), a strong normative argument for such high rates emerges naturally.

After reviewing the literature, in Section 2 we construct a simple version of our model to develop the economic intuition for our quantitative results. Sections 3, 4 and 5 set out the quantitative model, discuss its calibration and its fit to the micro data, respectively. In Section 6 we present and decompose our optimal tax results, and Section 7 contains sensitivity analyses with respect to the key parameters characterizing household preferences and labor productivity. Section 8 concludes, and the appendix contains all theoretical derivations as well as details of the calibration and the computational algorithm.

Related Literature The point of departure for this paper is the static literature on optimal taxation of labor income, starting from Mirrlees (1971), Diamond (1998) and Saez (2001). Diamond and Saez (2011) discuss the practical implications of this literature and provide a concrete policy recommendation that advocates for taxing labor earnings at the high end of the distribution at very high marginal rates, in excess of 70%. On the empirical side the literature that motivates our analysis includes the papers by Piketty and Saez (2003) and Alvaredo et al. (2013) who document an increasing concentration of labor earnings and income at the top end of the distribution, and argue that this trend coincides with a reduction of marginal tax rates for top income earners. Their work thus provides the empirical underpinning for the policy recommendation by Diamond and Saez (2011) of increasing top marginal income tax rates substantially.

Methodologically, our paper is most closely related to the quantitative dynamic optimal taxation literature.[4] Examples include Domeij and Heathcote (2004), Conesa and Krueger (2006), Conesa et al. (2009), Bakis et al. (2015), Fehr and Kindermann (2015) and Hubmer et al. (2020). A subset of this literature (see e.g. Guner, Lopez-Daneri and Ventura, 2016, Holter, Krueger and Stepanchuk (2019) or Badel, Huggett and Luo, 2020) characterizes the relationship between tax rates and tax revenues (that is, the Laffer curve). In this paper we show that although the welfare-optimal top marginal tax rate is quantitatively smaller than the revenue-maximizing rate (from the top 1%), it is fairly close.[5]

Our measure of social welfare disentangles aggregate efficiency gains from the redistributive benefits of progressive taxation and thus departs from the classical notion of utilitarianism. It builds on Benabou (2002) who studies optimal progressive taxation and education subsidies in an endogenous growth model driven by human capital accumulation but abstracts from the accumulation of non-human wealth.

Especially relevant for our work is the paper by Badel et al. (2020) who build on the human capital model of Benabou (2002). These authors study a dynamic economy with endogenous human capital accumulation to quantify the effects of high marginal income tax rates at the top of the distribution on the aggregate level of economic activity as well as the distribution of wages (which is endogenous in their model, due to the human capital accumulation decision of households) and household incomes. They stress the negative long-run effect of top marginal tax rates on human capital accumulation and conclude that the revenue-maximizing tax rate on top earners is about 15 percentage points lower than in a comparable model with exogenous human capital.

The complementary work of Brüggemann and Yoo (2015) studies the aggregate and distributional steady state consequences of an increase in the top marginal tax rate from the status quo to 70%, and consistent with our findings, reports substantial adverse aggregate and large positive distributional consequences, resulting in net welfare gains from the policy reform they study. Brüggemann (2019) explores the importance of entrepreneurial activity for the taxation of top income earners. Finally, for our quantitative analysis to be credible, it is crucial for the model to deliver an empirically plausible earnings and wealth distribution at the low and especially at the right tail of the distribution. We therefore build on the literature studying the mechanisms to generate sufficient wealth concentration in dynamic general equilibrium models, especially Castaneda et al. (2003), but also Quadrini (1997), Krusell and Smith (1998), Cagetti and DeNardi (2006).

[4] A comprehensive recent survey of the dynamic taxation literature is contained in Stantcheva (2020).

[5] We study optimal progressive labor income taxes, thereby sidestepping the question whether capital income taxation is a useful redistributive policy tool. The benchmark result by Chamley (1986) and especially Judd (1985) suggest that positive capital income taxation is suboptimal, at least in the long run, even if the social welfare function places all the weight on households not owning capital. The ensuing theoretical literature on using capital income taxes for redistribution and social insurance includes Bassetto (2014), Vogelgesang (2000) and Jacobs and Schindler (2012). Our paper also connects to the theoretical literature on optimal taxation over the life cycle, e.g. Erosa and Gervais (2002).

2 Building Intuition: A Static Model of Labor Supply

In this section, we build some intuition for our results using a simplified static version of the full model employed in the quantitative analysis. We first set out the basic model and tax experiment, which allows us to characterize the peak of the Laffer curve for top 1% tax payers analytically using a policy elasticity and distribution parameters in the tradition of the sufficient statistics literature. We then derive closed-form solutions for the policy elasticity in terms of preference parameters for the utility function used throughout the paper. Finally, we shift focus from tax revenue maximization (the Laffer curve) to social welfare maximization by clarifying the social insurance benefits of high marginal taxes at the top of the income distribution. The purpose of this section is to lay the foundation for the intuition of (and establish the notation for) the results from the dynamic model, and to justify why modeling wealth dynamics along the transition between steady states is quantitatively important for the determination of the optimal top marginal tax rate.

2.1 A Simple Static Model

There is a continuum of ex ante identical agents, but ex post a share Φ_l has productivity e_l and a share Φ_h (e.g. the top 1%) has productivity $e_h > e_l$. Thus e_h/e_l is a measure of productivity- and thus income inequality. Households of type $i \in \{l, h\}$ choose labor supply n_i and consumption c_i to maximize the utility function

$$U(c,n) = \frac{c^{1-\gamma}}{1-\gamma} - \lambda \frac{n^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}. \tag{1}$$

The parameter $\chi \geq 0$ governs the Frisch elasticity of labor supply and thus the importance of the substitution effect on labor supply when top marginal tax rates change. The parameter $\gamma \geq 0$ determines both the magnitude of the income effect on labor supply from tax rate changes as well as risk aversion. For simplicity, in this section we set $\lambda = 1$, but will use it for calibration purposes in the full quantitative model.

Households pay taxes $T(z_i)$ on their labor earnings $z_i = e_i n_i$. We assume that high productivity earners always face the marginal tax rate τ_h. Income is only taxable above an earnings threshold $\bar{z}$, assumed to be larger than labor earnings of the low productivity agents, i.e. $z_h \geq \bar{z} \geq z_l$ when both groups choose labor supply optimally. In addition, individuals receive a lump-sum transfer R, so that the labor tax function reads as

$$T(z) = \tau_h \max(z - \bar{z}, 0) - R.$$

Consequently, the budget constraints of the two earners in our economy are

$$c_l = e_l n_l + R \quad \text{and} \quad c_h = \bar{z} + (1-\tau_h)(e_h n_h - \bar{z}) + R.$$

We conduct the following tax experiment: We increase the top marginal tax rate by an amount $d\tau_h$. This triggers a behavioral labor supply reaction of top earners that leads to a change in top labor earnings dz_h. At the same time, the lump-sum transfer adjusts by dR to keep the tax reform revenue-neutral. Such a tax reform leads to a redistribution of resources from top earners to low income households (and thus to social insurance from an ex-ante perspective) as long as τ_h is on the increasing part of the Laffer curve. We now seek to determine what level of the top marginal tax rate τ_h maximizes (i) tax revenue from the top 1% earners and (ii) maximizes suitably defined social welfare.

2.2 The Top Laffer Curve

We first focus on the tax revenue generated from the top income earners as a function of the top tax rate τ_h. We call this relation the Top Laffer curve. Evaluating this fiscal effect is of first-order importance for a welfare analysis, as it defines the upper limit of redistribution by the government.[6] Using the assumptions on the tax schedule, taxes paid by an individual with earnings $z_h \geq \bar{z}$ can be written as

$$T(z_h) = \tau_h(z_h - \bar{z}) - R. \tag{2}$$

Our objective is to characterize the top marginal tax rate τ_h that maximizes tax revenue from the top earners $T(z_h)$. The next proposition characterizes this rate.[7]

Proposition 1 *The revenue-maximizing top marginal rate $\tau_h = \tau_{Laffer}$ is given by*

$$\tau_{Laffer} = \frac{1 - (a-1) \cdot \tau_a(\bar{z}) \cdot \epsilon(\tau_a(\bar{z}))}{1 + a \cdot \epsilon(z_h)} \tag{3}$$

where

$$\frac{a}{a-1} = \frac{z_h}{\bar{z}} \quad \text{and} \quad \tau_a(\bar{z}) = \frac{-R}{\bar{z}}$$

are the ratio between top earnings z_h and the top tax threshold $\bar{z}$ as measured by the Pareto coefficient a, and the average tax rate at $\bar{z}$, respectively, and

$$\epsilon(z_h) = \frac{dz_h}{d(1-\tau_h)} \cdot \frac{1-\tau_h}{z_h} \quad \text{and} \quad \epsilon(\tau_a(\bar{z})) = \frac{d\tau_a(\bar{z})}{d(1-\tau_h)} \cdot \frac{1-\tau_h}{\tau_a(\bar{z})}$$

are the elasticities of z_h and $\tau_a(\bar{z})$ with respect to the net-of-tax rate $1 - \tau_h$.

Proof: See Appendix A.

---

[6] Since in our quantitative analysis we also study tax reforms that are budget neutral (i.e. use revenue generated from the top 1% to lower tax rates in other parts of the income distribution) we do not focus on the overall income tax Laffer curve, in contrast to, e.g. Saez (2001), Badel and Huggett (2017), or Guner et al. (2014). Nevertheless, our analytical results for the peak of the Laffer curve resemble, and in the absence of income effects on labor supply, are identical to theirs.

[7] This rate is also the welfare-maximizing rate if top earners receive no weight in the social welfare function.

Corollary 2 *If there are no adjustments of the labor earnings tax schedule below $\bar{z}$, i.e. $dR = \epsilon(\tau_a(\bar{z})) = 0$, the revenue-maximizing tax rate is given by*

$$\tau_{\text{Laffer}} = \frac{1}{1 + a \cdot \epsilon(z_h)}. \tag{4}$$

Note that, as in Saez (2001) or Badel and Huggett (2017), this is model-free in the sense that it applies to arbitrary models of labor supply. Of course, the elasticities and distributional statistics in the formula depend on the specific model under consideration. We will provide an example to make this concrete below.

In order to interpret the components of the revenue maximizing tax rate in (3), consider first the case in Corollary 2 and assume that a change in the top marginal rate is not associated with a change in the other elements of the tax function, $\epsilon(\tau_a(\bar{z})) = 0$. Formula (4), first derived by Saez (2001), is used by Diamond and Saez (2011) to make explicit policy recommendations about actual income taxes at the top of the distribution. It shows the trade-off between the two main effects of an increase in the top marginal tax rate above the threshold $\bar{z}$: (i) a *mechanical* increase in tax revenue of $d\tau_h(z_h - \bar{z})$ and (ii) a negative *behavioral* response $\tau_h dz_h$ stemming from the adjustment of labor supply and thus earnings z_h by top earners to changes in the top marginal tax rate. The less elastic earnings are to the tax rate (the lower is $\epsilon(z_h)$) and the fatter the right tail of the earnings distribution (the lower is a), the higher is the revenue-maximizing rate τ_{Laffer}. Note that the elasticity $\epsilon(z_h)$ is a *policy elasticity* in the sense of Hendren (2016), that is, $\epsilon(z_h)$ summarizes the earnings reaction of top earners to *the specific tax experiment* considered here.

The full formula in (3) includes an additional *mechanical* effect that emerges when a change in the top rate is associated with a change of other elements of the tax schedule (in our simple model an increase in the lump-sum transfer R), or when it is used to reduce marginal tax rates at lower incomes, as in our quantitative analysis below. In this case $\epsilon(\tau_a(\bar{z})) \neq 0$, and an increase in τ_h and associated decline in the top marginal net-of-tax rate $1 - \tau_h$ changes the average tax rate at the threshold income level $\bar{z}$. It thus also changes tax revenues from top earners on their earnings below the threshold $\bar{z}$.

In summary, there are three major determinants of the size of the Laffer tax rate in this simple model: (i) average incomes above the top tax threshold $\bar{z}$ summarized by the statistic a, (ii) the extent to which households react with their labor earnings to the tax experiment summarized in the elasticity $\epsilon(z_h)$ and (iii) the extent to which the lower part of the tax schedule reacts to changes in the top marginal tax rate summarized by $\epsilon(\tau_a(\bar{z}))$.

2.3 The Policy Elasticity

Although the formula in (3) is general, its ingredients, especially the policy elasticity $\epsilon(z_h)$ of earnings with respect to the top net-of-tax rate $1 - \tau_h$ but also the Pareto coefficient a depend on the model under consideration, and is typically not invariant to τ_h. To clarify which features and parameters of the model these statistics depend on, we now study the household optimization problem. In the context of this simple model we can analytically characterize the policy elasticity $\epsilon(z_h)$ that determines the peak of the top 1% Laffer curve in Proposition 1.

Proposition 3 *The policy elasticity is*

$$\epsilon(z_h) = \epsilon_h^u - \eta_h \cdot \frac{\bar{z}}{z_h} \cdot \left[1 + \frac{\tau_a(\bar{z})}{1-\tau_h} \cdot \epsilon(\tau_a(\bar{z}))\right],$$

where ϵ_h^u and η_h denote the uncompensated labor supply elasticity and income elasticity, given by

$$\epsilon_h^u = \frac{(1-\gamma)(1-\tau_h)z_h + \tau_h\bar{z} + R}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{\tau_h\bar{z}+R}{\chi}}. \tag{5}$$

and

$$\eta_h = \frac{dz_h}{dR}(1-\tau_h) = \frac{-\gamma(1-\tau_h)z_h}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{\tau_h\bar{z}+R}{\chi}}.$$

Proof: see Appendix A.

Corollary 4 *Suppose the tax system is purely proportional, such that $\bar{z} = 0$ and $R = 0$. Then*

$$\epsilon(z_h) = \frac{1-\gamma}{\gamma + \frac{1}{\chi}}. \tag{6}$$

Proposition 3 and Corollary 4 characterize the determinants of the labor supply reaction to the tax policy experiment. It shows that the elasticities of labor supply with respect to wages and exogenous income, in turn, primarily depend on the structural parameters of the underlying model, including the preference parameters controlling income and substitution effects γ and χ, as well as those governing the income distribution as summarized by a.

Consider first the case of a purely proportional tax system as in Corollary 4. In this case the policy elasticity $\epsilon(z_h) = \epsilon_h^u$ is a function only of the structural preference parameters capturing the standard substitution effect χ and income effect γ of a change in net wages on labor supply. If $\gamma = 0$ the income effect is absent and the policy elasticity is $\epsilon(z_h) = \epsilon_h^u = \chi$. If $\gamma = 1$, then the policy elasticity is zero as income and substitution effect cancel.

However, the simple model also clarifies that the precise tax reform matters. If $\bar{z} > 0$, then the income effect from the tax change is smaller since taxes are only lowered above the threshold $\bar{z}$. In fact, household exactly at the threshold $z_h = \bar{z}$ experience no income effect at all (recall we still assume $\tau_a(\bar{z}) = 0$), and thus the policy elasticity $\epsilon(z_h) = \epsilon_h^u - \eta_h \cdot \frac{\bar{z}}{z_h} = \epsilon_h^u - \eta_h = \epsilon_h^c$ is governed exclusively by the Hicksian compensated labor supply elasticity. A household with a greater z_h experiences a larger negative income effect on leisure (a larger positive income effect on labor). Consequently, the extra $\eta_h \cdot \frac{1-a}{a} = \eta_h \cdot \frac{\bar{z}}{z_h}$ declines with z_h, and the labor earnings reaction becomes smaller. If z_h is large relative to $\bar{z}$, the labor earnings reaction is approximately the same as in a proportional tax system, as the tax payment $T(\bar{z})$ on income below the threshold is small relative to the total tax bill $T(z)$. Finally, if in addition other parts of the tax schedule adjust to the change in the top rate, then $\epsilon(\tau_a(\bar{z})) \neq 0$. Even if the top earner was exactly at the top threshold $\bar{z}$, she would experience an additional income effect on labor supply due to the tax change for her income below the threshold. In the case of $\epsilon(\tau_a(\bar{z})) < 0$, these additional income effects make labor supply more elastic to the tax reform (i.e. increase the policy elasticity $\epsilon(z_h)$), and thus, ceteris paribus, reduce the maximal tax revenue and tax rate at which the peak of the top Laffer curve is attained.

Summing up, the policy elasticity of top labor earnings with respect to a change in the tax system varies with the adjustment of the tax code below the threshold, and is determined by the ratio of top labor earnings and the top tax threshold $z_h/\bar{z}$ as encoded in the Pareto parameter a, the uncompensated labor supply elasticity ϵ_h^u, the income effect η_h, and the sensitivity of $\tau_a(\bar{z})$ with respect to the net-of-tax rate $1-\tau_h$. These in turn depend crucially on three parameters. First, the Frisch labor supply elasticity χ parameterizing the substitution effect, second, the parameters shaping the top of the earnings distribution, as summarized by a, and third, the parameter γ determining the size of the income effect. We will therefore place special focus on calibrating these parameters in Section 4.

2.4 From the Laffer Curve to Welfare

Maximizing tax revenues and thus the size of transfers R is welfare maximizing if and only if top earners receive no weight in the social welfare function. Now we demonstrate this analytically and also show that the welfare-maximizing rate depends positively on the degree of inequality e_h/e_l in the economy as well as the value of social insurance represented by parameter ν, as discussed below. To obtain an analytical characterization of the welfare-maximizing tax rate we assume that there are no income effects in the utility function (1), $\gamma = 0$, so that the household maximization problem has a closed-form solution given by

$$\begin{aligned} n_l^* &= [e_l]^\chi && \text{and} \quad c_l^* = [e_l]^{1+\chi} + R \\ n_h^* &= [(1-\tau_h)e_h]^\chi && \text{and} \quad c_h^* = [(1-\tau_h)e_h]^{1+\chi} + \tau_h \bar{z} + R \end{aligned}$$

and expected utility, as a function of the tax rate τ_h

$$V_l(\tau_h) = \frac{[e_l]^{1+\chi}}{1+\chi} + R \quad \text{and} \quad V_h(\tau_h) = \frac{[(1-\tau_h)e_h]^{1+\chi}}{1+\chi} + \tau_h \bar{z} + R.$$

Finally, the tax and the transfer are related by

$$R = R(\tau_h) = \Phi_h \left[\tau_h e_h n_h - \tau_h \bar{z}\right] = \Phi_h \tau_h \left[[(1-\tau_h)]^\chi \, [e_h]^{1+\chi} - \bar{z} \right]$$

Define social welfare as

$$\mathcal{W}(\tau_h) = (1-\Phi_h)\,\frac{V_l(\tau_h)^{1-\nu}}{1-\nu} + \Phi_h \frac{V_h(\tau_h)^{1-\nu}}{1-\nu}$$

$$= (1-\Phi_h)\frac{\left(\frac{[e_l]^{1+\chi}}{1+\chi}+R\right)^{1-\nu}}{1-\nu}+\Phi_h\frac{\left(\frac{[(1-\tau_h)e_h]^{1+\chi}}{1+\chi}+\tau_h\bar{z}+R\right)^{1-\nu}}{1-\nu}.$$

This social welfare function can be interpreted in two ways. First, one can think of $\mathcal{W}(\tau_h)$ as utilitarian social welfare where ν measures inequality aversion of society. Alternatively, $\mathcal{W}(\tau_h)$ is the expected utility of a household with a Greenwood, Hercovitz and Huffman (1988) utility function in consumption and labor with curvature parameter ν. In either interpretation, ν measures the importance of social insurance by determining how costly it is to be born with low productivity e_l. The following proposition then characterizes the welfare-maximizing top tax rate τ_h^* and shows how it depends on the desire for social insurance parameterized by ν and inequality as measured by e_h/e_l.

Proposition 5 *Assume that* $\frac{e_h}{e_l} > 1+\frac{1}{\chi}$.[8] *Then the welfare maximizing tax rate* τ_h^* *is*

1. *strictly below the revenue-maximizing tax rate:* $\tau_h^* < \tau_h^{Laffer}$
2. *equal to zero in the absence of valued insurance,* $\tau_h^* = 0$ *if* $\nu = 0$ *and strictly positive as long as society values social insurance: if* $\nu > 0$*, then* $0 < \tau_h^* < \tau_h^{Laffer}$
3. *strictly increasing in the preference for social insurance* ν *and the degree of productivity inequality* e_h/e_l *and thus income inequality*[9] z_h/z_l.

Proof: see Appendix A.

To get a first sense of the magnitude of the optimal tax rate, how it relates to the revenue-maximizing rate, and how it depends on inequality in the economy and the desire for social insurance, we now provide a back-of-the-envelope calculation. We calibrate the model, motivated by Diamond and Saez (2011), such that the peak of the Laffer curve lies at exactly the 73% rate these authors advocate for. To make the back-of-the-envelope calculations consistent with our quantitative work, we choose a policy elasticity of $\epsilon(z_h) = 0.21$ and a Pareto coefficient of $a = 1.79$. These numbers are within the range of typical values reported in the literature.[10] Section 4.3.2 offers more details on how we arrive at these choices. Therefore in this section we choose $\chi = 0.21$ since, in the absence of income effects, the policy elasticity equals χ, and target incomes of $z_l = 0.63, \bar{z} = 7.62, z_h = 17.18$, resulting in a Pareto coefficient of $a = 1.79$ and a ratio of labor incomes between the top 1% and bottom 99% of 27, as in the quantitative model.[11]

---

8 This assumption ensures that even at the revenue-maximizing tax rate τ_h^{Laffer} utility of high-productivity individuals is larger than that of low-productivity individuals.

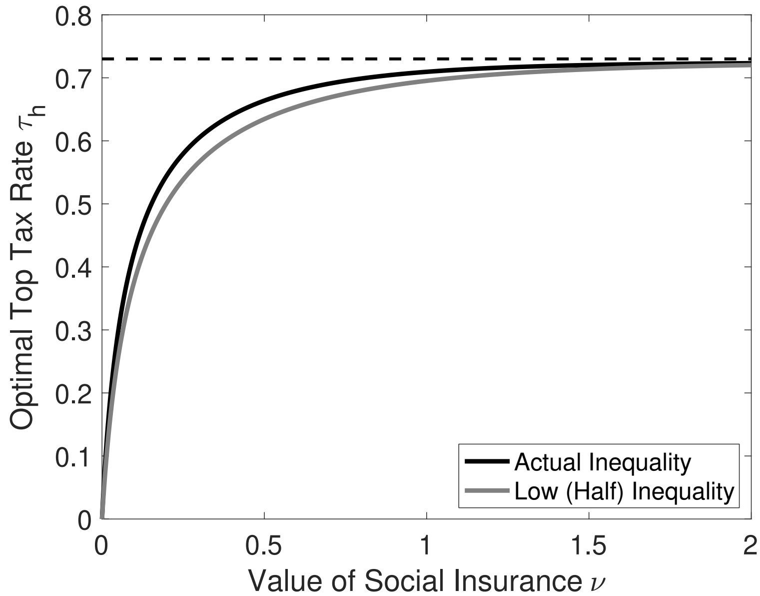
9 When varying e_h/e_l in the comparative statics exercise we also vary $\bar{z}$ to keep $z_h/\bar{z}$ constant.

10 Saez (2001) and Diamond and Saez (2011) argue for values between 1.5 and 2.0 for the Pareto parameter, depending on the exact definition of taxable income. The literature on the elasticity of labor supply is much broader and offers a wide range of values for this micro elasticity. A value between 0.20 and 0.25 is in line with early estimates by MaCurdy (1981) and typical in the life-cycle labor literature, see also Keane (2011).

11 In order to obtain these income ratios in the simple model requires (at a current marginal rate of $\tau_h = 39.6\%$) productivities $e_l = 0.68, e_h = 11.45$.

With these numbers, the peak of the Laffer curve by construction is at 73%.[12] The solid line of Figure 1 displays the welfare-maximizing rate, as a function of the value of social insurance ν. For example, the optimal rate is $\tau_h^* = 66.4\%$ for $\nu = 0.5$ and $\tau_h^* = 71.0\%$ for $\nu = 1$. In contrast, if there is no desire for social insurance ($\nu = 0$), then the optimal rate is at zero, as the proposition above has demonstrated theoretically. The figure also shows how the optimal tax rate depends on inequality in the economy. The gray line displays the optimal tax rate in an economy where the ratio of top to bottom incomes z_h/z_l is cut in half.[13] Not surprisingly, the optimal tax rate increases with inequality, holding the desire for social insurance ν fixed. For example, at $\nu = 1$ the optimal rate declines from $\tau_h^* = 71.0\%$ to 69.5%, and at $\nu = 0.5$ it declines from 66.4% to 63.5%

Figure 1: Optimal Tax Rate for Different Degrees of Income Inequality $\frac{z_h}{z_l}$ - Actual and Half $\frac{z_h}{z_l}$



The results in this section are meant to convey two basic insights. First, the optimal top income tax rate depends crucially on the shape of the top income distribution, the earnings policy elasticity, the desire for social insurance and the use of the extra tax revenue, represented in the simple model by the statistics $(a, \epsilon(z_h), \nu, \tau_a(\bar{z}))$. Second, even though the welfare-maximizing rate is in general lower than the revenue-maximizing rate as long as the social desire for insurance ν is finite, these rates are quantitatively close for even moderate ν, given the observed differences in incomes z_h/z_l and implied low marginal utilities of income-rich relative to income poor individuals.

Finally, even in the static model these key objects are not invariant to the top marginal tax rate. In the dynamic model they are neither constant over time. Specifically, we will show that the dynamics of the wealth distribution matters considerably for the size of the policy elasticity $\epsilon(z_h)$ and, hence, for the location of the peak of the Laffer curve.

[12] We adjust $\bar{z}$ such that at the revenue-maximizing choice $a = 1.79$ continues to be true.

[13] We adjust $\bar{z}$ so that the peak of the Laffer curve remains unchanged.

Since wealth is a slow moving object over time, it is crucial to consider the transitional dynamics induced by a tax policy reform; we will show that the revenue-maximizing rate differs greatly in the short- and in the long run. The dynamic response of the economy ultimately also shapes the socially optimal top tax rate. To show this we now introduce the full dynamic model.

3 The Quantitative Model

We now study a standard large-scale overlapping generations model in the spirit of Auerbach and Kotlikoff (1987), augmented by exogenous ex-ante heterogeneity across households by education levels as well as ex-post heterogeneity due to uninsurable idiosyncratic labor productivity and thus wage risk, as in Conesa, Kitao and Krueger (2009). Given the focus of the paper it is especially important that the endogenous earnings and wealth distributions predicted by the model well approximate their empirical counterparts, both at the low and the high end of the distribution.

We first set out the model using recursive language and define a stationary recursive competitive equilibrium. We then turn to a description of the potential policy reforms and the transition dynamics induced by it.

3.1 Technology

The final good is produced by a representative, competitive firm that hires capital and labor on competitive spot markets to operate the constant returns to scale technology

$$Y = \Omega K^{\epsilon} L^{1-\epsilon}, \tag{7}$$

where $\Omega \geq 0$ parameterizes the level of technology and the parameter $\epsilon \in [0, 1]$ measures the elasticity of output with respect to capital. Capital depreciates at rate δ_k in every period. Given our assumptions of perfect competition in all markets and constant returns to scale in production the number of operative firms and their size is indeterminate, and without loss of generality we can assume the existence of a representative, competitively behaving firm producing according to the aggregate production function (7).

3.2 Preferences and Endowments

Households are finitely lived, with maximal life span given by J and generic age denoted by j. In each period a new age cohort is born whose size is $1 + g_n$ as large as the previous cohort, so that g_n is the constant and exogenous population growth rate. We denote by ψ_{j+1} the conditional probability of survival of each household from age j to age $j + 1$. At age $j_r < J$ households become unproductive and thus retire after age j_r.

Households have preferences defined over stochastic streams of consumption and labor $\{c_j, n_j\}$ determined by the period utility function $U(c_j, n_j)$ in (1) and the time discount factor β. They maximize expected (with respect to longevity- and idiosyncratic wage risk) lifetime utility, and are ex-ante heterogeneous with respect to the education they have acquired, a process we do not model endogenously. Let $s \in \{n, c\}$ denote the education level of a household, with $s = c$ denoting some college education and $s = n$ representing high school education (or less). The fraction of college educated households is exogenously given by ϕ_s. In addition, prior to labor market entry households draw a fixed effect[14] α from an education-specific distribution $\phi_s(\alpha)$. The wage a household faces in the labor market is given by

$$w \cdot e(j, s, \alpha, \eta)$$

where w is the aggregate wage per labor efficiency unit and $e(j, s, \alpha, \eta)$ captures idiosyncratic wage variation that is a function of the age, education status and fixed effect of the household as well as a random component η that follows an education specific first-order Markov chain with states $\eta \in \mathcal{E}_s$ and transition matrix $\pi_s(\eta'|\eta)$.

Idiosyncratic wage risk determined by the process for η and mortality risk parameterized by the survival probabilities ψ_j cannot be explicitly insured because of market incompleteness as in Bewley (1986), Huggett (1993) or Aiyagari (1994). However, households can self-insure against these risks by saving at a risk-free after-tax interest rate $r_n = r(1 - \tau_k)$. In addition to saving $a' - a$ the household spends her income, composed of earnings $z = we(j, s, \alpha, \eta)n$, capital income $r_n a$ and transfers $b_j(s, \alpha, \eta)$[15] on consumption $(1 + \tau_c)c$, including consumption taxes, and on paying labor income taxes $T(z)$ as well as payroll taxes $T_{ss}(z)$. Implicit in these formulations is that the consumption- and capital income tax are assumed to be linear, whereas the labor earnings tax is given by the potentially non-linear function $T(.)$.

The individual state variables of the household thus include (j, s, α, η, a), the exogenous age, education and idiosyncratic wage shock, as well as the endogenously chosen asset position. For given (time-invariant) prices, taxes and transfers, the dynamic programming problem of the household then reads as

$$v(j, s, \alpha, \eta, a) = \max_{c,n,a'} U(c, n) + \beta\psi_{j+1} \sum_{\eta'} \pi_s(\eta'|\eta) v(j + 1, s, \alpha, \eta', a') \tag{8}$$

subject to

$$\begin{gathered}(1 + \tau_c)c + a' + T(z) + T_{ss}(z) = (1 + r_n)a + b_j(s, \eta) + z \\ \text{with } z = we(j, s, \alpha, \eta)n\end{gathered} \tag{9}$$

[14] Both education and the fixed effect shift life cycle wage profiles in a deterministic fashion, so we could have combined them into a single fixed effect. However, when mapping the model to wage data it is more transparent to distinguish between the two components impacting the deterministic part of wages. In addition, education affects the mean age profile of labor productivity and variance of shock to it, whereas the fixed effect has no impact on these two features in the model.

[15] Transfers include social security for the retired and accidental bequests for all working households.

and subject to the borrowing limit $a' \geq 0$. This results in a value function v and policy functions c, n, a' as functions of the state (j, s, α, η, a) of a household.

3.3 Government Policy

The government uses tax revenues from labor earnings, capital income and consumption to finance an exogenous stream of government expenditures G and the interest payments on government debt B. In addition it runs a balanced-budget pay-as-you-go social security system. Finally it collects accidental bequests and redistributes them among the surviving population in a lump-sum fashion. Since the population is growing at a constant rate g_n in this economy (G, B) should be interpreted as per capita variables that are constant in a stationary recursive competitive equilibrium.

Letting Φ denote the cross-sectional distribution[16] of households, constant in a stationary equilibrium, the budget constraint of the government in a stationary recursive competitive equilibrium with population growth reads as

$$\begin{aligned} & r\tau_k \int a'(j,s,\alpha,\eta,a)d\Phi + \tau_c \int c(j,s,\alpha,\eta,a)d\Phi + \int T(we(j,s,\alpha,\eta)n(j,s,\alpha,\eta,a))d\Phi \\ = \; & G + (r - g_n)B \end{aligned} \tag{10}$$

In addition, the PAYGO social security system is characterized by a payroll tax rate τ_{ss}, an earnings threshold $\bar{z}_{ss}$ below which households pay social security taxes, and benefits $p(s, \alpha, \eta)$ that depend on the last realization of the persistent wage shock η of working age,[17] education s and the fixed effect α (which determine expected wages over the life cycle). Thus $(\tau_{ss}, \bar{z}_{ss})$ completely pin down the payroll tax function T_{ss}. The specific form of the function $p(s, \alpha, \eta)$ is discussed in Section 4.

The budget constraint of the social security system then reads as

$$\int p(s,\alpha,\eta) \cdot \mathbf{1}_{\{j>j_r\}} d\Phi = \tau_{ss} \int \min\{\bar{z}_{ss}, we(j,s,\alpha,\eta)n(j,s,\alpha,\eta,a))\} d\Phi. \tag{11}$$

Finally, we assume that accidental bequests are redistributed lump sum among the surviving working age population, and thus

$$Tr = \frac{\int (1+r^n)(1-\psi_{j+1})a'(j,s,\alpha,\eta,a)d\Phi}{\int \mathbf{1}_{\{j \leq j_r\}} d\Phi}. \tag{12}$$

so that transfers received by households are given as

$$b(j,s,\alpha,\eta) = \begin{cases} Tr & \text{if } j \leq j_r \\ p(s,\alpha,\eta) & \text{if } j > j_r. \end{cases} \tag{13}$$

16 Formally, Φ is a measure and the total mass of households of age $j = 1$ is normalized to 1.

17 This formulation has the advantage that we can capture the feature of the actual system that social security benefits are increasing in earnings during working age, without adding an additional continuous state variable (such as average earnings during the working age). Since benefits depend on the exogenous η rather than endogenous labor earnings, under our specification households do not have an incentive to massively increase labor supply in their last working period to boost pension payments.

3.4 Recursive Competitive Equilibrium (RCE)

Definition 6 *Given government expenditures G, government debt B, a tax system characterized by (τ_c, τ_k, T) and a social security system characterized by $(\tau_{ss}, \bar{z}_{ss})$, a stationary recursive competitive equilibrium with population growth is a collection of value and policy functions (v, c, n, a') for the household, optimal input choices (K, L) of firms, transfers b, prices (r, w) and an invariant probability measure Φ with the following properties:*

1. *[Household maximization]: Given prices (r, w), transfers b_j given by (13) and government policies $(\tau_c, \tau_k, T, \tau_{ss}, \bar{z}_{ss})$, the value function v satisfies the Bellman equation (8), and (c, n, a') are the associated policy functions.*

2. *[Firm maximization]: Given prices (r, w), the optimal choices of the representative firm satisfy*

$$\begin{aligned} r &= \Omega\epsilon \left[\frac{L}{K}\right]^{1-\epsilon} - \delta_k \\ w &= \Omega(1-\epsilon)\left[\frac{K}{L}\right]^{\epsilon}. \end{aligned}$$

3. *[Government Budget Constraints]: Government policies satisfy the government budget constraints (10) and (11).*

4. *[Market clearing]:*

   (a) The labor market clears:

$$L = \int e(j, s, \alpha, \eta) n(j, s, \alpha, \eta, a) d\Phi$$

   (b) The capital market clears

$$(1 + g_n)(K + B) = \int a'(j, s, \alpha, \eta, a) d\Phi$$

   (c) The goods market clears

$$Y = \int c(j, s, \alpha, \eta, a) d\Phi + (g_n + \delta)K + G$$

5. *[Consistency of Probability Measure Φ]: The invariant probability measure is consistent with the population structure of the economy, with the exogenous processes π_s, and the household policy function $a'(.)$. A formal definition is provided in Appendix C.*

3.5 Transition Paths

Our thought experiments involve unexpected changes in government tax policy that induce the economy to undergo a deterministic transition path from the initial benchmark stationary recursive competitive equilibrium to a final RCE associated with the new long-run policy. At any point in time the aggregate economy is characterized by a cross-sectional probability measure Φ_t over household types. The household value functions, policy functions, prices, policies and transfers are now also indexed by time, and the key equilibrium conditions, the government budget constraint and the capital market clearing conditions now read as

$$\begin{aligned} G + (1 + r_t)B_t =& (1 + g_n)B_{t+1} + r_t\tau_k(K_t + B_t) + \tau_c \int c_t(j, s, \alpha, \eta, a)d\Phi_t \\ &+ \int T_t(w_t e(j, s, \alpha, \eta)n_t(j, s, \alpha, \eta, a))d\Phi_t \end{aligned}$$

and

$$(1 + g_n)(K_{t+1} + B_{t+1}) = \int a'_t(j, s, \alpha, \eta, a)d\Phi_t$$

Note that, in line with the policy experiments conducted below, the labor earnings tax function T_t and government debt are now permitted to be functions of time.[18]

4 Mapping the Model into Data

Conceptually, we proceed in two steps to map the initial stationary equilibrium of our model into U.S. data. We first choose a subset of the parameters based on model-exogenous information. Then we calibrate the remaining parameters such that the initial stationary equilibrium is consistent with selected aggregate and distributional statistics of the U.S. economy.

Most of the calibration is fairly standard for quantitative OLG models with idiosyncratic risk. However, given the purpose of the paper there are two essential issues that require special attention. First, it is important that the model-generated cross-sectional earnings and wealth distribution is characterized by the same concentration at the top as in the data. We follow Castaneda et al. (2003) and augment fairly standard stochastic wage processes derived from the PSID with labor productivity states that occur with low probability, but induce persistently large earnings when they occur. This allows the model to match the high earnings concentration and the even higher wealth concentration at the top of the distribution. In addition, the explicit life-cycle structure, including a fully articulated social security system, permits us to generate a distribution of earnings and wealth at the bottom and the middle of the distribution that matches the data quite well.

Second, we ensure that the reaction of top earners to changes in the tax system is consistent with empirical estimates provided, for example, in Diamond and Saez (2011). We already argued in Section 2 that the policy elasticity of top earnings with respect to the top marginal tax rate is one key determinant of the peak of the Laffer curve of top 1% labor earnings tax payers, and hence also an important determinant of the welfare-maximizing tax policy. We will therefore calibrate the utility parameter γ so as to obtain a realistic top earnings behavior.

[18] For a complete formal definition of a dynamic equilibrium with time varying policies in an economy very close to ours, see e.g. Conesa, Kitao and Krueger (2009).

4.1 Demographics

We set the population growth rate to $g_n = 1.1\%$, the long-run average value for the U.S. Data on survival probabilities from the Human Mortality Database for the US in 2010 are used to determine the age-dependent survival probabilities $\{\psi_j\}$.

4.2 Technology

The production side of the model is characterized by the three parameters $(\Omega, \epsilon, \delta_k)$. We set the capital share to $\epsilon = 0.33$ and normalize the level of technology Ω such that the equilibrium wage rate per efficiency unit of labor is $w = 1$. The depreciation rate of capital δ_k is set such that the initial equilibrium interest rate in the economy is $r = 4\%$; this requires an annual rate of $\delta_k = 7.5\%$.

4.3 Endowments and Preferences

4.3.1 Labor Productivity

One unit of work time earns the household a wage $we(j, s, \alpha, \eta)$, where $e(j, s, \alpha, \eta)$ is the idiosyncratic labor productivity (and thus the idiosyncratic part of the wage) which depends on the age j, education s and the fixed effect α of the household as well as an idiosyncratic shock η.

We assume that $\eta \in \mathcal{E}_s$ can take on 7 education-specific values. We associate an $\eta \in \{\eta_{s,1}, \ldots, \eta_{s,5}\}$ with "normal" labor earnings observed in household data sets such as the PSID, and reserve $\{\eta_{s,6}, \eta_{s,7}\}$ for the very high labor productivity and thus earnings realizations at the top of the cross-sectional distribution, but not captured by any observations in the PSID. Log-wages are specified as

$$\ln e(j, s, \alpha, \eta) = \begin{cases} \alpha + \varepsilon_{j,s} + \eta & \text{if } \eta \in \{\eta_{s,1}, \ldots, \eta_{s,6}\} \\ \eta & \text{if } \eta = \eta_{s,7} \end{cases}$$

That is, as long as the labor productivity shock $\eta \in \{\eta_{s,1}, \ldots, \eta_{s,6}\}$, idiosyncratic wages are (in logs) the sum of the fixed effect α that is constant over the life cycle, an education-specific age-wage profile $\varepsilon_{j,s}$ and the random component η, as is fairly standard in quantitative life cycle models with idiosyncratic risk (see e.g. Conesa et al., 2009). On the other hand, if a household becomes highly productive, $\eta = \eta_{s,7}$, wages are independent of education and the fixed effect. We think of these states as representing, in a reduced form, successful entrepreneurial or artistic opportunities that yield very high earnings and that are independent of the education level and fixed effect of the household.[19]

[19] Conceptually, nothing prevents us from specifying $e(j, s, \alpha, \eta) = \exp(\alpha + \varepsilon_{j,s} + \eta)$ for $\eta = \eta_7$, but our chosen formulation provides a better fit to the earnings and wealth distributions.

We now specify the seven states of the Markov chain $\{\eta_{s,1}, \ldots, \eta_{s,7}\}$ and the transition matrices π_s. In addition we need to determine the education-specific distribution of the fixed effect $\phi_s(\alpha)$ and the deterministic, education-specific age-wage profile $\{\varepsilon_{j,s}\}$. For the latter we use the estimates from the PSID by Krueger and Ludwig (2013). Furthermore we assume that for each education group $s \in \{n, c\}$ the fixed effect α can take two values $\alpha \in \{-\sigma_{\alpha,s}, \sigma_{\alpha,s}\}$ with equal probability, $\phi_s(-\sigma_{\alpha,s}) = \phi_s(\sigma_{\alpha,s}) = 0.5$. For the "normal" labor productivity states $\{\eta_{s,1}, \ldots, \eta_{s,5}\}$ we use a discretized (by the Rouwenhorst method) Markov chain of a continuous, education-specific AR(1) process with persistence ρ_s and (conditional) variance $\sigma^2_{\eta,s}$. Thus the parameters governing this part of the labor productivity process are the education-specific variances of the fixed effect, the variances and persistence parameters $\{\sigma^2_{\alpha,s}, \sigma^2_{\eta,s}, \rho_s\}$ of the AR(1) processes, together with the share of college-educated households ϕ_s. Table 1 summarizes our choices.

Table 1: Labor Productivity Process

	ρ_s	$\sigma^2_{\eta,s}$	$\sigma^2_{\alpha,s}$	ϕ_s
$s = n$	0.9850	0.0346	0.2061	0.59
$s = c$	0.9850	0.0180	0.1517	0.41

In order to account for very high earnings realizations we augment the 5-state Markov process and its transition matrices $\pi_s = (\pi_{ij,s})$ by two more states $\{\eta_{s,6}, \eta_{s,7}\}$. The transition matrix of the extended process is given by:

$$\pi_s = \begin{bmatrix} \pi_{11,s}(1-\pi_{16,s}) & \ldots & \pi_{13,s}(1-\pi_{16,s}) & \ldots & \pi_{15,s}(1-\pi_{16,s}) & \pi_{16,s} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ \pi_{51,s}(1-\pi_{56,s}) & \ldots & \pi_{53,s}(1-\pi_{16,s}) & \ldots & \pi_{55,s}(1-\pi_{56,s}) & \pi_{56,s} & 0 \\ 0 & \ldots & 1-\pi_{66,s}-\pi_{67,s} & \ldots & 0 & \pi_{66,s} & \pi_{67,s} \\ 0 & \ldots & 0 & \ldots & 0 & 1-\pi_{77,s} & \pi_{77,s} \end{bmatrix}$$

We assume that $\pi_{16,s} = \ldots = \pi_{56,s} = \pi_{\cdot 6,s}$. Thus from each "normal" state $\{\eta_{s,1}, \ldots, \eta_{s,5}\}$ there is a (small) probability to climb to the high state $\eta_{s,6}$. The highest state $\eta_{s,7}$ can only be reached from state $\eta_{s,6}$, and households at the highest state can only fall to state $\eta_{s,6}$. If wage productivity falls back to the "normal" range, it falls to $\eta_{s,3}$ with probability 1. This transition matrix will permit us to match both the empirical earnings and wealth distribution (including at the top) very accurately. In addition, we assume that $\eta_{n,7} = \eta_{c,7}$ and $\pi_{77,n} = \pi_{77,c}$. This leaves us with ten additional parameters characterizing the labor productivity process which we summarize, along with the empirical targets, in Table 2. Appendix D gives the values of the transition probabilities and states of the Markov chains.[20]

Table 2: Earnings and Wealth Targets

Parameters		Targets
Prob. to high wage region ($s = n$)	$\pi_{\cdot 6,n}$	95-99% Earnings
Prob. to high wage region ($s = c$)	$\pi_{\cdot 6,c}$	99-100% Earnings
Persistence high shock ($s = n$)	$\pi_{66,n}$	Share college in 95-99% Earnings
Persistence high shock ($s = c$)	$\pi_{66,c}$	Share college in 99-100% Earnings
Prob. to highest wage ($s = n$)	$\pi_{67,n}$	Gini Earnings
Prob. to highest wage ($s = n$)	$\pi_{67,c}$	95-99% Wealth
Persistence highest shock	$\pi_{77,n} = \pi_{77,c}$	99-100% Wealth
High wage shock ($s = n$)	$\eta_{n,6}$	Share college in 95-99% Wealth
High wage shock ($s = c$)	$\eta_{c,6}$	Share college in 99-100% Wealth
Highest wage shock	$\eta_{n,7} = \eta_{c,7}$	Gini Wealth

4.3.2 Preferences

We assume that the period utility function is given by (1), with parameters (χ, γ, λ). The parameter χ governs the Frisch elasticity of labor supply and thus the importance of the substitution effect on labor supply when top marginal tax rates change. The parameter γ determines both the size of the income effect on labor supply from tax rate changes, as well as the benefits of social insurance.

We exogenously set $\chi = 0.6$, a medium range value for the Frisch elasticity that tries to incorporate empirical results for both men and women, see for example Keane (2011). We then calibrate γ using the following strategy: Diamond and Saez (2011), based on the simple formula discussed in Section 2,

$$\tau_{\text{Laffer}} = \frac{1}{1 + a \cdot \epsilon(z_h)},$$

argue for a revenue maximizing (and optimal) top marginal tax rate of $\tau_{\text{Laffer}} = 73\%$. Our goal is to ensure that if this formula were used to determine the optimal rate based on data generated by the steady state of our model, the resulting top marginal rate would precisely coincide with the value argued for by Diamond and Saez (2011).

The calibration for labor productivity that targets both the earnings and wealth distribu-

[20] Since in the data the share of households under the age of 30 with earnings in the top 1% is very small, we assume that only households aged 31 and older can climb up to the highest two productivity states.

tion implies a value of $a = 1.79$ for the Pareto coefficient (in the initial steady state).[21] With this value for a, a policy elasticity of $\epsilon(z_h) = 0.21$ is needed to obtain a recommended rate of 73% to reach the peak of the top 1% curve.[22] In our full dynamic general equilibrium model, we cannot provide a closed-form solution for the policy elasticity anymore.[23] Hence, we calculate the policy elasticity numerically within our model.[24] By choice of $\gamma = 1.509$ our model delivers a policy elasticity of $\epsilon(z_h) = 0.21$. A similar value for γ has been estimated by Storesletten et al. (2014) in a life-cycle model using cross-sectional data on earnings and consumption from PSID and CEX.

Finally, we choose the disutility of labor parameter λ so that households spend, on average, $\bar{n} = 1/3$ of time on market work, and the time discount factor β such that the capital-output ratio in the economy equals to 2.9.

4.4 Government Policies

The two government policies we model explicitly are the tax system and the social security system.[25] We discuss both in turn now.

4.4.1 The Tax System

We assume that the labor earnings tax function is characterized by the marginal tax rate function $T'(z)$ depicted in Figure 2. It is thus characterized by two tax rates τ_l, τ_h and two earnings thresholds $\bar{z}_l, \bar{z}_h$. As in the simple model of Section 2, earnings below $\bar{z}_l$ are not taxed and earnings above $\bar{z}_h$ are taxed at the highest marginal rate τ_h. For earnings in the interval $[\bar{z}_l, \bar{z}_h]$ marginal taxes increase linearly from τ_l to τ_h. This tax code strikes a balance between approximating the current income tax code in the U.S., being

[21] This value is within the range of values reported in the literature. While Diamond and Saez (2011) argue for $a = 1.5$ based on taxable income data (that might include other sources of income beyond labor earnings), Saez (2001) finds a value of $a = 2.0$ when looking at wage income data only.

[22] Note that a labor income- or taxable earnings elasticity between 0.20 and 0.25 is in line with early estimates by MaCurdy (1981) and quite typical in the life-cycle labor literature, see also Keane (2011). In Sections 7.2 and 7.3 we study the sensitivity of our results with the respect to the choice of this elasticity. We find that the peak of the Laffer curve is quite robust to changes in both γ and χ, the parameters that mainly govern labor supply choices.

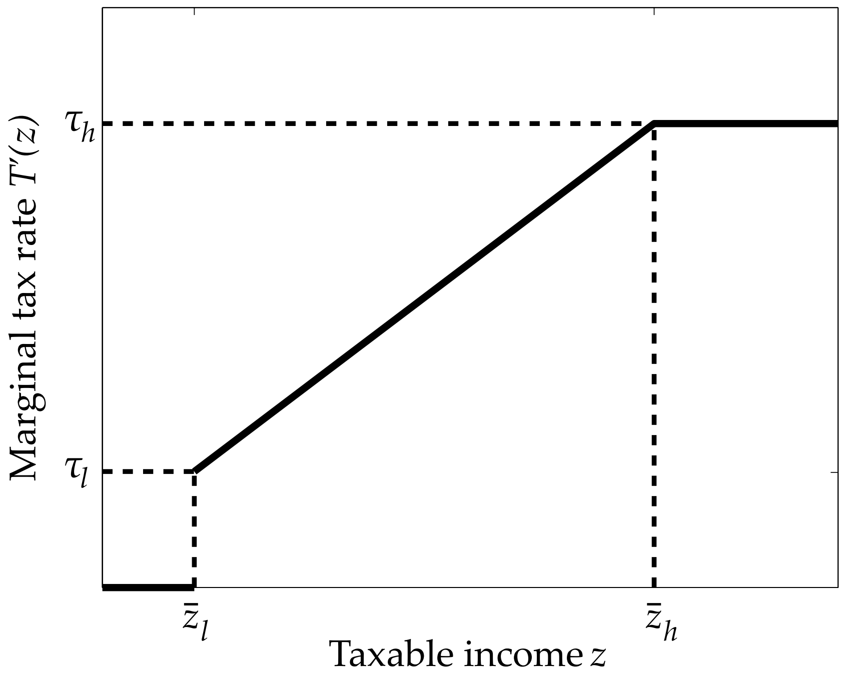
[23] In the full model there is a nondegenerate earnings distribution above the threshold $\bar{z}$, and the policy elasticity also depends on other factors beyond those delineated in Proposition 3, e.g. general equilibrium price effects, changes in other tax rates etc.

[24] To do so, we replace earnings of the single top earner z_h by the average earnings of all individuals within the top 1% earner bracket. Note that it is easy to show theoretically that the same formulas as in Proposition 1 apply in the case of a distribution of top earners. We calculate the change in top average earnings in period $t = 1$ of the transition resulting from small variations in the top 1% net-of-tax rate. We provide more details on the exact calculations in Appendix D.2.

[25] In addition the government collects and redistributes accidental bequests. This activity does not require the specification of additional parameters, however.

parameterized by few parameters and being continuously differentiable above the initial earnings threshold $\bar{z}_l$, which is crucial for our computational algorithm. Varying τ_h permits us to control the extent to which labor earnings at the top of the earnings distribution are taxed, and changing $\bar{z}_h$ controls at what income threshold the highest marginal tax rate sets in. Furthermore, if an increase in τ_h is met by a reduction of the lowest positive marginal tax rate τ_l (say, to restore government budget balance), the resulting new tax system is more progressive than the original one, as is the case in the simple model of Section 2.

Figure 2: Marginal Labor Income Tax Function



For the initial equilibrium we choose the highest marginal tax rate $\tau_h = 39.6\%$, equal to the highest marginal income tax rate of the federal income tax code prior to the 2018 federal income tax reform.[26] That tax rate applies to labor earnings in excess of 4 times average household *income*, or $\bar{z}_h = 4\bar{y}$. Households below 35% of median income do not pay any taxes,[27] $\bar{z}_l = 0.35y^{med}$ and we determine τ_l from budget balance in the initial stationary equilibrium, given the other government policies discussed below.[28] This requires $\tau_l = 11.2\%$, which lies in between the two lowest marginal tax rates of the current U.S. federal income tax code (10% and 12%).

26 This value for the highest marginal tax rate is also close to the value assumed by Diamond and Saez (2011) when Medicare taxes are abstracted from; we treat Medicare as part of the social security system.

27 In the data the income thresholds at which the lowest and highest marginal tax rates apply depend on the family structure and filing status of the household. Krueger and Ludwig (2013) argue that the value of the tax exemption and standard deduction constitute roughly 35% of median household income, fairly independent of household composition.

28 To interpret the upper income threshold $\bar{z}_h$, note that in the model about 2% of households in the initial equilibrium have earnings that exceed this threshold.

The initial proportional capital income tax rate is set to $\tau_k = 28.3\%$ and the consumption tax rate to $\tau_c = 5\%$. We choose exogenous government spending G such that it constitutes 17% of GDP; outstanding government debt B is set such that the debt-to-GDP ratio is 60% in the initial stationary equilibrium. These choices coincide with those in Krueger and Ludwig (2013) who argue that they reflect well U.S. policy prior to the Great Recession.

4.4.2 The Social Security System

We model the social security system as a flat labor earnings tax τ_{ss} up to an earnings threshold $\bar{z}_{ss}$, together with a benefit formula that ties benefits to past earnings, but without introducing an additional continuous state variable (such as average indexed monthly earnings). Thus we compute, for every state (s, α, η), average labor earnings in the population for that state, $\tilde{z}(s, \alpha, \eta)$, and apply the actual progressive social security benefit formula $f(z)$ to $\tilde{z}(s, \alpha, \eta)$. The social security benefit a household of type (s, α) with shock η_{65} in the last period of her working life receives is then given by

$$p(s, \alpha, \eta) = f(\tilde{z}(s, \alpha, \eta = \eta_{65})).$$

We discuss the details of the benefit formula $f(.)$ in Appendix D.3.

4.5 Calibration Summary

Tables 3 and 4 summarize the choice of the remaining exogenously set parameters as well as those endogenously calibrated within the model. The exogenously chosen parameters include policy parameters describing current U.S. fiscal policy, as well as the capital share in production ϵ and the preference parameter χ. The choices for these parameters are standard relative to the literature, with the possible exception of the Frisch labor supply elasticity $\chi = 0.6$, which is larger than the microeconomic estimates for white prime age males. However, it should be kept in mind that we are modeling household labor supply, including the labor supply of the secondary earner. Note that this choice implies, ceteris paribus, strong disincentive effects on labor supply from higher marginal tax rates at the top of the earnings distribution.

The set of parameters calibrated within the model include the technology parameters (δ_k, Ω), the preference parameters (β, γ, λ) as well as the entry marginal tax rate τ_l. The latter is chosen to assure government budget balance in the initial stationary equilibrium. The preference parameters are chosen so that the model equilibrium is consistent with a capital-output ratio of 2.9 and a share of time spent on market work equal to 33% of the total time endowment available to households. The technology parameters are then determined to reproduce a real (pre-tax) return on capital of 4% and a wage rate of 1, the latter being an innocuous normalization of Ω. Table 4 summarizes the associated values

Table 3: Exogenously Chosen Parameters

Parameter	Value	Target/Data
Survival probabilities $\{\psi_j\}$		HMD 2010
Population growth rate g_n	1.1%	
Capital share in production ϵ	33%	
Threshold positive taxation $\bar{z}_l$	35%	as fraction of y^{med}
Top tax bracket $\bar{z}_h$	400%	as fraction of $\bar{y}$
Top marginal tax rate τ_h	39.6%	
Consumption tax rate τ_c	5%	
Capital income tax τ_k	28.3%	
Government debt to GDP B/Y	60%	
Government consumption to GDP G/Y	17%	
Bend points b_1, b_2	0.184, 1.114	SS data
Replacement rates r_1, r_2, r_3	90%, 32%, 15%	SS data
Pension Cap $\bar{z}_{ss}$	200%	$\tau_p = 0.124$
Frisch elasticity χ	0.60	

of the parameters.[29]

Table 4: Endogenously Calibrated Parameters

Parameter	Value	Target/Data
Technology level Ω	0.920	$w = 1$
Depreciation rate δ_k	7.5%	$r = 4\%$
Initial marginal tax rate τ_l	11.2%	Budget balance
Time discount factor β	0.981	$K/Y = 2.9$
Disutility from labor λ	24	$\bar{n} = 33\%$
Coeff. of Relative Risk Aversion γ	1.509	$\epsilon(z_m) = 0.21$

[29] Even though it is understood that all model parameters impact all equilibrium entities, the discussion below associates those parameters to specific empirical targets that, in the model, impact the corresponding model statistics most significantly.

5 Characteristics of the Benchmark Economy

Prior to turning to our tax experiments we first briefly discuss the aggregate and distributional properties of the initial stationary equilibrium. This is perhaps more important than for most applications since a realistic earnings and wealth distribution, especially at the top of the distribution, is required to evaluate a policy reform that will entail potentially massive redistribution of the burden of taxation across different members of the population.

5.1 Macroeconomic Aggregates

In Table 5 we summarize the key macroeconomic aggregates implied by the initial stationary equilibrium of our model. It shows that the main source of government tax revenues are taxes on labor earnings.

Table 5: Macroeconomic Variables

Parameter	Value
Capital	288%
Government debt	60%
Consumption	58%
Investment	25%
Government Consumption	17%
Av. hours worked (in %)	33%
Interest rate (in %)	4%
Tax revenues	
- Consumption	2.9%
- Labor	11.9%
- Capital income	4.0%
Pension System	
Contribution rate (in %)	12.5%
Total pension payments	5.1%

All variables in % of GDP if not indicated otherwise

5.2 Earnings and Wealth Distribution

In this section we show that, given our earnings process with small but positive probability of very high earnings realizations, the model is able to reproduce an empirically realistic cross-sectional earnings and wealth distribution.

Table 6: Labor Earnings Distribution in Benchmark Economy

	Share of total sample (in %)								
	Quintiles					Top (%)			
	1st	2nd	3rd	4th	5th	90-95	95-99	99-100	Gini
Model	0.0	5.7	10.9	17.3	66.2	11.0	18.9	22.6	0.648
US Data	-0.1	4.2	11.7	20.8	63.5	11.7	16.6	18.7	0.636

Table 6 displays the model-implied earnings distribution and Table 7 contains the wealth distribution. When comparing the model-implied earnings and wealth quintiles to the corresponding data statistics[30] we observe that the model fits the data very well, even at the top of the distribution. The same is true for the earnings and wealth Gini coefficients.

Table 7: Wealth Distribution in Benchmark Economy

	Share of total sample (in %)								
	Quintiles					Top (%)			
	1st	2nd	3rd	4th	5th	90-95	95-99	99-100	Gini
Model	0.0	0.9	4.3	11.6	83.3	14.1	25.3	30.4	0.808
US Data	-0.2	1.1	4.5	11.2	83.4	11.1	26.7	33.6	0.816

We do not view the ability of the model to reproduce the earnings and wealth distributions as a success per se, since the stochastic wage process (and especially the two high-wage states) were designed for exactly that purpose. However, that fact that our approach is indeed successful gives us some confidence that ours is an appropriate model to study tax policy experiments that are highly redistributive across households at different parts of the earnings and wealth distribution.

6 Quantitative Results

In this section we set out our main results. We first describe the thought experiment we consider, and then turn to the optimal tax analysis. We do so in three steps. First we display top income Laffer curves, showing at what top marginal tax rate revenue from the top 1% earners is maximized, and relate our findings to the static analysis of Saez (2001) and Diamond and Saez (2011). However, revenue maximization does not imply welfare maximization in our dynamic general equilibrium model, partly because

30 As reported by Diaz-Gimenez et al. (2011), based on the 2007 Survey of Consumer Finances.

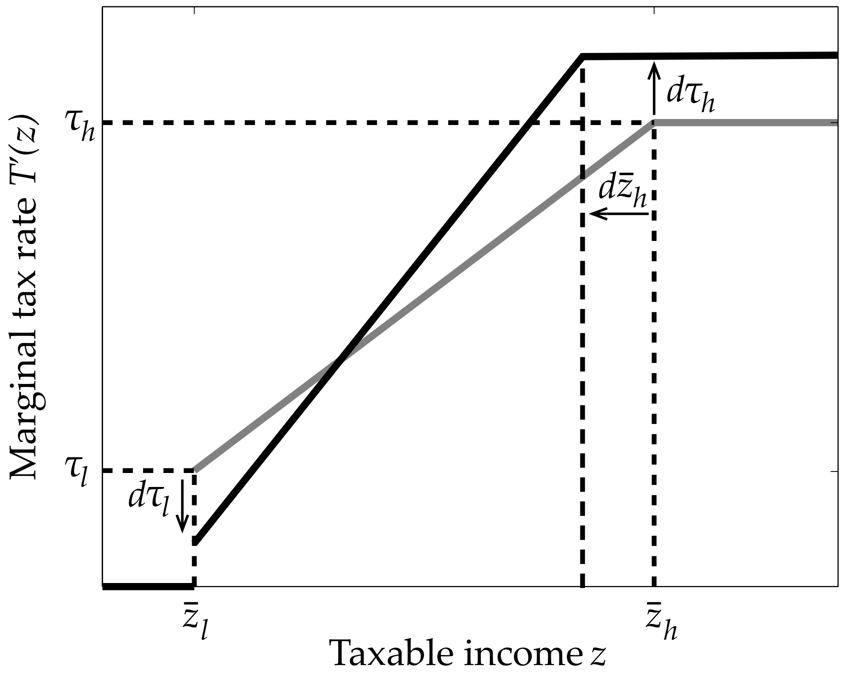
the top 1% of the population might enter social welfare, but also because their behavioral response triggers potentially important general equilibrium effects. In a second step we argue that the welfare-maximizing top marginal tax rate is lower but quantitatively fairly close to the revenue maximizing rate. In a third step we then dissect the sources of the substantial welfare gains from the optimal tax reform by a) documenting the magnitude of the adverse impact on macroeconomic aggregates of significantly raising top marginal rates, and b) quantifying the distributional benefits of such tax reforms, both in terms of enhanced ex-ante redistribution among different education and productivity groups as well as in terms of insurance against ex-post labor productivity risk. We will conclude that the significant welfare gains from increasing top marginal labor income tax rates above 80% stem primarily from enhanced insurance against not ascending to the very top of the earnings ladder, and only secondarily from redistribution across ex-ante heterogeneous households, and that these gains outweigh the macroeconomic costs (as measured by the decline in aggregate consumption) of the reform. In a last subsection we show that these conclusions are robust to alternative preference specifications of households, but that they do crucially depend on a productivity and thus earnings process that delivers the empirically observed earnings *and* wealth inequality in the data.

6.1 The Thought Experiments

We now describe our fiscal policy thought experiments. Starting from the initial steady state fiscal constitution we consider one-time, unexpected (by private households and firms) tax reforms that change the top marginal labor earnings tax rate. The unexpected reform induces a transition of the economy to a new stationary equilibrium, and we model this transition path explicitly. Given the initial outstanding debt and the change in τ_h, the government in addition permanently adjusts the entry marginal tax rate τ_l (but not the threshold $\bar{z}_l$) as well as $\bar{z}_h$ to assure both that the intertemporal budget constraint holds and that the top 1% earners are defined by the threshold $\bar{z}_h$ in the first period of the policy-induced transition path, see Figure 3 for an illustration. An appropriate sequence of government debt along the transition path ensures that the sequential government budget constraints hold for every period t along the transition.

In the aggregate, a transition path is thus characterized by deterministic sequences of interest rates, wages and government debt $\{r_t, w_t, B_{t+1}\}_{t=1}^{T}$ converging to the new stationary equilibrium indexed by a new policy $(\tau_l, \tau_h, \bar{z}_l, \bar{z}_h)$. For every period $t \geq 1$ along the transition path the analysis delivers new lifetime utilities $v_t(j, s, \alpha, \eta, a)$ of households with individual states (j, s, α, η, a). The optimal tax experiment then consists in maximizing a weighted sum of these lifetime utilities over τ_h, using adjustments in τ_l to ensure that the intertemporal government budget constraint is satisfied.

Figure 3: Laffer Curve of Labor Income Tax Receipts from Top 1%



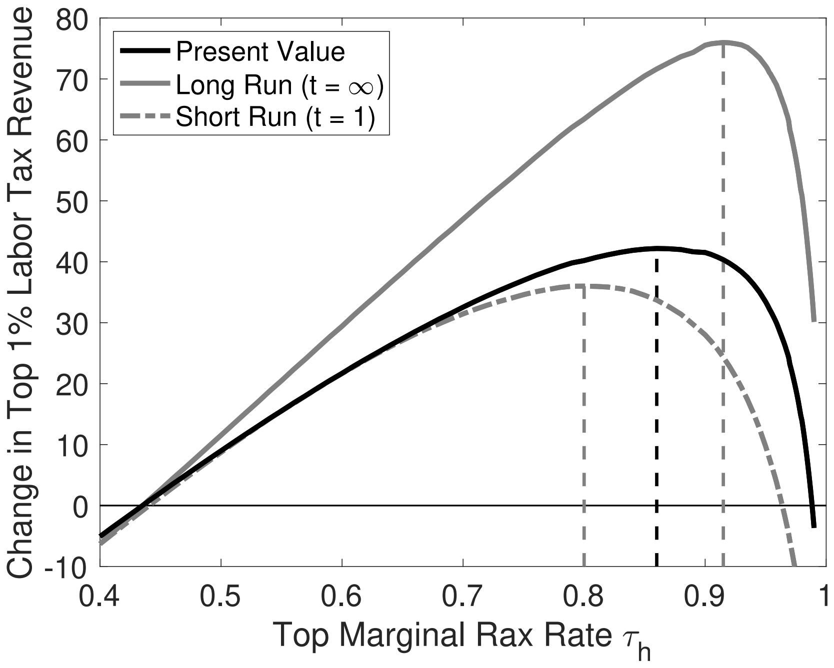
6.2 The Top 1% Laffer Curve

In Figure 4 we plot (in % deviation from the initial stationary equilibrium) labor income tax receipts from the top 1% earners against the top marginal labor income tax rate.[31] The three lines correspond to tax revenues in the first period of the transition (the "Short Run"), new steady state tax revenues (the "Long Run") and the present discounted value of tax receipts along the entire transition path (and the final steady state), where the discount rates used are the time-varying interest rates along the transition path.

From this figure we observe that the revenue maximizing top marginal tax rate, independent of the time horizon used, is very high, in excess of 80%. However, we also note that the time horizon does matter significantly: when maximizing tax revenue from top 1% earners in the short run the revenue-maximizing rate is 80% and the extra revenue that can be generated is roughly 35% higher than in the benchmark economy. As we will show, households reduce their wealth holdings along the transition, and become more *inelastic* when faced with higher top marginal tax rates. Consequently, the longer the time horizon, the higher is the revenue-maximizing top rate, and the larger are the extra revenues that can be generated by this rate. If one restricts attention solely to a steady state analysis, then the peak of the top 1% Laffer curve is attained at a tax rate of 92%, with 70% higher tax revenues from the highest income earners than in the initial stationary equilibrium. The peak of the Laffer curve when maximizing the present discounted

[31] Since in the benchmark tax system the top marginal tax rate does not apply exactly to the top 1% income earners, whereas in our tax experiments we ensure that it does, the Laffer curve does not intersect the zero line at exactly 39.6%, but rather at a slightly higher level. This is of course irrelevant for the question where the peak of the Laffer curve (and the optimal rate) is located.

Figure 4: Laffer Curve of Labor Income Tax Receipts from Top 1%



value of tax revenues, which is most informative for our welfare calculations, lies in the middle between the short- and long-run results (at a rate of 87%). Thus we deduce two main points from Figure 4: first, revenue-maximizing rates are very high, relative to the status quo. Second, and more importantly, the time horizon plays a crucial role for the quantitative results due to endogenous wealth accumulation, a finding that can only be uncovered through an explicit analysis of the transition path of a dynamic model with endogenous capital accumulation.

Revenue-maximizing tax rates need not be welfare maximizing, even when the current top 1% earners have no weight in the social welfare function. Therefore we move to an explicit characterization of socially optimal rates next. Prior to this analysis we first explore why the revenue-maximizing tax rates we find in our dynamic general equilibrium model are even higher than the 73% rate Diamond and Saez (2011) have advocated for.

6.3 Connection to Sufficient Statistics Approach

In this section, we reconnect our simulation results to the sufficient statistics approach literature from Saez (2001) and Diamond and Saez (2011). Recall that the original Saez (2001) formula is given by

$$\tau_{\text{Laffer}} = \frac{1}{1 + a \cdot \epsilon(z_h)}. \tag{14}$$

The first column in Table 8 summarizes the ingredients into this formula based on simulated statistics from our quantitative model. In our calibration the Pareto parameter a, which summarizes the ratio between average top 1% earnings and the top 1% earnings

Table 8: Sufficient Statistic vs. Numerical Simulation

	Original	Augmented Formula			
		$t=1$		$t=\infty$	
		initial	final	initial	final
a_t	1.79	1.79	1.58	1.79	1.49
$\epsilon(z_{h,t})$	0.21	0.21	0.26	0.09	0.11
$\epsilon(\tau_a(\bar{z}_t))$		-0.78	-0.60	-0.78	-0.30
$\tau_a(\bar{z}_t)$		0.27	0.36	0.27	0.51
$\tau_{\text{Laffer}}^{\text{predicted}}$	**0.73**	**0.85**	**0.80**	**1.00**	**0.92**
$\tau_{\text{Laffer}}^{\text{simulated}}$		**0.80**		**0.92**	

threshold, takes a value of 1.79, which is right in the middle between the value of 1.5 assumed by Diamond and Saez (2011) and the value of 2 used by Saez (2001). The policy elasticity, as targeted in the calibration of the preference parameter γ is 0.21, slightly lower than in Diamond and Saez. As a consequence, the peak of the Laffer curve, as predicted by the original formula, is precisely at 73%, as recommended by Diamond and Saez (2011). In other words, if our model is the true data-generating process, and if one were to base policy recommendations on (14), one would arrive exactly at their recommendation.[32]

However, as we have already pointed out in the simple model of Section 2, the tax experiment matters. Proposition 1 gave an augmented formula for the peak of the Laffer curve whenever the tax schedule is adjusted below the top 1% earnings threshold as we do in our thought experiments. Second, even with the right formula, its inputs are in general not invariant to the tax system in place, but change with adjustments in household behavior and general equilibrium factor prices along the policy induced transition of the economy to a new steady state. Endogenous adjustments in wealth accumulation over time due to changes in the top marginal rate will prove especially relevant in this regard.

The remaining columns of Table 8 demonstrate these points. They are based on the augmented formula given in Proposition 1:

$$\tau_{\text{Laffer}} = \frac{1-(a-1)\cdot\tau_a(\bar{z})\cdot\epsilon(\tau_a(\bar{z}))}{1+a\cdot\epsilon(z_h)}$$

and show the various ingredients of the formula computed from model-generated data in the first period of the transition ($t=1$) and the final steady state ($t=\infty$). The columns "initial" and "final" calculate these ingredients at the initial status quo and the final (i.e. revenue-maximizing) tax system, respectively. The additional (relative to the original formula) term $(a-1)\cdot\tau_a(\bar{z})\cdot\epsilon(\tau_a(\bar{z}))$ summarizes the effects that stem from an adjust-

[32] As described in Section 4 and Appendix D.2, this is of course how we chose γ in the first place.

ment of the tax schedule below the top 1% earnings threshold, in our case of changes in τ_l. Recall that $\tau_a(\bar{z})$ is the average tax rate of a household at the income threshold and $\epsilon(\tau_a(\bar{z}))$ is the elasticity of that rate with respect to $1 - \tau_h$.

Focusing first on the short run (columns 2 and 3), we find that at the status quo tax system (with a top marginal rate of 39.6%) this elasticity equals -0.78. That is, as the top marginal tax rate τ_h increases, so does the average tax rate at the top 1% threshold.[33] This boosts tax revenue collected from this group and hence the Laffer curve peaks at a higher rate. Based on the sufficient statistics estimates from our model, we predict a Laffer peak at $\tau_h = 85\%$, see the second column ("initial") of Table 8.

Comparing this prediction to the actual peak rate of 80% shows that the sufficient statistics formula only imprecisely predicts the peak of the Laffer curve. The third column ("final") of the table shows why. There, we summarize the sufficient statistics when the tax system features the Laffer tax rate of $\tau_h = 80\%$. The results show that these statistics are far from invariant to the very tax system under which they were calculated. With the higher top rate the Pareto parameter a drops by 0.21 (i.e. the right tail of the earnings distribution becomes fatter), the policy elasticity rises to 0.26, the average tax rate at the threshold increases by nine percentage points and its elasticity declines in absolute value. As a consequence of these changes, the augmented formula generates an estimate for the Laffer tax rate of 80 percent when the sufficient statistics are calculated under the Laffer tax system, exactly equal to the true Laffer tax rate.

The remaining two columns, which display the statistics in the long run (the eventual stationary equilibrium) demonstrate that restricting attention to a steady state analysis can be quite misleading. For a given tax system, the key distinction between the short- and the long run is that the policy elasticity $\epsilon(z_h)$ is much smaller in the long run ($\epsilon(z_h) = 0.09$) than in the short run ($\epsilon(z_h) = 0.21$). This is due to the fact that top earners enter the first period of the transition after a surprise change in tax policy with significant amount of wealth, which was accumulated under the old tax system with low top rates. A sudden increase in the top marginal tax rate leads these households to lower their labor supply substantially and finance their consumption through their wealth. This makes top 1% earnings very elastic to the marginal tax rate. In contrast, households in the new long-run equilibrium face a higher top marginal rate for their entire life-time, leading the top 1% to accumulate less wealth. The reduction of wealth dampens their labor supply reaction and, consequently, leads to a smaller labor earnings elasticity. Therefore, the predicted peak of the Laffer curve is significantly larger in the long- than in the short run, at 100% (see column 4). As in the short run, the sufficient statistics strongly depends on the tax system at which they are evaluated, with the peak rate falling to 92%.

These results serve to reinforce three points already made analytically in the simple

[33] This is the net of two opposing effects on $\tau_a(\bar{z})$. First, the entry tax rate τ_l declines, which reduces the tax burden on all individuals above the entry threshold. On the other hand, the upper earnings threshold $\bar{z}$ declines to subject exactly all top 1% households to the top marginal tax rate. This second effect leads marginal tax rates to increase faster with income, and therefore causes an increase in $\tau_a(\bar{z})$. In our numerical simulations the latter effect dominates and thus $\tau_a(\bar{z})$ is negative.

model of Section 2. First, as the difference between the first and the second (and fourth) column suggests, the changes in the tax code below the top rate financed with the extra revenue from higher top rates are crucial determinants of the peak rate. Second, the distinction between the short-run transition and the long-run steady state analysis is quantitatively very important for the determination of the top marginal rate, at least as long as wealth accumulation is endogenous. Finally, while the optimal tax formulae based on the sufficient statistics approach work well based on model-simulated data, these statistics respond strongly to the tax system in place. Therefore, in our view, these formulae are useful primarily for describing the forces that govern revenue-maximizing (and possibly optimal) rates, rather than for prescriptive purposes when these prescriptions are based on statistics emerging from current tax systems.

6.4 Welfare-Maximizing Tax Rates

After having discussed the revenue implications from increasing top marginal tax rates we now turn to our analysis of *socially optimal* rates. To do so we now describe how we measure social welfare.

6.4.1 Measuring Social Welfare

The welfare measure we employ is constructed as follows. After solving for the equilibrium path of a specific tax reform, we calculate the amount of initial wealth transfers needed to make an individual indifferent between the status quo and the policy reform, ex post for the currently living and ex ante for future generations.[34] These transfers $\Psi_0(j, s, \alpha, \eta, a)$ satisfy, for currently alive individuals,

$$v_1(j, s, \alpha, \eta, a + \Psi_0(j, s, \alpha, \eta, a)) = v_0(j, s, \alpha, \eta, a)$$

where v_0 denotes the value function in the initial steady state. For households born in period $t \geq 1$ we find the number Ψ_t such that

$$Ev_t(j = 1, s, \alpha, \bar{\eta}, \Psi_t) = Ev_0(j = 1, s, \alpha, \bar{\eta}, 0)$$

where expectations are taken with respect to initial fixed effect and education. Note that a positive Ψ constitutes a welfare *loss* from a given reform, relative to the status quo. The total present discounted value of all transfers is then given by

$$W = \int \frac{\Psi_0(j, s, \alpha, \eta, a)}{1 + r_0} \, d\Phi_0 + \mu_1 \sum_{t=1}^{\infty} \left(\frac{1 + g_n}{1 + r_0} \right)^t \Psi_t.$$

[34] These wealth transfers induce behavioral responses which we capture when computing the transfers. However, we abstract from the general equilibrium effects these hypothetical transfers induce. For future cohort the transfer is one number per cohort, for currently alive households the transfers differ by characteristics (j, s, α, η, a). Future transfers are discounted at rate $\frac{1+n}{1+r_0}$ where r_0 is the interest rate in the initial stationary equilibrium and our aggregate welfare measure is the sum of these transfers.

When top 1% households are excluded from the social welfare function only transfers to the bottom 99% of the current earnings distribution are included in the calculations.

In order to turn this wealth-based welfare measure into a consumption flow measure we express the present discounted value of the transfers as an annuity C that pays a constant flow of consumption through the transition and in the new steady state, and express the size of this annuity as a percent of initial aggregate consumption. That is, we calculate

$$C \sum_{t=0}^{\infty} \left(\frac{1+g_n}{1+r_0} \right)^t = -W$$

Recall that if the transfers W are positive, this signals welfare losses from the reform, negative W mean welfare gains. We express welfare gains in percent of initial consumption

$$CEV = 100 * C/C_0.$$

This idea of calculating the welfare consequences of policy reforms follows closely that of Huang et al. (1997) or Benabou (2002), and more generally, the hypothetical lump-sum redistribution authority originally envisioned by Auerbach and Kotlikoff (1987).[35] [36]

Finally, we also compute and report a steady state welfare measure that asks what uniform (over time and across states) percentage increase in consumption a household born into the old steady state, under the veil of ignorance, prior to the realization of the education level s and fixed effect α, would need to receive to be indifferent to being born into the steady state associated with a new policy.[37]

6.4.2 Optimal Size of the Top Marginal Earnings Tax Rate

In this section we document the optimal top marginal labor earnings tax rate. In Figure 5 we plot two welfare measures against the top marginal tax rate τ_h. The black line plots the aggregate welfare measure CEV, whereas the gray line instead displays "steady state welfare" as described in the previous subsection.

The optimal top marginal tax rate is indeed very high, around 80% under both welfare measures. Welfare CEV including the top 1% households, and including the transition

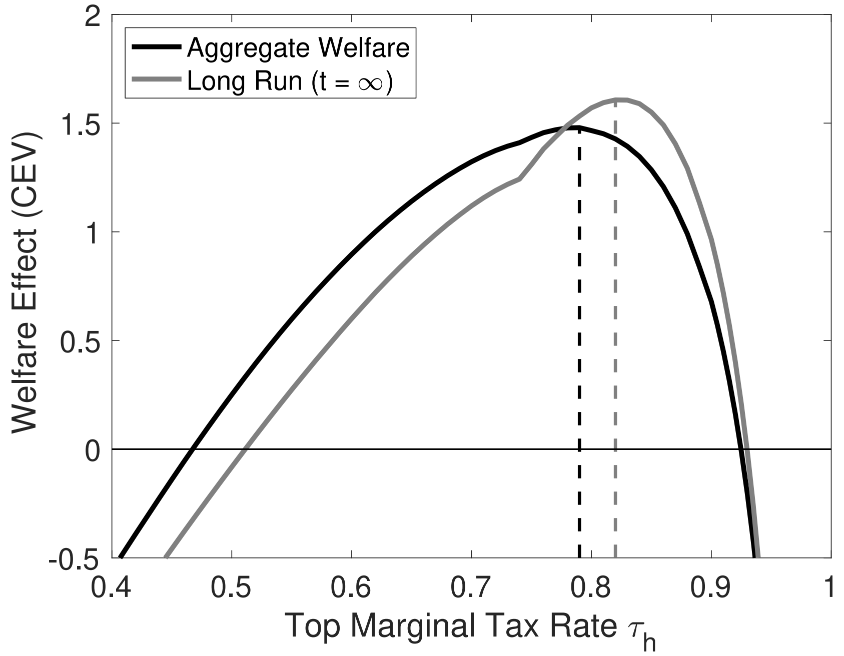
---

[35] Whereas Benabou (2002) evaluates aggregate efficiency by calculating a certainty equivalent consumption sequence for each individual and then summing it across individuals and over time, we determine the wealth equivalent of changes in the life cycle allocation of consumption and labor supply for each individual and then sum across households. The advantage of both of these closely related approaches over using social welfare functions is that both Benabou's (2002) as well as our measure separates aggregate efficiency considerations from the potential desire of the policy maker (as built into the social welfare function) to engage in intergenerational or intragenerational redistribution.

[36] Fehr and Kindermann (2015) show that, to a first approximation of the value function, maximizing our welfare measure is equivalent to maximizing the weighted sum of (remaining) lifetime utilities, with weights given by the inverse of the marginal utility of wealth in the value function, or equivalently (by the envelope theorem) the inverse of the marginal utility of current consumption.

[37] Conesa et al. (2009) employ the same long-run welfare measure in their study of optimal capital taxes.

Figure 5: Three Aggregate Welfare Measures as Functions of τ_h



effects is hump-shaped and maximized at $\tau_h = 79\%$. Recall that the top marginal tax rate that maximizes the present discounted value of tax revenues from the top 1% earners is 87%, higher than this welfare-optimal rate, but quantitatively close.[38] Focusing exclusively on welfare in the long run the optimal top marginal rate is even larger, at $\tau_h = 82\%$. Note that the welfare gains induced by these high marginal tax rates are very substantial, in the order of 1.5% of permanent consumption. In these thought experiments, as we vary τ_h we adjust the upper threshold $\bar{z}_h$ above which the highest marginal tax rate applies so that in the first period of the transition the top 1% earners face this rate. The government intertemporal budget constraint is balanced by adjusting the entry marginal rate τ_l, holding fixed the lower bend point $\bar{z}_l$.[39]

6.5 Understanding the Welfare Gains

In order to understand the reported welfare gains from the optimal tax reform we proceed in two steps. First we display the transition paths of key macroeconomic variables that the tax reform induces, documenting the significant adverse consequences on out-

[38] Since this welfare measure include short- and long-run welfare effects, a comparison with the present discounted value Laffer curve is most informative.

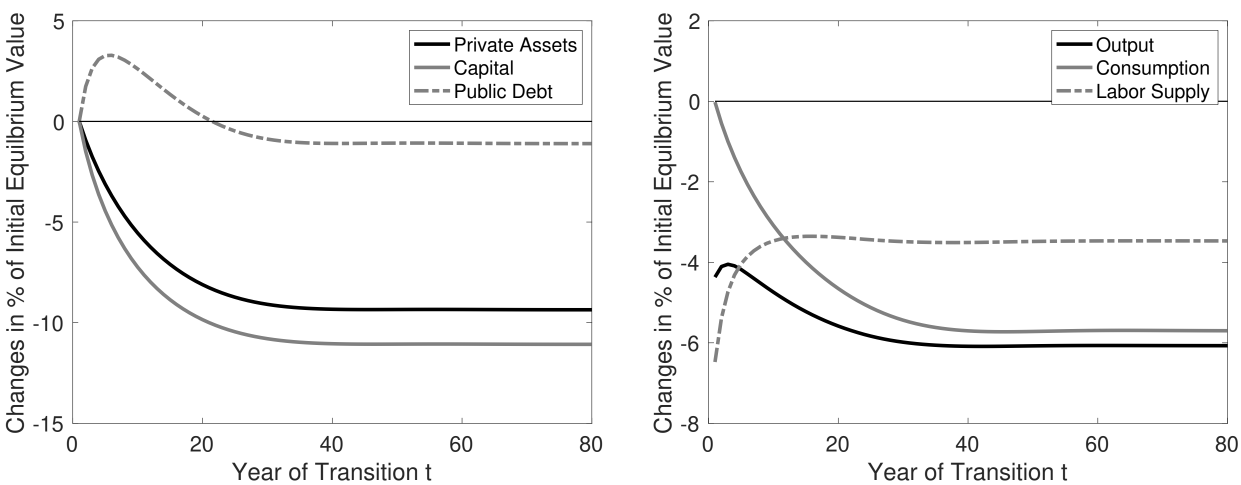
[39] If the required τ_l is non-negative, all households with earnings below $\bar{z}_l$ pay zero taxes, if τ_l is negative, all households with earnings below $\bar{z}_l$ receive a subsidy of τ_l per dollar earned, akin to the Earned Income Tax Credit in the U.S. This slight asymmetry about how income below $\bar{z}_l$ is treated induces a small kink in the welfare plot when τ_l turns from positive to negative. This is of course irrelevant for the determination of the optimal tax code, as the kink occurs far to the left of the optimal τ_h.

put, *aggregate* consumption and the capital stock in the economy. Second, we quantify the redistributive and insurance benefits of the reform, arguing that the latter are crucial for understanding our welfare results.

6.5.1 The Dynamics of Aggregates Along the Transition

In Figure 6 we plot the evolution of key macroeconomic aggregates along the transition from the old to the new stationary equilibrium. All variables are expressed in % deviations from their initial steady state values. Figure 7 displays the transition path of hours worked, separately for the bottom 99% and the top 1% of the earnings distribution, as well as the time path of wages and interest rates in the economy. Finally, Figure 8 shows how revenues for consumption, labor income, and capital income taxes as well as pre-tax earnings and wealth inequality (as measured by the Gini coefficient) evolve over time.

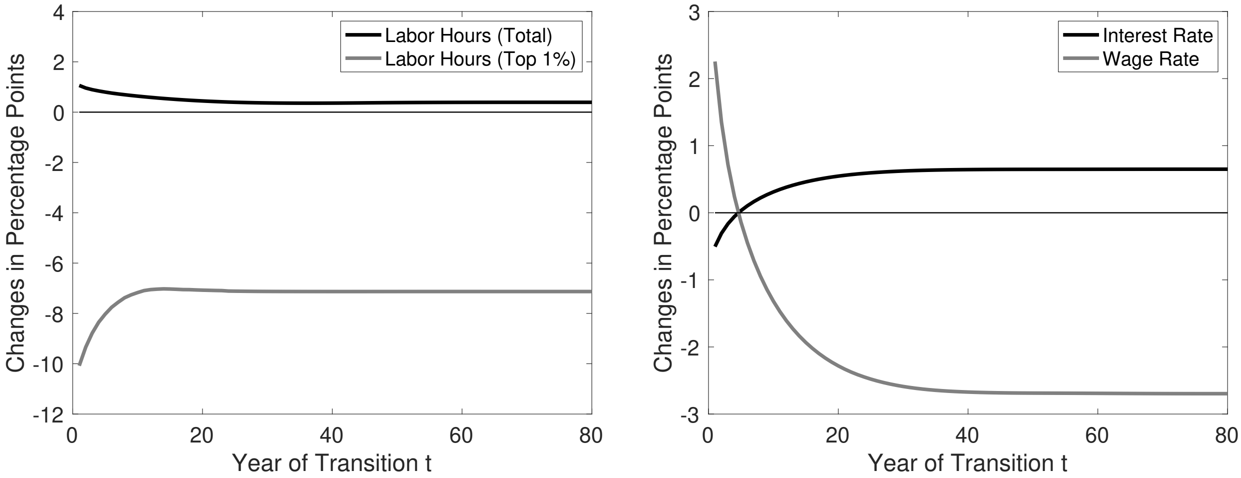
Figure 6: Capital, Assets, Gov. Debt; Labor Supply, Consumption, Output along Transition



The right panel of Figure 6 shows that on impact the massive increase in marginal tax rates at the top of the earnings distribution leads to a contraction of labor input by close to 7% and a corresponding fall of output by 4% (since capital is predetermined and thus fixed in the short run). The left panel of Figure 7 indicates that the collapse in labor input is entirely due to the reduction in hours worked by the highly productive top 1% of the earnings distribution, whose hours fall on average by 10 percentage points. Thus even though this group is small, because of their massive behavioral response and their high relative productivity this 1% of earners drives down aggregate labor input substantially. The ensuing partial recovery is owed to wages rising above initial steady state levels temporarily (see the right hand panel of Figure 7) as the capital-labor ratio falls early in the transition. Furthermore, over time the top group reduces its wealth holdings: a negative wealth effect on leisure (positive wealth effect on labor supply) results.

In the medium run the capital stock falls significantly, partially being crowded out by higher public debt used to finance the tax transition, but mainly driven by the decline in private saving of the high earners that are now subject to a significantly higher marginal

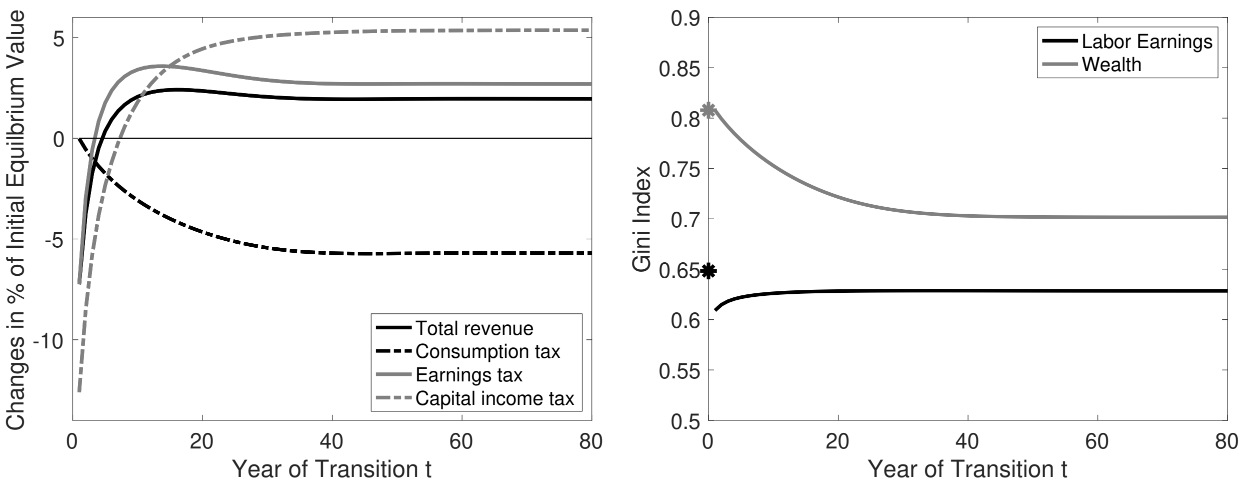
Figure 7: Hours and Prices along Transition



(and thus average) labor earnings tax rate under the new tax system. Whereas in the short run most of the loss in output is absorbed by lower investment, in the long run aggregate consumption declines strongly as well, by about 6% (right panel of Figure 6).

The left panel of Figure 8 displays the evolution of tax revenue along the transition. Even though overall economic activity falls in response of the tax reform, tax revenues of the government decline only temporarily which in turn explains the temporary increase in government debt present in Figure 6. The composition of tax revenue changes substantially as well. Since aggregate consumption falls, so does revenue from taxing it. On the other hand, once hours of the top 1% earners have partially recovered, labor income tax revenues increase, on account of the significantly higher taxes these individuals pay. In fact, in the long run this group accounts for close to 80% of all revenue from the labor earnings tax. Revenues from capital income taxes also rise due to the higher return a lower capital-labor ratio implies, despite the decline in the capital income tax base.

Figure 8: Tax Revenues and Inequality along Transition



Finally, the right panel of Figure 8 shows that the tax reform leads to a reduction of both earnings and wealth inequality. The Gini index for pre-tax labor earnings falls significantly on impact, reflecting primarily the decline in hours worked and thus earnings of the top 1% earners. As hours of this group partially recover, so does earnings inequality, without reaching its pre-reform level. Wealth inequality, on the other hand, is monotonically and very substantially declining over time as the lower labor earnings of the households at the top of the distribution translates into lower wealth holdings of that group and thus a lesser net worth concentration in the population. In the long run, the wealth Gini is 10 percentage points lower than under the benchmark tax system, indicating that when a wide labor earnings distribution is the main culprit for high wealth inequality, tackling earnings inequality with high marginal earnings taxation at the top is an effective tool for curbing wealth inequality. This is, of course, not an explicit policy goal of the government, but rather a side effect of its desire to provide social insurance and ex-ante redistribution, as discussed in Section 6.5.2.

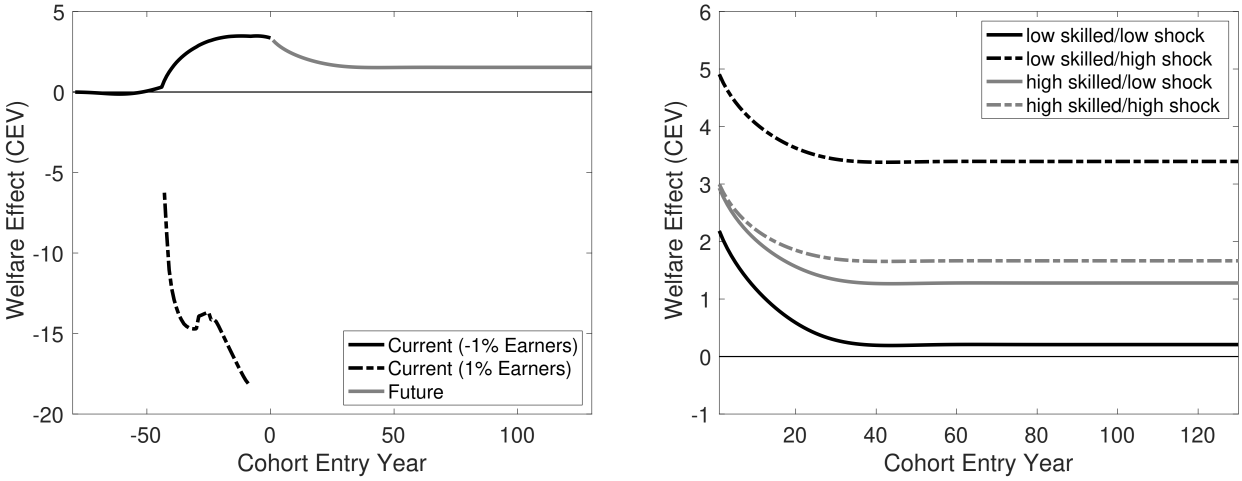
To summarize, the aggregate statistics indicate a massive decline in aggregate output and a somewhat delayed fall in aggregate consumption, coupled with a reduction of hours worked at the top of the earnings distribution. Furthermore, earnings and wealth inequality are significantly lower under the tax system featuring very high marginal tax rates at the top. These aggregate statistics suggest that the sources of the welfare gains from the tax reform documented in Section 6.4.2 come from enhanced social insurance and redistribution rather than from stimulating aggregate economic activity. In the next section we will provide a decomposition to argue that the main source of the welfare gains along the transition, but especially in the new steady state, comes from better consumption insurance rather than more ex-ante redistribution under the new tax system with high marginal tax rates at the top. These insurance benefits offset the aggregate consumption losses, since these losses accrue exclusively to those few households that rise to the very top of the earnings distribution.

6.5.2 Ex-Ante Redistribution or Ex-Post Insurance?

In order to understand why the optimal tax system implies substantial welfare gains despite its adverse impact on macroeconomic aggregates we first display the welfare consequences from the tax reform for households with different characteristics. The left panel of Figure 9 plots these gains against the age of a household cohort; all cohorts to the left of zero on the x-axis are already alive at the time of the reform, everyone to the right is born into the transition. For cohorts currently alive we distinguish between welfare for the top 1% earners (in the initial steady state) and welfare of the rest, always aggregated as discussed in Section 6.1.

The welfare impact of the reform on the top 1% earners currently alive is very strongly negative, whereas the reform has very little impact on current retirees (the cohorts economically born 45 years prior to the reform or earlier). For current non-top earners the welfare gains are larger the younger they are since younger workers spend a larger share of their working life under the new tax regime. Finally, the welfare impact of future gen-

Figure 9: Welfare Effects of Policy Reform by Age Cohort



erations is positive, in the order of 1.5-2.5% of lifetime consumption. It falls along the transition as the economy consumes part of its capital stock.[40]

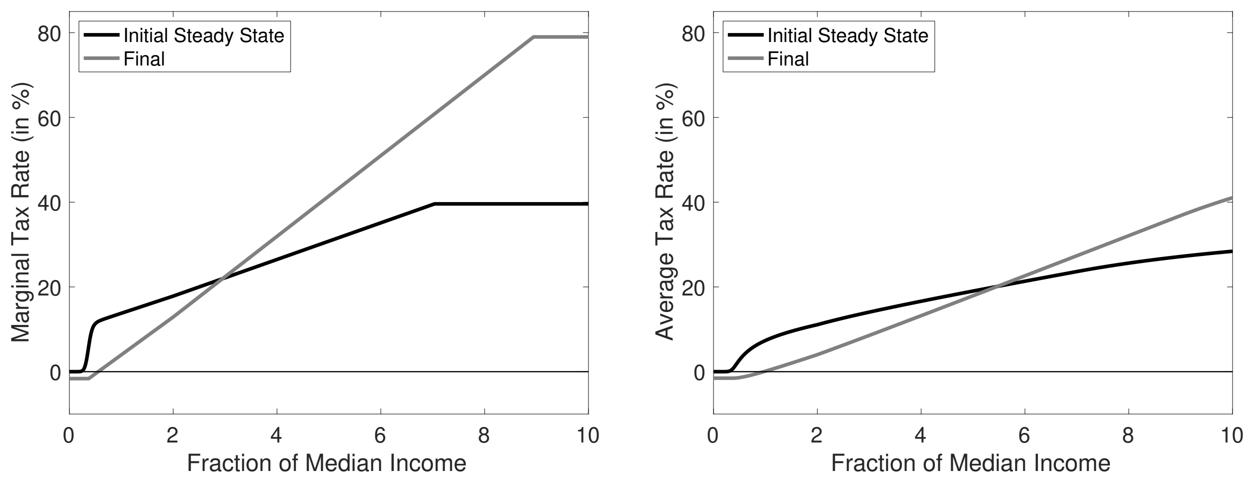
The right panel of Figure 9 focuses on generations born after the implementation of the reform, but takes an ex-post (after household type has been realized) perspective by disaggregating welfare gains from the tax reform by household type. Recall that our economy is populated by households that differ by education and by a productivity fixed effect. Thus a total of four ex-ante heterogeneous household types are born in every transition period. The right panel of Figure 9 displays the lifetime welfare gain from the reform for each of these types. We make three observations: First, and consistent with the left panel, for all types the welfare gains are declining over time, reflecting the reduction in aggregate consumption induced by a fall in the aggregate capital stock. Second, the welfare consequences are substantially positive for *all four* household types, clarifying that the welfare gains do not stem primarily from socially beneficial redistribution towards low-skilled households. Third, the welfare gains display considerable heterogeneity across the types. Notably, the welfare gains of one group, the low-skilled with high fixed effect, is significantly larger than the gains of the other groups.

To understand this last finding, it is instructive to display how marginal and average tax rates changes between the benchmark and the optimal tax system. Figure 10 plots marginal (left panel) and average (right panel) tax rates against labor earnings in the initial and the final steady state.

It shows that households with up to about six times median earnings in the initial steady state face lower average taxes whereas high earners face massively higher marginal and

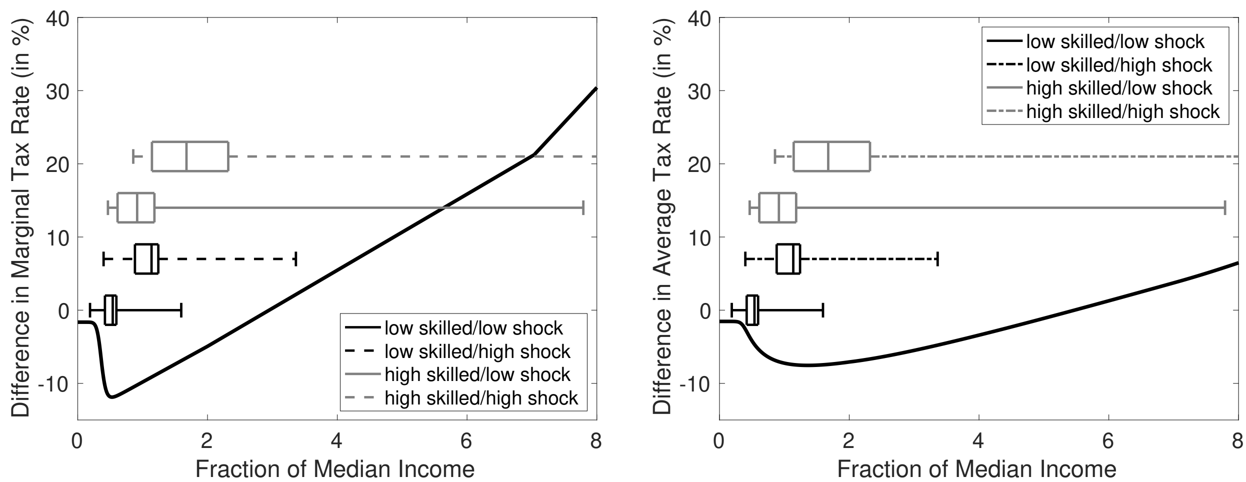
[40] The aggregate welfare measures in Section 6.4.2 aggregated the welfare impact of all current and future generations, and thus is a convex combination of small welfare gains of retired households, large welfare losses of the current top 1% (if included in welfare), sizable welfare gains for current working age households, and substantial welfare gains of future generations. The steady state welfare gains in contrast only capture the large gain of future generations, and thus display a larger benefit from the tax reform than the welfare measures that include transitional generations.

Figure 10: Marginal Tax Schedules, Average Tax Schedule: Benchmark and Optimum



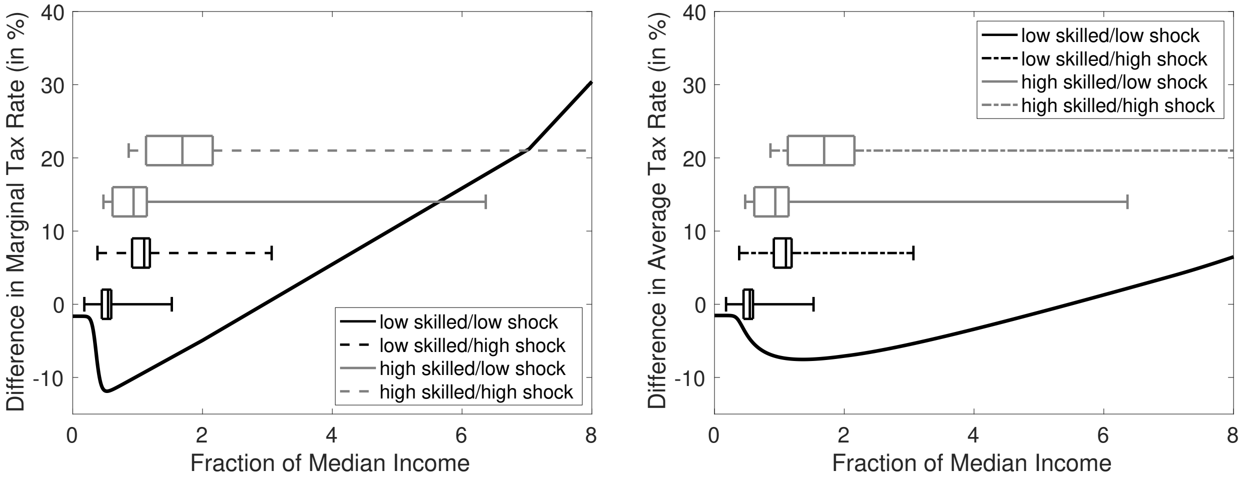
average taxes. In Figures 11 and 12 we display the differences in marginal and average tax rates between the two tax codes in conjunction with box plots to summarize the earnings distribution in the model in the initial (Figure 11) and final (figure 12) steady state. As our model is populated by four ex-ante heterogeneous types who differ by their education and earnings fixed effect, each panel includes four box plots associated with the earnings distribution of each of the four types. The box in the middle contains 50% of the probability mass, with median earnings of the group represented by the vertical line in the middle of the box. The ends of the box plots give the positions of the 2.5%-tile and the 97.5%-tile of the earnings distribution.

Figure 11: Difference in Tax Schedules and Earnings Distribution (by Type) in Initial Steady State



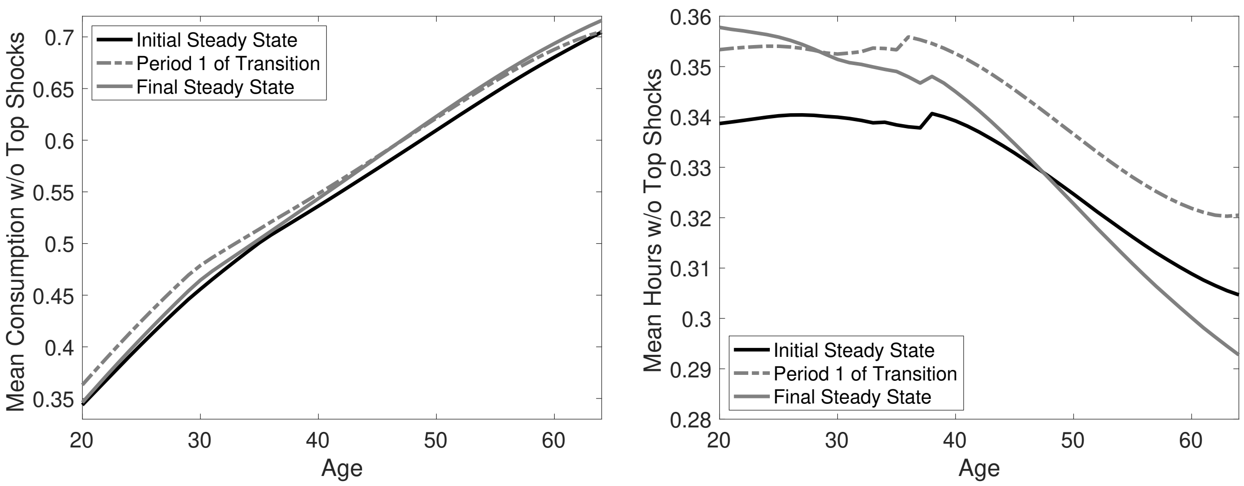
We make three main observations. First, the overwhelming majority of households is located in parts of the earnings distribution that faces lower average and marginal tax rates under the optimal, relative to the benchmark tax system. Second, the earnings distributions shift to the left between Figure 11 and Figure 12, indicating a decline in overall *pre-tax* labor earnings induced by the tax reform. Third, the largest reduction in marginal

Figure 12: Difference in Tax Schedules and Earnings Distribution (by Type) in Final Steady State



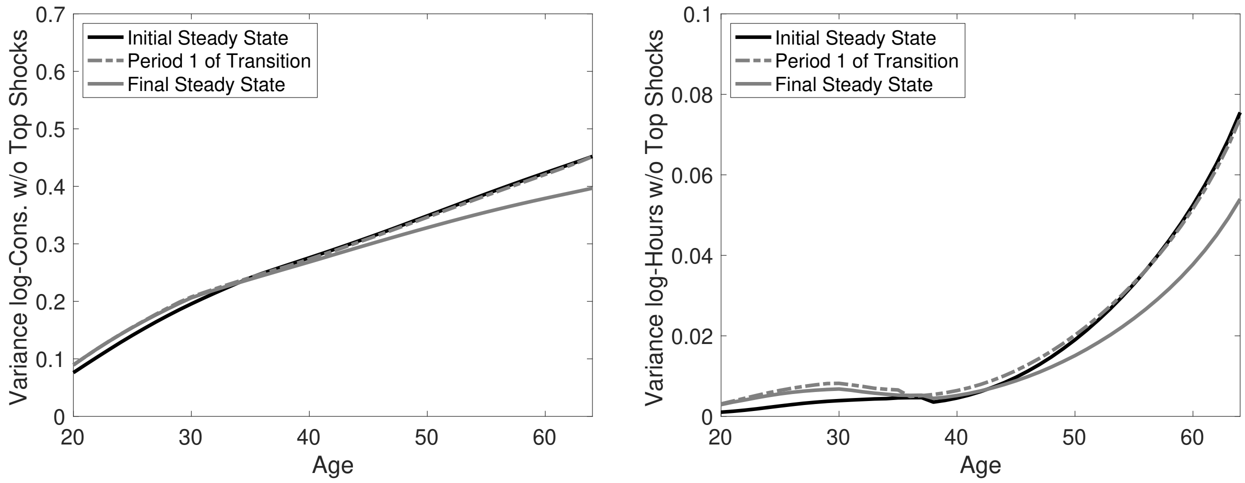
and especially average tax rates occurs among the middle class, households with earnings between 50% and 200% of median income. This naturally makes the low-skilled, high fixed effect group and the high-skilled, low fixed effect group the largest beneficiaries of the reform, see the box plots of these two groups. The main difference between these two groups is that high-skilled (college) households have a nontrivial chance of rising to the very top of the earnings distribution (where they are hurt by the high marginal tax rates), whereas the low-skilled face an essentially zero change of experiencing the same fate; compare the location of the 97.5%-tile of the earnings distribution for each of the two groups. This combination – middle class earnings in expectation and almost no chance of becoming very earnings rich – makes this group benefit dis-proportionally from the proposed tax reform.

Figure 13: Average Consumption and Hours over the Life Cycles, w/o Top 2 Shocks



The previous discussion does not clarify what are the *common* sources of the welfare gains of each of these four groups. To identify these sources, in Figure 13 we plot mean consumption and hours worked over the life cycle, not counting consumption and hours occurring when households have one of the two high labor productivity shocks (that is, roughly, excluding hours and consumption of the top 1%). Figure 14 does the same for the variance of consumption of hours, and Figures 15 and 22 in Appendix E repeat the same for the entire population, that is, now *including* the top productivity states in the calculation of means and variances.

Figure 14: Variance of Consumption and Hours over the Life Cycles, w/o Top 2 Shocks

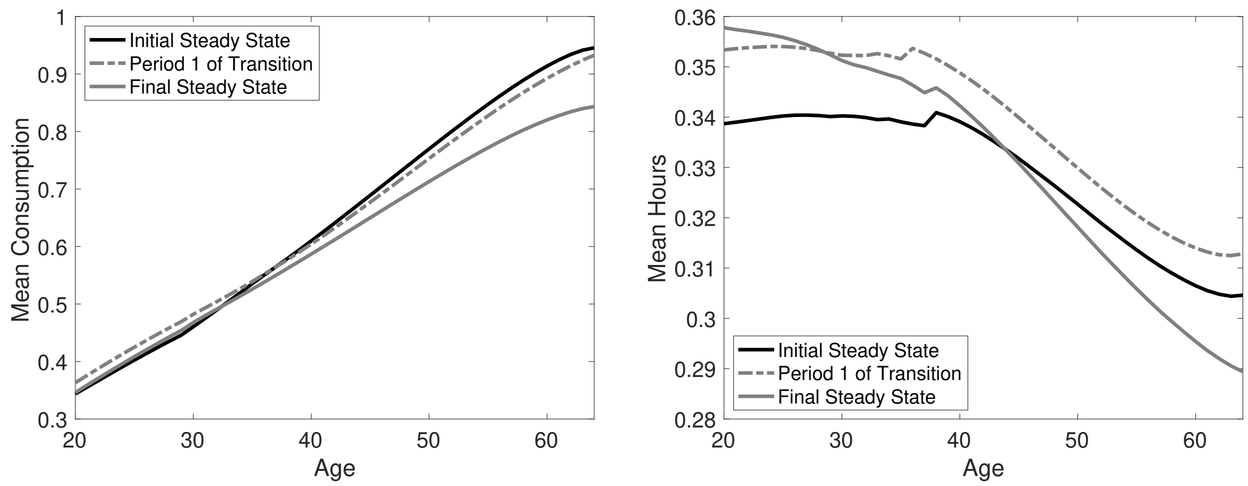


The key observation comes from comparing Figures 13 and 15. Average consumption of households outside the top 1% is uniformly larger over the life cycle under the new, relative to the old tax system (comparing steady states), despite the fact that aggregate consumption is 6% lower. As Figure 15 shows, the reduction of consumption is heavily concentrated among older household at the top of the earnings distribution. In addition, hours worked remain roughly constant in the new, relative to the old steady state. Coupled with a sizeable reduction of lifetime consumption risk, approximated by the within-cohort consumption variance, see the left panel of Figure 14 or 22, the 1.5% steady state welfare gains emerge.

7 Sensitivity Analysis

In this section we discuss the sensitivity of our results to the key parameter choices we have made so far. The next subsection explores the importance of the size and persistence of the labor productivity process producing top income earners in the model, and Section 7.2 summarizes sensitivity analyses with respect to the key preference parameters governing the elasticity of earnings with respect to taxes. Details on how we adjust the model to produce the results in this section are in Appendix F.

Figure 15: Average Consumption and Hours over the Life Cycles, Entire Population



7.1 The Productivity Process Generating Top Income Earners

The key quantitative model ingredient to generate the very high earnings at the top of the income distribution and the even more concentrated wealth distribution, is the presence of high and persistent labor productivity states. It it well known that this model element is sufficient to generate these distributions. However, we have argued here that the implied desire to provide social insurance against not ever becoming an earnings superstar or falling back to normal earnings provides a rationale for very high marginal tax rates on these top earners being optimal. We now show that in the absence of this model element the implications for optimal tax rates at the top change dramatically.

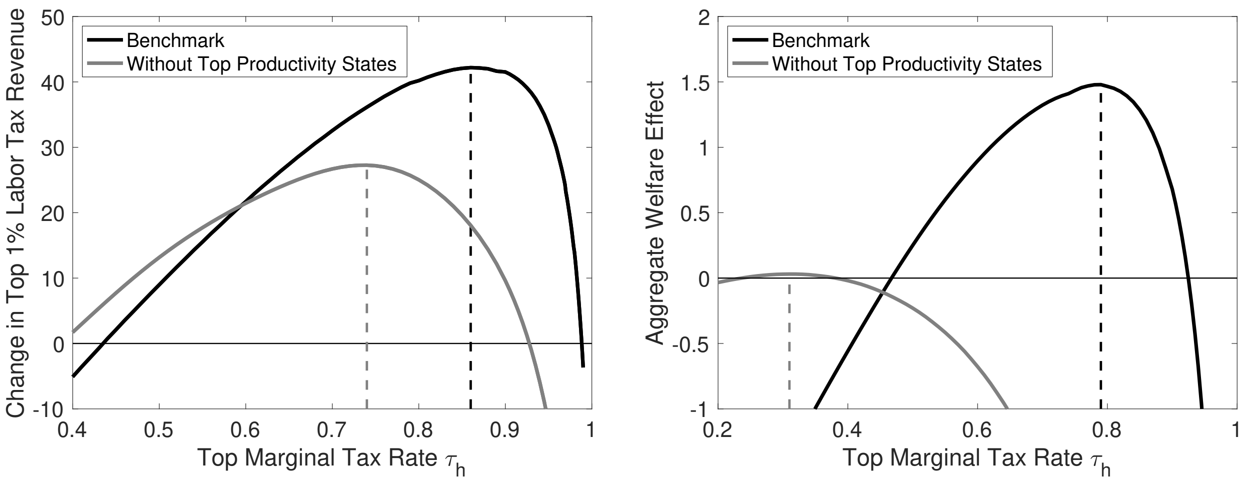
7.1.1 The Model without Superstars

Suppose first that households face a labor productivity process that does *not* contain the small chance of very high wage and thus earnings realizations.[41] In this version of the model the earnings, income and wealth distributions do not display the degree of concentration observed in U.S. data, and thus the model does not paint an accurate picture of the economic circumstances of the top 1%.

Figure 16 displays the top 1% Laffer curve (left panel) and welfare (right panel). As the figure shows, in the absence of the top two productivity shocks, and thus in the absence of a realistic degree of earnings and wealth dispersion, the optimal top marginal labor earnings tax rate falls and is fairly close to the current U.S. top rate. This happens for two reasons. First, the revenue-maximizing top marginal tax rate falls to 74% (rather than above 87%, as in the benchmark economy), on account of a smaller income effect of the

[41] One interpretation of this economy is that it describes the 1960's and early 1970's, the period prior to the large increase in the top 1% income share. Hsu and Yang (2013) study steady state optimal (piece-wise) linear income taxation in an infinite horizon model very similar to this economy.

Figure 16: Laffer Curves and Welfare as Function of τ_h, Absent Top Productivity Shocks



now less-earnings rich top 1% that makes labor supply more elastic to the top marginal tax rate. Most importantly, now the divergence between the revenue-maximizing (from the top 1%) top tax rate (above 70%) and welfare maximizing top rate (below 40%) is much more significant. Since the largest productivity realizations are now much less severe, the large social insurance benefit of high tax progressivity vanishes. Thus, our main result of very high marginal tax rates for top earners depends crucially on a productivity process capable of producing earnings- and wealth-rich households as in the data.

7.1.2 Persistence of High Productivity States

To what extent do our results depend on the fact that the large productivity shocks are persistent, but far from permanent (and thus a progressive tax system provides both insurance against the risk of never becoming highly productive and becoming unproductive again after a spell of stardom)? To answer this question we model permanent superstars, but we remain consistent with our benchmark model in which the probability of becoming very productive is essentially zero before age 30. Specifically, we proceed as follows.

In the benchmark model, starting at age 30, household may receive the high productivity shock η_6 and subsequently the superstar shock η_7 according to the Markov transition matrix specified in the calibration section[42]. In constrast, we now assume that at age 30 a share of households randomly but *permanently* draws shocks η_6 and η_7. These shares are chosen such that the share of households with these productivities in the population are the same as in the benchmark model. In this way we vary the persistence of the superstar states (by making it permanent) without changing the cross-sectional productivity distribution, relative to the benchmark model. Note that whereas the earnings Gini remains close to its empirical counterpart, the wealth Gini falls from 0.81 to 0.74 and the wealth

[42] Which implies a good chance of reverting back to the normal part of the productivity distribution.

share of the top 1% decreases from 30% to 21% in the model with permanent superstars.

Figure 17: Laffer Curve and Aggregate Welfare, Persistent vs. Permanent Highest Wage Shocks

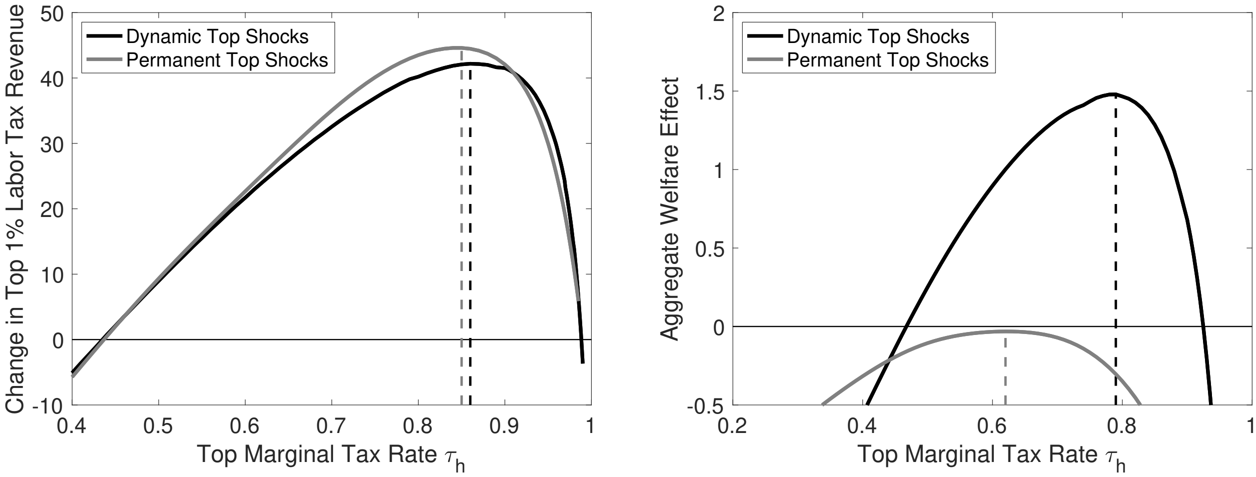
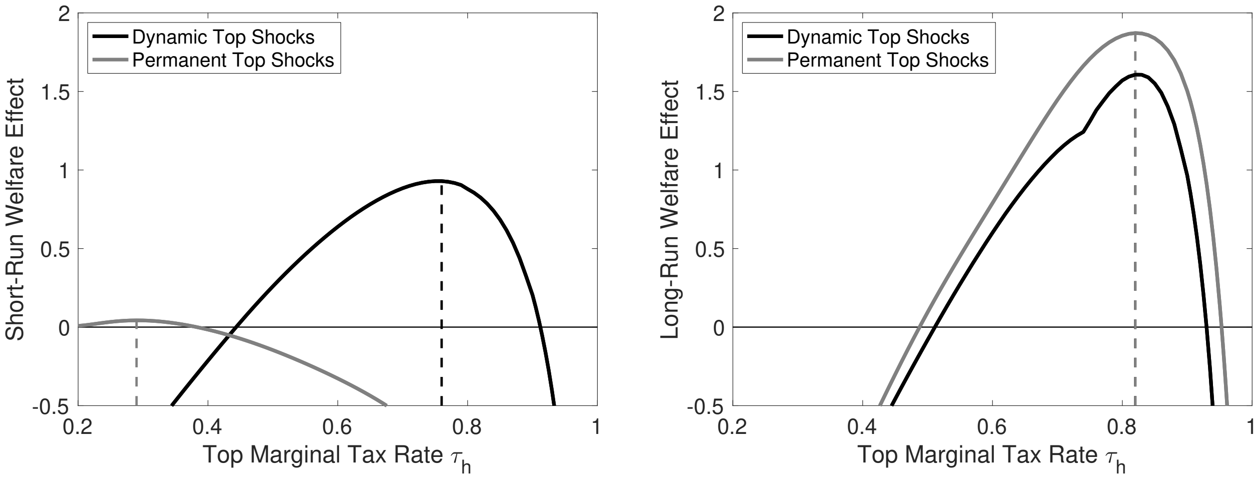


Figure 18: Short- and Long-Run Welfare Effects, Persistent vs. Permanent Highest Wage Shocks



The main change relative to the benchmark model (and consistent with the previous section) is that with permanent top income shocks, the gap between the revenue-maximizing and the welfare-maximizing top rate is significantly larger. Specifically, the welfare-maximizing tax rate is significantly smaller with permanently high productivity states. Effectively, now being an earnings superstar is a permanent trait, and a high marginal tax rate on these individuals (with associated lower rates on everyone else) no longer provides useful insurance against reverting back to the lower part of the earnings distribution.

Interestingly, the short- and long-run welfare consequences of a high marginal rates are pointedly different, comparing both panels of Figure 18. For future generations high marginal rates on the top provide social insurance against not becoming a permanent earnings superstar, just as in the benchmark economy. In fact, the long-run welfare results are very similar in both versions of the model. However, in the initial period of the transition, who is permanently earnings-rich is already determined among the alive generations, and thus moderate welfare gains of those not in the highest earnings states are completely offset by massive losses of the permanently top 1% households who now face higher marginal rates and do not benefit from social insurance against falling back down in the earnings distribution. Consequently, and in contrast to the benchmark model, high marginal rates are suboptimal and lead to sizeable aggregate welfare losses

7.1.3 Evidence on Top Income Earners

The previous section has argued that mean reversion of earnings at the very top of the distribution is crucial for our optimal tax results. There is significant empirical support for this assumption. For example, Guvenen et al. (2014) investigate data on the top 1% and the top 0.1% of wage earners from the social security administration. They estimate the likelihood that a top earner remains in the same earnings bracket in the following year as well as over a five year horizon. Their data display a significant extent of transitions in and out of the top earnings brackets. In the 2000s, an individual in the top 0.1% earner bracket only had about a 57% chance of staying there in the next year (40% over the next 5 years). An individual in the next 0.9% only stays within this group of earners with a probability of 65% (46% over a five year horizon.) In our calibration, the probability of remaining in the very high productivity state for another year (71% over one year, 18% over five years) strikes a balance between these short-run and long-run estimates.[43]

7.2 Labor Supply Elasticity

As most clearly seen in the simple model in Section 2 the revenue-maximizing and optimal top marginal tax rate depend on the parameters governing the elasticity of labor supply with respect to tax rates. Therefore we now conduct sensitivity analyses with respect to the Frisch elasticity parameter χ governing the size of the substitution effect as well as of risk aversion γ which also controls the size of the income effect on labor supply, as well as the importance of the social insurance benefits progressive taxes have.

In Table 9 and Figure 23 in Appendix E we document how our optimal tax and welfare results depend on the Frisch labor supply elasticity. The key finding is that, although the positive and normative results change in the expected direction (a larger elasticity reduces the size of the top marginal tax rate and the associated welfare gains from the policy reform), the differences are quantitatively fairly small. Even with a household-level Frisch labor supply elasticity of 1.5, arguably at the upper bound of empirical estimates the optimal top marginal tax rate exceeds 70%.

As shown analytically in Section 2.3 and quantitatively in Section 6.3 the policy elasticity

[43] Auten et al. (2013) use tax return data and adminstrative records in the IRS Compliance Data Warehouse to document that for individuals who were in the top 1% group of tax payers in 2005, 65% were still top 1% tax payers in 2006 and only 27% in 2010, confirming the results of Guvenen et al. (2014).

Table 9: Sensitivity with Respect to Frisch Labor Supply Elasticity

Scenario	τ_h	τ_l	K	L	LR Wel.	Agg Wel.
Frisch elasticity = 0.25	83%	-0.2%	-8.7%	-2.5%	1.6%	1.4%
Benchmark = 0.60	79%	-1.6%	-11.1%	-3.5%	1.5%	1.5%
Frisch elasticity = 1.50	74%	-3.9%	-12.8%	-4.3%	1.6%	1.7%

of labor supply, and thus the optimal tax rate is strongly affected not only by the substitution effect, but also by the income effect of households at the very top of the earnings distribution. We now document how changes in its magnitude affect our results. To this end we change the parameter governing income effects, γ, from 1.509 to 1 (log-utility) in consumption, making our preference specification consistent with balanced growth. A smaller value of γ implies smaller income effects and thus stronger responses of labor supply at the top to changes in marginal tax rates.

Figure 19: Laffer Curve and Aggregate Welfare, $\gamma = 1.509$ vs. $\gamma = 1$

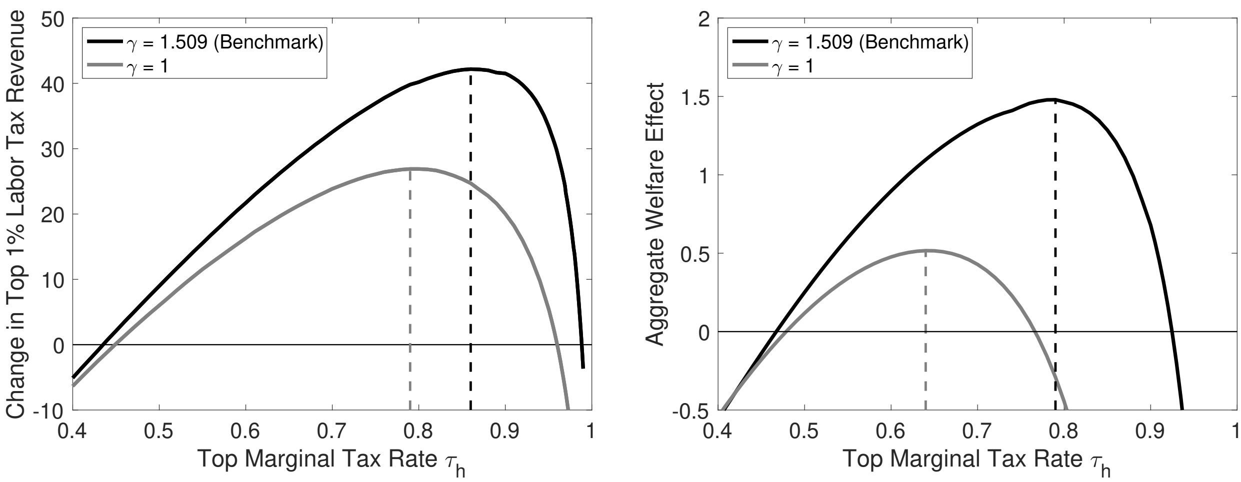


Figure 19 plots the top earner Laffer curve (the present discounted value version), both for risk aversion of $\gamma = 1$ and $\gamma = 1.509$, whereas the right panel does the same for aggregate welfare. We observe that the magnitude of the income effect is quantitatively important for our findings, but that the key result (top marginal tax rate significantly above current levels) remains unaffected. With log-utility the *revenue maximizing* top rate is 79% and the welfare maximizing rate is 64%.

Through the lens of the sufficient statistics approach, a value of $\gamma = 1$ produces a short-run policy elasticity of $\epsilon(z_{m,1}) = 0.30$ and a long-run elasticity of $\epsilon(z_{m,\infty}) = 0.23$. Both values are significantly higher than the original ones shown in Table 8. The Pareto parameter rises to $a = 1.89$. Thus a smaller income effect increases the elasticity of aggregate top earnings with respect to the top marginal net-of-tax rate. As a consequence, the revenue-maximizing tax rate falls from 0.87 to 0.79. Second, the divergence between revenue maximization and welfare maximization again becomes more important as lower risk aversion shrinks the insurance benefits of highly progressive labor income taxes. Thus the socially optimal top rate is even lower. Yet, it still remains at a sizable 64%, substantially higher than the current values in the U.S. We think of the parameter configuration with log-utility as delivering a plausible lower bound for what the top marginal tax rate should be, since logarithmic utility implies a risk aversion at the low end of commonly used values and leads to a high elasticity of earnings with respect to taxes at the upper bound of empirical estimates.[44]

Overall, we conclude from our sensitivity analysis that variations in preference parameters within empirically plausible bounds leave our main conclusions intact, whereas a labor productivity process with *persistent but not permanent* "superstar states" is crucial, in the context of our model, for generating both an empirically plausible income and wealth distribution, as well as the high optimal marginal tax rates on these superstars.

8 Conclusion

In this paper we have numerically characterized the optimal marginal earnings tax rate τ_h faced by the top 1% of the cross-sectional earnings distribution. We found it to be very high, in the order of 80%, fairly independently of whether the top 1% is included or excluded in the social welfare function. We have argued that such high marginal tax rates provide optimal social insurance in a world where very high labor incomes are generated by rare, but persistent earnings opportunities, coupled with endogenous, and fairly elastic, labor supply choices of households.

The crucial model ingredient that generates realistic earnings and wealth inequality is a policy-invariant labor productivity process where individuals with small probability receive very high realizations, and these realizations are mean reverting but persistent. Given the centrality of this assumption for our result, important next steps of inquiry are to empirically assess for which share of earners at the very top of the distribution such an abstraction is plausible. Sports and entertainment stars as well as some entrepreneurs are likely well-described by our model, whereas high earnings professionals for whom long-term human capital investment decisions are crucial entry tickets into the Top 1% are likely not. Furthermore, it would be interesting to conduct the same tax reform analysis in other models known to be able to generate a realistic earnings and wealth distribution, such as the model of entrepreneurial choice of Quadrini (1997), Cagetti and De Nardi (2006), or the human capital model analyzed in Badel et al. (2020).

[44] As an important additional distinction, with log preferences the correlation between hours worked and labor productivity is positive whereas in the benchmark that correlation was slightly negative.

References

[1] Aiyagari, R. (1994), "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109, 659-684.

[2] Alvaredo, F., A. Atkinson, T. Piketty and E. Saez (2013), "The Top 1 Percent in International and Historical Perspective , " *Journal of Economic Perspectives*, 27, 3-20.

[3] Auerbach, A. and L. Kotlikoff (1987), *Dynamic Fiscal Policy*, Cambridge University Press.

[4] Auten, G., Gee, G. and N. Turner (2013), "Income Inequality, Mobility, and Turnover at the Top in the US, 1987?2010,"*American Economic Review*, 103, 168-172.

[5] Badel, A. and M. Huggett (2017), "The Sufficient Statistic Approach: Predicting the Top of the Laffer Curve," *Journal of Monetary Economics*, 87, 1-12.

[6] Badel, A., M. Huggett and W. Luo (2020), "Taxing Top Earners: A Human Capital Perspective," *forthcoming: Economic Journal*.

[7] Bakis, O., B. Kaymak and M. Poschke (2015), "Transitional dynamics and the optimal progressivity of income redistribution, " *Review of Economic Dynamics*, 18, 679-693.

[8] Bassetto, M. (2014), "Optimal Fiscal Policy with Heterogeneous Agents" *Quantitative Economics*, 5, 675-704.

[9] Benabou, R. (2002), "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency, " *Econometrica*, 70, 481-517.

[10] Bewley, T. (1986), "Stationary Monetary Equilibrium With a Continuum of Independently Fluctuating Consumers," in: W. Hildenbrand and A. Mas-Colell (eds.): *Contributions to Mathematical Economics in Honor of Gerard Debreu*, North-Holland, Amsterdam, 79-102.

[11] Brüggemann, B. and J. Yoo (2015), "Aggregate and Distributional Effects of Increasing Taxes on Top Income Earners," *Working Paper*.

[12] Brüggemann, B. (2019), "Higher Taxes at the Top: The Role of Entrepreneurs, " *Working Paper*.

[13] Cagetti, M. and M. De Nardi (2006), "Entrepreneurship, Frictions and Wealth," *Journal of Political Economy*, 106, 835-870.

[14] Castaneda, A., J. Diaz-Gimenez and V. Rios-Rull (2003), "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111, 818-857

[15] Chamley, C. (1986), "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 54, 607-622.

[16] Conesa, J. and D. Krueger (2006), "On the Optimal Progressivity of the Income Tax Code," *Journal of Monetary Economics*, 53, 1425-1450.

[17] Conesa, J., S. Kitao and D. Krueger (2009), "Taxing Capital: Not a Bad Idea after All!," *American Economic Review*, 99, 25-48.

[18] Diamond, P. (1998),"Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83-95.

[19] Diamond, P. and E. Saez (2011), "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165-190.

[20] Diaz-Gimenez, J., A. Glover and V. Rios-Rull (2011), "Facts on the Distributions of Earnings, Income, and Wealth in the United States: 2007 Update," *Federal Reserve Bank of Minneapolis Quarterly Review*, 34 (1), 1-31.

[21] Domeij, D. and J. Heathcote (2004), "On the Distributional Effects of Reducing Capital Taxes," *International Economic Review*, 45, 523-554.

[22] Erosa, A. and M. Gervais (2002), "Optimal Taxation in Life Cycle Economies," *Journal of Economic Theory*, 105, 338-369.

[23] Fehr, H. and F. Kindermann (2015), "Optimal Taxation with Current and Future Cohorts," *Journal of Economicy Dynamics & Control*, 51, 64-77.

[24] Guner, N., M. Lopez-Daneri and G. Ventura (2016), "Heterogeneity and Government Revenues: Higher Taxes at the Top?" *Journal of Monetary Economics*, 80, 69-85.

[25] Guvenen, F., Kaplan, G. and J. Song (2014), "The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners, 1981?2012,"*NBER Working Paper* 20560.

[26] Guvenen, F. and G. Kaplan (2017), Top Income Inequality in the 21st Century: Some Cautionary Notes," *Federal Reserve Bank of Minneapolis Quarterly Review*, 38 (1), 1-15.

[27] Hendren, N. (2016), "The Policy Elasticity," *Tax Policy and the Economy*, 30, 51-89.

[28] Holter, H., D. Krueger and S. Stepanchuk (2019), "How Does Tax Progressivity and Household Heterogeneity Affect Laffer Curves?," *Quantitative Economics*, 10, 1317-1356.

[29] Hsu, M. and C.C. Yang (2013), "Optimal Linear and Two-Bracket Income Taxes with Idiosyncratic Earnings Risk," *Journal of Public Economics*, 105, 58-71.

[30] Huang, H., T.J. Sargent and S. Imrohoroglu (1997), "Two Computations to Fund Social Security," *Macroeconomic Dynamics*, 1, 7-44.

[31] Hubmer, J., P. Krusell and A. Smith (2020) "Sources of U.S. Wealth Inequality: Past, Present, and Future," forthcoming in *NBER Macroeconomics Annual*, 35.

[32] Huggett, M. (1993), "The Risk-free rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics & Control*, 17, 953-969.

[33] Jacobs, B. and D. Schindler (2012), "On the Desirability of Taxing Capital Income in Optimal Social Insurance," *Journal of Public Economics*, 96, 853-868.

[34] Jones, C. and J. Kim (2018), "A Schumpeterian Model of Top Income Inequality," *Journal of Political Economy*, 126, 1785-1826.

[35] Judd, K. (1985), "Redistributive Taxation in a Simple Perfect Foresight Model," *Journal of Public Economics*, 28, 59-83.

[36] Keane, M.P. (2011), "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49, 961-1075.

[37] Krueger, D. and A. Ludwig (2013), "Optimal Progressive Labor Income Taxation and Education Subsidies When Education Decisions and Intergenerational Transfers are Endogenous," *American Economic Review*, 103, 496-501.

[38] Krusell, P. and Smith, A. (1998), "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867-896.

[39] MaCurdy, T. E. (1981), "An empirical model of labor supply in a life-cycle setting," *Journal of Political Economy*, 89, 1059-1085.

[40] Mirrlees, J. (1971), "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38, 175-208.

[41] Piketty, T. (2014), *Capital in the Twenty-First Century*. Harvard University Press, Cambridge, MA.

[42] Piketty, T. and E. Saez (2003), "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 118, 1-39.

[43] Quadrini, V. (1997), "Entrepreneurship, Saving and Social Mobility," *Review of Economic Dynamics*, 3, 1-40.

[44] Reich, R. (2010), *Aftershock*. Random House, New York, NY.

[45] Saez, E. (2001), "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205-229.

[46] Stantcheva, S. (2020), "Dynamic Taxation," forthcoming in *Annual Review of Economics*.

[47] Storesletten, K., J. Heathcote and G.L. Violante (2014), "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104, 1-52.

[48] Vogelgesang, U. (2000), "Optimal Capital Income Taxation and Redistribution," *Finanzarchiv*, 57, 412-434.

A Appendix: Proofs of Propositions

A.1 Proof of Proposition 1

We start from the definition of top 1% labor earnings tax revenue

$$T(\tau_h) = \tau_h(z_h - \bar{z}) - R,$$

which we can, for the purpose of notation, also write as

$$T(\tau_h) = T(\bar{z}) + \tau_h(z_h - \bar{z}),$$

with $T(\bar{z}) = -R$. Total differentiation yields

$$dT(\tau_h) = dT(\bar{z}) + d\tau_h(z_h - \bar{z}) + \tau_h dz_h.$$

By some rearranging, we obtain

$$\begin{aligned} dT(\tau_h) = & \frac{d\left[\frac{T(\bar{z})}{\bar{z}}\right]}{d(1-\tau_h)} \cdot \frac{(1-\tau_h)}{T(\bar{z})/\bar{z}} \cdot \frac{T(\bar{z})}{\bar{z}} \cdot \frac{\bar{z}}{1-\tau_h} \cdot d(1-\tau_h) \\ & - (z_h - \bar{z})d(1-\tau_h) + \frac{dz_h}{d(1-\tau_h)} \cdot \frac{1-\tau_h}{z_h} \cdot \frac{\tau_h z_h}{1-\tau_h} \cdot d(1-\tau_h). \end{aligned}$$

With the definitions as in Proposition 1, we immediately get

$$\frac{dT(\tau_h)}{d(1-\tau_h)} = \epsilon(\tau_a(\bar{z})) \cdot \frac{\tau_a(\bar{z})\bar{z}}{1-\tau_h} - (z_h - \bar{z}) + \epsilon(z_h)z_h \cdot \frac{\tau_h}{1-\tau_h}.$$

The peak of the Laffer curve can then be found by setting $\frac{dT(\tau_h)}{d(1-\tau_h)} \overset{!}{=} 0$, which yields

$$\epsilon(\tau_a(\bar{z})) \cdot \frac{\tau_a(\bar{z})}{1-\tau_h} - \left(\frac{z_h}{\bar{z}} - 1\right) + \epsilon(z_h) \cdot \frac{z_h}{\bar{z}} \cdot \frac{\tau_h}{1-\tau_h} = 0.$$

Using $\frac{z_h}{\bar{z}} = \frac{a}{a-1}$ and solving for τ_h gives us

$$(a-1) \cdot \epsilon(\tau_a(\bar{z})) \cdot \tau_a(\bar{z}) - (1-\tau_h) + \epsilon(z_h) \cdot a \cdot \tau_h = 0$$

from which we immediately get

$$\tau_{\text{Laffer}} := \tau_h = \frac{1 - (a-1) \cdot \epsilon(\tau_a(\bar{z})) \cdot \tau_a(\bar{z})}{1 + a \cdot \epsilon(z_h)}.$$

□

A.2 Proof of Proposition 3

The optimization problem of a household with high labor productivity e_h reads

$$\max_{c_h, n_h} \frac{c_h^{1-\gamma}}{1-\gamma} - \lambda \frac{n_h^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}} \quad \text{s.t.} \quad c_h = e_h n_h - \tau_h(e_h n_h - \bar{z}) + R.$$

The first-order conditions of this problem read

$$c_h^{-\gamma} = \mu \quad \text{and} \quad \lambda n_h^{\frac{1}{\chi}} = \mu(1-\tau_h)e_h,$$

where μ is the Lagrange multiplier on the budget constraint. Combining these equations with the budget constraint yields the labor supply equation

$$n_h e_h - \tau_h(n_h e_h - \bar{z}) + R - \left[\frac{(1-\tau_h)e_h}{\lambda}\right]^{\frac{1}{\gamma}} n_h^{-\frac{1}{\gamma\chi}} = 0.$$

Uncompensated labor supply elasticity Total differentiation with respect to e_h yields

$$\left\{(1-\tau_h)z_h + \frac{1}{\gamma\chi}\left[z_h - \tau_h(z_h - \bar{z}) + R\right]\right\}\frac{dn_h}{n_h}$$
$$+ \left\{(1-\tau_h)z_h - \frac{1}{\gamma}\left[z_h - \tau_h(z_h - \bar{z}) + R\right]\right\} = 0.$$

Rearranging leads to

$$\begin{aligned}\epsilon_h^u = \frac{dn_h}{de_h}\cdot\frac{e_h}{n_h} &= \frac{-(1-\tau_h)z_h + \frac{1}{\gamma}\left[z_h - \tau_h(z_h - \bar{z}) + R\right]}{(1-\tau_h)z_h + \frac{1}{\gamma\chi}\left[z_h - \tau_h(z_h - \bar{z}) + R\right]} \\ &= \frac{(1-\gamma)(1-\tau_h)z_h + \tau_h\bar{z} + R}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{\tau_h\bar{z}+R}{\chi}}.\end{aligned}$$

Income elasticity of labor supply Total differentiation of the labor supply equation with respect to R yields

$$\left\{(1-\tau_h)z_h + \frac{1}{\gamma\chi}\left[z_h - \tau_h(z_h - \bar{z}) + R\right]\right\}\frac{dn_h}{n_h} + dR = 0$$

which immediately gives

$$\begin{aligned}\eta_h = \frac{dz_h}{dR}(1-\tau_h) &= \frac{e_h\cdot dn_h}{dR}(1-\tau_h) \\ &= \frac{-(1-\tau_h)z_h}{(1-\tau_h)z_h + \frac{1}{\gamma\chi}\left[z_h - \tau_h(z_h - \bar{z}) + R\right]} \\ &= \frac{-\gamma(1-\tau_h)z_h}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{\tau_h\bar{z}+R}{\chi}}.\end{aligned}$$

Policy elasticity Before taking the total differential with respect to our tax reform, it is useful to formulate the labor supply equation in labor earnings terms as

$$z_h - T(\bar{z}) - \tau_h(z_h - \bar{z}) - \left[\frac{(1-\tau_h)}{\lambda}\right]^{\frac{1}{\gamma}} \cdot e_h^{\frac{1}{\gamma}\left[1+\frac{1}{\chi}\right]} z_h^{-\frac{1}{\gamma\chi}} = 0.$$

Note that we again use the notation $T(\bar{z}) = -R$ from Appendix A.1. Total differentiation with respect to the policy experiment then yields

$$\begin{aligned} & dz_h - dT(\bar{z}) - d\tau_h(z_h - \bar{z}) - \tau_h dz_h \\ & \qquad - \frac{1}{\gamma} \cdot \frac{d(1-\tau_h)}{1-\tau_h} \cdot \left[\frac{(1-\tau_h)}{\lambda}\right]^{\frac{1}{\gamma}} \cdot e_h^{\frac{1}{\gamma}\left[1+\frac{1}{\chi}\right]} z_h^{-\frac{1}{\gamma\chi}} \\ & \qquad\quad + \frac{1}{\gamma\chi} \cdot \frac{dz_h}{z_h} \cdot \left[\frac{(1-\tau_h)}{\lambda}\right]^{\frac{1}{\gamma}} \cdot e_h^{\frac{1}{\gamma}\left[1+\frac{1}{\chi}\right]} z_h^{-\frac{1}{\gamma\chi}} = 0. \end{aligned}$$

Some rearranging gives us

$$\begin{aligned} & \left\{\gamma(1-\tau_h)z_h + \frac{1}{\chi}\left[(1-\tau_h)z_h + (\tau_h - \tau_a(\bar{z}))\bar{z}\right]\right\} \frac{dz_h}{z_h} \\ & \quad + \left\{\gamma(1-\tau_h)(z_h - \bar{z}) - \left[(1-\tau_h)z_h + (\tau_h - \tau_a(\bar{z}))\bar{z}\right]\right\} \cdot \frac{d(1-\tau_h)}{1-\tau_h} \\ & \qquad - \gamma \frac{d\tau_a(\bar{z})}{d(1-\tau_h)} \cdot \frac{1-\tau_h}{\tau_a(\bar{z})} \cdot \tau_a(\bar{z}) \cdot \bar{z} \cdot \frac{d(1-\tau_h)}{1-\tau_h} = 0. \end{aligned}$$

Hence, we obtain with $-\tau_a(\bar{z})\bar{z} = R$

$$\begin{aligned} \epsilon(z_h) &= \frac{dz_h}{d(1-\tau_h)} \cdot \frac{1-\tau_h}{z_h} \\ &= \frac{(1-\gamma)(1-\tau_h)z_h + (\tau_h - \tau_a(\bar{z}))\bar{z}}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{(\tau_h - \tau_a(\bar{z}))\bar{z}}{\chi}} + \frac{\gamma(1-\tau_h)\bar{z}_h + \gamma\tau_a(\bar{z})\bar{z}\epsilon(\tau_a(\bar{z}))}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{(\tau_h - \tau_a(\bar{z}))\bar{z}}{\chi}} \\ &= \frac{(1-\gamma)(1-\tau_h)z_h + \tau_h\bar{z} + R}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{\tau_h\bar{z}+R}{\chi}} + \frac{\gamma(1-\tau_h)\bar{z}_h + \gamma\tau_a(\bar{z})\bar{z}\epsilon(\tau_a(\bar{z}))}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{\tau_h\bar{z}+R}{\chi}} \\ &= \epsilon_h^u - \frac{-\gamma(1-\tau_h)z_h}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h)z_h + \frac{\tau_h\bar{z}+R}{\chi}} \cdot \frac{\bar{z}}{z_h} \cdot \left[1 + \frac{\tau_a(\bar{z})}{1-\tau_h} \cdot \epsilon(\tau_a(\bar{z}))\right] \\ &= \epsilon_h^u - \eta_h \cdot \frac{\bar{z}}{z} \cdot \left[1 + \frac{\tau_a(\bar{z})}{1-\tau_h} \cdot \epsilon(\tau_a(\bar{z}))\right]. \end{aligned}$$

□

Comparison with the Saez (2001) result The formula for the Laffer tax rate can hence be written as

$$\tau_{\text{Laffer}} = \frac{1 - (a-1) \cdot \tau_a(\bar{z}) \cdot \epsilon(\tau_a(\bar{z}))}{1 + a \cdot \epsilon_h^u - \eta_h \cdot (a-1) \cdot \left[1 + \frac{\tau_a(\bar{z})}{1-\tau_h} \cdot \epsilon(\tau_a(\bar{z}))\right]}.$$

With $\epsilon(\tau_a(\bar{z})) = 0$ as in Saez (2001), the formula reduces to

$$\tau_{\text{Laffer}} = \frac{1}{1 + a\epsilon_h^u - \eta_h \cdot (a-1)} = \frac{1}{1 + \epsilon_h^u + \left[\epsilon_h^u - \eta_h\right] \cdot (a-1)} = \frac{1}{1 + \epsilon_h^u + \epsilon_h^c \cdot (a-1)},$$

with $\epsilon_h^u = \epsilon_h^c + \eta$ and ϵ_h^c being the compensated labor supply elasticity. This is the same as in equation (9) in Saez (2001, p. 212) with $\bar{g} = 0$.

A.3 Proof of Proposition 5

A.3.1 Revenue Maximization: The Laffer Curve Revisited

The revenue maximization problem, given the optimal labor supply choice of top income earners, can be stated as

$$\max_{\tau_h} \tau_h \left[\left[(1-\tau_h)\right]^{\chi} \left[e_h\right]^{1+\chi} - \bar{z} \right]$$

with first order condition

$$\left[(1-\tau_h)\right]^{\chi} \left[e_h\right]^{1+\chi} - \bar{z} = \chi \tau_h \left[1-\tau_h\right]^{\chi-1} \left[e_h\right]^{1+\chi}.$$

Thus[45]

$$\frac{1-\tau_h}{\tau_h} = \chi + \frac{\bar{z}}{\tau_h \left[1-\tau_h\right]^{\chi-1} \left[e_h\right]^{1+\chi}} > \chi$$

and thus the revenue maximizing tax rate satisfies

$$\tau_h^{\text{Laffer}} < \frac{1}{1+\chi} = \tau_h^{\text{Laffer}}(\bar{z}. = 0)$$

To state the revenue-maximizing rate more concisely, recall that the Pareto coefficient is defined as

$$\frac{a}{a-1} = \frac{e_h n_h}{\bar{z}} = \frac{\left[1-\tau_h\right]^{\chi} \left[e_h\right]^{1+\chi}}{\bar{z}} \quad \text{or} \quad \frac{a-1}{a} = \frac{\bar{z}}{\left[1-\tau_h\right]^{\chi} \left[e_h\right]^{1+\chi}}.$$

Note that if $\bar{z} = 0$, then $a = 1$ and as $\bar{z} \to z_h$ then $a \to \infty$. Then the revenue-maximizing tax rate satisfies

$$\frac{1-\tau_h}{\tau_h} = \chi + \frac{a-1}{a}\frac{1-\tau_h}{\tau_h} \quad \text{or} \quad \tau_h = \frac{1}{1 + a(\tau_h)\chi}. \tag{15}$$

precisely as predicted by the Saez formula. But it is important to note that

$$\frac{a-1}{a} = \frac{\bar{z}}{\left[1-\tau_h\right]^{\chi} \left[e_h\right]^{1+\chi}} = \frac{a-1}{a}(\tau_h) \quad \text{or}$$

[45] The previous equation also insures that at the revenue-maximizing tax rate (and thus at any tax rate lower than that) labor income of the top income earners $n_h e_h$ is strictly higher than the threshold $\bar{z}$.

$$a(\tau_h;\bar{z}) = \frac{[1-\tau_h]^{\chi}\,[e_h]^{1+\chi}}{[1-\tau_h]^{\chi}\,[e_h]^{1+\chi} - \bar{z}} = \frac{1}{1 - \frac{\bar{z}}{[1-\tau_h]^{\chi}[e_h]^{1+\chi}}}.$$

We observe that $a(\tau_h;\bar{z})$ is a strictly increasing function of the tax rate τ_h and a strictly increasing function of the threshold $\bar{z}$. Thus the right hand side of (15) is continuous and strictly decreasing in τ_h, strictly positive at $\tau_h = 0$ and tends to 0 as τ_h tends to 1. Thus there is a unique positive revenue-maximizing tax rate τ_h^{Laffer} characterized by (15), and since the right hand side is strictly decreasing in $\bar{z}$, so is τ_h^{Laffer}. At $\bar{z} = 0$, we find $\tau_h^{\text{Laffer}}(\bar{z} = 0) = \frac{1}{1+\chi}$.

A.3.2 Insuring that High Productivity Households Are Better Off

Proving our welfare results requires that high-productivity individuals are (weakly) better off than low-productivity individuals for all relevant tax rates. For a given tax rate τ_h this requires that

$$\frac{[e_l]^{1+\chi}}{1+\chi} + R \leq \frac{[(1-\tau_h)e_h]^{1+\chi}}{1+\chi} + \tau_h\bar{z} + R$$

and thus

$$\frac{[e_l]^{1+\chi}}{1+\chi} \leq \frac{[(1-\tau_h)e_h]^{1+\chi}}{1+\chi} + \tau_h\bar{z}.$$

Since the welfare-maximizing tax rate cannot exceed the revenue- and thus transfer-maximizing tax rate, a sufficient (but by no means necessary) condition for the welfare analysis that this condition is satisfied at the peak of the Laffer curve rate $\tau_h^{\text{Laffer}}(\bar{z} = 0) = \frac{1}{1+\chi}$ since for all $\bar{z} \geq 0$ and all $\tau_h \leq \tau_h^{\text{Laffer}}(\bar{z} = 0)$

$$\frac{[(1-\tau_h)e_h]^{1+\chi}}{1+\chi} + \tau_h\bar{z} \geq \frac{\left[\frac{\chi}{1+\chi}e_h\right]^{1+\chi}}{1+\chi}.$$

Thus a sufficient condition such that (for all $\bar{z}$) high-income individuals have higher welfare from post-tax consumption and labor than low-income individuals is

$$\frac{[e_l]^{1+\chi}}{1+\chi} \leq \frac{\left[\frac{\chi}{1+\chi}e_h\right]^{1+\chi}}{1+\chi}$$

or

$$e_l \leq \frac{\chi}{1+\chi}e_h$$
$$\frac{e_h}{e_l} \geq 1 + \frac{1}{\chi}.$$

This is the condition stated in the proposition. Note again that this is a (potentially very loose) sufficient condition, and a much tighter $\bar{z}$ specific condition could be obtained.

A.3.3 Welfare Maximization

Now turn to welfare, defined in the main text as

$$\mathcal{W}(\tau_h) = (1-\Phi_h)\frac{\left(\frac{[e_l]^{1+\chi}}{1+\chi}+R\right)^{1-\nu}}{1-\nu} + \Phi_h\frac{\left(\frac{[(1-\tau_h)e_h]^{1+\chi}}{1+\chi}+\tau_h\bar{z}+R\right)^{1-\nu}}{1-\nu}$$
$$R(\tau_h) = \Phi_h\tau_h\left[[1-\tau_h]^{\chi}[e_h]^{1+\chi}-\bar{z}\right].$$

Taking first order conditions with respect to the tax rate τ_h and rearranging yields

$$\Psi(\tau_h) := \left(\frac{[(1-\tau_h)e_h]^{1+\chi}+\tau_h\bar{z}+(1+\chi)R(\tau_h)}{[e_l]^{1+\chi}+(1+\chi)R(\tau_h)}\right)^{\nu}$$
$$= \frac{\Phi_h}{1-\Phi_h}\left(\frac{[e_h]^{1+\chi}[1-\tau_h]^{\chi}-\bar{z}}{\frac{dR(\tau_h)}{d\tau_h}}-1\right) := \Gamma(\tau_h)$$

with

$$\frac{dR(\tau_h)}{d\tau_h} = (1-\Phi_l)\left([1-\tau_h]^{\chi}[e_h]^{1+\chi}-\bar{z}-\tau_h\chi[1-\tau_h]^{\chi-1}[e_h]^{1+\chi}\right)$$
$$= (1-\Phi_l)[e_h]^{1+\chi}[1-\tau_h]^{\chi}\left(\left[1-\frac{\bar{z}}{[1-\tau_h]^{\chi}[e_h]^{1+\chi}}\right]-\frac{\chi\tau_h}{1-\tau_h}\right)$$

and thus

$$\Psi(\tau_h) := \left(\frac{[(1-\tau_h)e_h]^{1+\chi}+\tau_h\bar{z}+(1+\chi)R(\tau_h)}{[e_l]^{1+\chi}+(1+\chi)R(\tau_h)}\right)^{\nu}$$
$$= \frac{\Phi_h}{1-\Phi_h}\left(\frac{1}{\Phi_h\left(1-\frac{a(\tau_h)\chi\tau_h}{1-\tau_h}\right)}-1\right) := \Gamma(\tau_h).$$

The existence, uniqueness and comparative statics then depend on the properties of the functions $(\Gamma(\tau_h),\Psi(\tau_h))$. Since $a(\tau_h)$ is strictly increasing in τ_h, the function $\Gamma(\tau_h)$ is continuous, strictly increasing on $[0,\tau_h^{\text{Laffer}})$ and with

$$\Gamma(\tau_h=0) = \frac{\Phi_h}{1-\Phi_h}\left(\frac{1}{\Phi_h}-1\right) = 1$$
$$\lim_{\tau_h\to\tau_h^{\text{Laffer}}}\Gamma(\tau_h) = \frac{\Phi_h}{1-\Phi_h}\left(\frac{1}{\Phi_h\left(1-\frac{\frac{a\chi}{1+a\chi}}{\frac{a\chi}{1+a\chi}}\right)}-1\right) = \infty.$$

Finally, $\Gamma(\tau_h)$ is independent of ν and e_h/e_l, but depends on e_h through $a(\tau_h)$. Therefore, in the comparative statics results with respect to inequality e_h/e_l we need to state that when changing e_h the threshold $\bar{z}$ is also changed such that top income relative to threshold income $n_h e_h/\bar{z} = z_h/\bar{z}$ and thus a remains unchanged.

Turning to the function $\Psi(\tau_h)$ we first note that it is continuous and strictly decreasing on $[0, \tau_h^{\text{Laffer}}]$, with

$$\Psi(\tau_h = \tau_h^{\text{Laffer}}) < \infty$$
$$\Psi(\tau_h = 0) = \left(\frac{[e_h]^{1+\chi}}{[e_l]^{1+\chi}}\right)^{\nu} \geq 1,$$

with equality only if $\nu = 0$. $\Psi(\tau_h)$ is strictly decreasing in τ_h since, taking the derivatives of the numerator and the denominator, we obtain

$$\begin{aligned}
&\frac{d\left[(1-\tau_h)e_h\right]^{1+\chi} + \tau_h \bar{z} + (1+\chi)R(\tau_h)}{d\tau_h} \\
&\qquad = -(1+\chi)(1-\tau_h)^{\chi}\left[e_h\right]^{1+\chi} + \bar{z} + (1+\chi)\frac{dR(\tau_h)}{d\tau_h} \\
&\qquad\qquad \leq (1+\chi)\frac{dR(\tau_h)}{d\tau_h} = \frac{d\left[e_l\right]^{1+\chi} + (1+\chi)R(\tau_h)}{d\tau_h}
\end{aligned}$$

since for all $\tau_h \leq \tau_h^{\text{Laffer}}$ we have $(1-\tau_h)^{\chi}\left[e_h\right]^{1+\chi} = z_h \geq \bar{z}$.

Finally, since the ratio defining $\Psi(\tau_h)$ is strictly larger than 1, an increase in ν shifts $\Psi(\tau_h)$ up without changing $\Gamma(\tau_h)$, and thus increases the optimal tax rate τ_h^*. Finally, assume that e_l/e_h increases, reducing inequality, and further assume that $a = n_h e_h/\bar{z} = z_h/\bar{z}$ remains unchanged. Then $\Gamma(\tau_h)$ remains unchanged and

$$\Psi(\tau_h) = \left(\frac{\frac{[1-\tau_h)e_h]^{1+\chi} + \tau_h \bar{z}}{[1-\tau_h]^{\chi}[e_h]^{1+\chi}} + (1+\chi)\Phi_h \tau_h \left[1 - \frac{a-1}{a}\right]}{\frac{[e_l]^{1+\chi}}{[1-\tau_h]^{\chi}[e_h]^{1+\chi}} + (1+\chi)\Phi_h \tau_h \left[1 - \frac{a-1}{a}\right]}\right)^{\nu} = \left(\frac{1 - \frac{\tau_h}{a} + (1+\chi)\Phi_h \frac{\tau_h}{a}}{\frac{[e_l/e_h]^{1+\chi}}{[1-\tau_h]^{\chi}} + (1+\chi)\Phi_h \frac{\tau_h}{a}}\right)^{\nu}$$

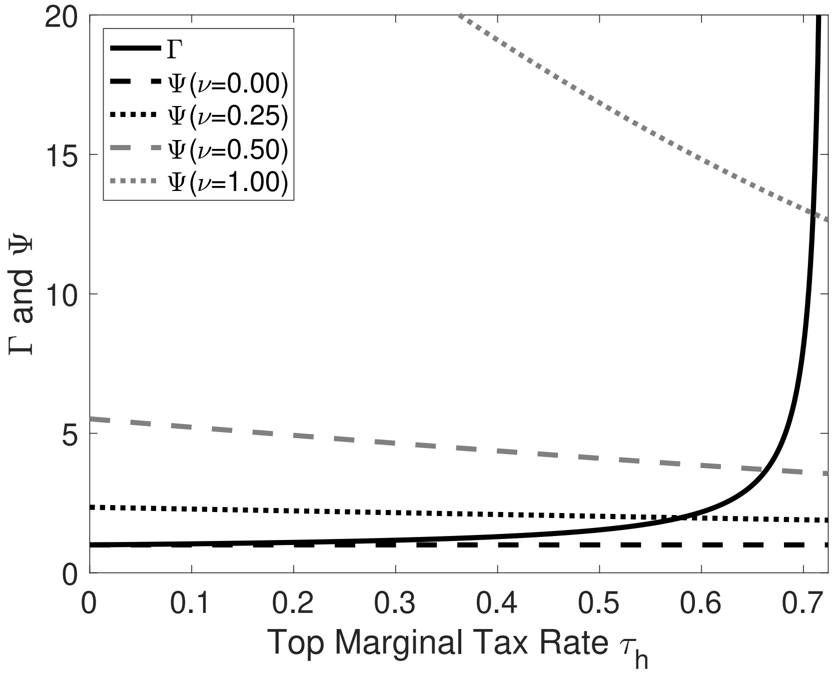
and thus the $\Psi(\tau_h)$ curve shifts down, reducing the optimal tax rate τ_h^*.

Figure 20 shows the determination of the optimal top tax rate for different values of social insurance ν. For each ν there is a unique intersection between the $\Gamma(\tau_h)$ curve and the $\Psi(\tau_h)$ curve. As ν increases, the $\Psi(\tau_h)$ curve shifts up, and thus the welfare maximizing top marginal tax rate rises. For $\nu = 0$ the optimal tax rate is zero, and for ν approaching ∞ it approaches the revenue maximizing peak of the Laffer curve rate.

B Appendix: Details of the Computational Approach

In order to solve the model outlined in this paper, we need three distinct algorithms: one that determines policy and value functions, one that solves for equilibrium quantities and prices, and one that delivers compensation payments.

Figure 20: Determination of Optimal Tax Rate for Different Degrees of Social Risk Aversion ν



B.1 Computation of Policy and Value Functions

We solve for policy and value functions using the method of endogenous grid points Formally, these functions exist on the state space

$$(j, s, \alpha, \eta, a) \in \{1, \ldots, J\} \times \{n, c\} \times \{-\sigma_\alpha, +\sigma_\alpha\} \times \{\eta_{s,1}, \ldots, \eta_{s,7}\} \times [0, \infty].$$

In order to be able to represent them on a computer, we however have to discretize the continuous elements of the state space, namely the asset dimension. For this purpose we chose a set of discrete points $\{\hat{a}^1 \ldots, \hat{a}^{100}\}$ such that the state space above can be approximated by

$$(j, s, \alpha, \eta, a) \in \{1, \ldots, J\} \times \{n, c\} \times \{-\sigma_\alpha, +\sigma_\alpha\} \times \{\eta_{s,1}, \ldots, \eta_{s,7}\} \times \{\hat{a}^1 \ldots, \hat{a}^{100}\}.$$

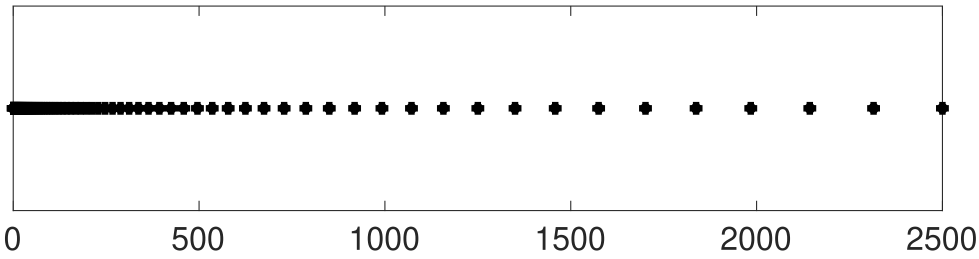
Note that the choice of $\hat{a}^i$ is not straightforward. Specifically we let

$$\hat{a}^i = \bar{a} \cdot \frac{(1+g_a)^{i-1} - 1}{(1+g_a)^{99} - 1},$$

which leaves us with two parameters that define our discrete grid space. $\bar{a}$ is the upper limit of the asset grid which we chose such that no individual in our simulated model would like to save more than this amount.[46] A g_a of 0 would result in equidistantly spaced grid points Setting $g_a > 0$ the distance between two successive grid points $\hat{a}^i$ and $\hat{a}^{i+1}$ grows at the rate g_a in i. In our preferred parameterization we let $g_a = 0.08$. We consequently located many grid points at the lower end of the grid space where

[46] In our model this leads to $\bar{a} = 1800$.

Figure 21: Discretized asset state space



borrowing constraints may occur and therefore policy functions may have kinks or be sharply curved. At the upper end of the grid space where policy and value functions are almost linear, we consequently use a much smaller amount of points. Figure 21 visualizes our discrete asset grid.

The discretization of the asset state space makes the solution for policy and value functions feasible via backward induction. We start out by solving the optimization problem at the last possible age an individual may have J. Since the agent is retired and dies with certainty, she will consume all her remaining resources and work zero hours,

$$c(J,s,\alpha,\eta,\hat{a}^i) = \frac{(1+r_n)\hat{a}^i + p(s,\alpha,\eta)}{1+\tau_c} \quad , \quad n(J,s,\alpha,\eta,\hat{a}^i) = 0 \quad ,$$
$$a'(J,s,\alpha,\eta,\hat{a}^i) = 0 \quad \text{for all } i = 1,\ldots,100.$$

In order to simplify the computation of the value function we will actually keep track of two different value functions, the one for consumption and the one for labor. This is possible due to the additive separability assumption we made. Consequently we have

$$v_c(J,s,\alpha,\eta,\hat{a}^i) = \frac{[c(J,s,\alpha,\eta,\hat{a}^i)]^{1-\gamma}}{1-\gamma} \quad \text{and} \quad v_n(J,s,\alpha,\eta,\hat{a}^i) = 0.$$

Knowing the policy and value function in the last period of life, we can now iterate backward over ages to determine the remaining household decisions. Since the algorithm is very similar for retired and working individuals, we will restrict ourselves to the case of workers. Assume that we had already calculated policy and value functions at age $j+1$. The problem we need to solve for an individual at state (j,s,α,η,a) then reads

$$\max_{c,n,a'} \quad \frac{c^{1-\gamma}}{1-\gamma} - \alpha\frac{n^{1+\chi}}{1+\chi} + \beta\psi_{j+1}\sum_{\eta'}\pi_s(\eta'|\eta)\left[v_c(j+1,s,\alpha,\eta',a') - v_l(j+1,s,\alpha,\eta',a')\right]$$

subject to the constraints

$$(1+\tau_c)c + a' + T(we(j,s,\alpha,\eta)n) + T_{ss}(we(j,s,\alpha,\eta)n)$$
$$= (1+r_n)a + b_j(s,\eta) + we(j,s,\alpha,\eta)n$$

as well as $0 \leq n \leq 1$ and $a' \geq 0$. The first order conditions (ignoring the constraint on n and the borrowing constraint) then are

$$c = [\lambda(1+\tau_c)]^{-1/\gamma}$$

$$\alpha n^{\chi} = \lambda we(j,s,\alpha,\eta)\left[1 - T'(we(j,s,\alpha,\eta)n) - T'_{ss}(we(j,s,\alpha,\eta)n)\right]$$
$$\lambda = \beta\psi_{j+1}(1+r'_n)(1+\tau'_c)\sum_{\eta'} c(j+1,s,\alpha,\eta',a'^{-\gamma}.$$

We now apply the method of endogenous grid points as follows: We assume that savings for tomorrow would amount to $a' = \hat{a}^i$ for all $i = 1,\dots,100$. Under this assumption, we can compute for each combination of (s,α,η) the respective λ from the last first-order condition. λ then defines a certain level of consumption $c^e(j,s,\alpha,\eta,\hat{a}^i)$ and labor supply $n^e(j,s,\alpha,\eta,\hat{a}^i)$.[47] Plugging these into the budget constraint, we can determine the endogenous grid point as

$$a^e(j,s,\alpha,\eta,\hat{a}^i) = \frac{1}{1+r_n}\Big[(1+\tau_c)c^e(j,s,\alpha,\eta,\hat{a}^i) + a' + T(we(j,s,\alpha,\eta)n) + T_{ss}(we(j,s,\alpha,\eta)n) - b_j(s,\eta) - we(j,s,\alpha,\eta)n^e(j,s,\alpha,\eta,\hat{a}^i)\Big].$$

Finally, we can compute the value functions as

$$v^e_c(j,s,\alpha,\eta,\hat{a}^i) = \frac{[c^e(j,s,\alpha,\eta,\hat{a}^i)]^{1-\gamma}}{1-\gamma} + \beta\psi_{j+1}\sum_{\eta'}\pi_s(\eta'|\eta)v_c(j+1,s,\alpha,\eta',\hat{a}_i)$$
$$v^e_n(j,s,\alpha,\eta,\hat{a}^i) = \frac{\alpha[n^e(j,s,\alpha,\eta,\hat{a}^i)]^{1+\chi}}{1+\chi} + \beta\psi_{j+1}\sum_{\eta'}\pi_s(\eta'|\eta)v_n(j+1,s,\alpha,\eta',\hat{a}_i).$$

Using the interpolation data

$$\left\{a^e(j,s,\alpha,\eta,\hat{a}^i), c^e(j,s,\alpha,\eta,\hat{a}^i)\right\}_{i=1}^{100} \quad , \quad \left\{a^e(j,s,\alpha,\eta,\hat{a}^i), n^e(j,s,\alpha,\eta,\hat{a}^i)\right\}_{i=1}^{100} \quad ,$$
$$\left\{a^e(j,s,\alpha,\eta,\hat{a}^i), v^e_c(j,s,\alpha,\eta,\hat{a}^i)\right\}_{i=1}^{100} \quad , \quad \left\{a^e(j,s,\alpha,\eta,\hat{a}^i), v^e_n(j,s,\alpha,\eta,\hat{a}^i)\right\}_{i=1}^{100} \quad ,$$

we can finally determine the (discrete) policy and value functions

$$c(j,s,\alpha,\eta,\hat{a}^i) \quad , \quad n(j,s,\alpha,\eta,\hat{a}^i) \quad , \quad v_c(j,s,\alpha,\eta,\hat{a}^i) \quad \text{and} \quad v_n(j,s,\alpha,\eta,\hat{a}^i)$$

for each today's asset value $\hat{a}^i, i = 1,\dots,100$ by piece-wise linear interpolation.[48]

Before applying this interpolation scheme, we however check for the occurrence of liquidity constraints. Liquidity constraints occur if $a^e(j,s,\alpha,\eta,0) > 0$. In this case, we

[47] Note that we can not solve for labor supply analytically due to the non linearity of the labor earnings tax schedule. Instead we use a quasi-Newton root finding routine to determine the solution to the respective first order condition. We thereby have to respect the constraint $0 \leq n \leq 1$ as well as the fact that there is a cap on contributions to the social security system. However, due to the additive separability of the utility function in consumption and labor supply, the constraints on n will not affect the individual's choice of consumption c.

[48] We do not interpolate v^e_c and v^e_n directly, but rather $[(1-\gamma)v^e_c]^{1/(1-\gamma)}$ and $[(1+\chi)v^e_n]^{1/(1+\chi)}$ and then transform them back to their original shape. This leads to much more accurate results in the high curvature region of the asset grid.

extend the above interpolation data by another point of value 0 on the left. The policy and value functions at this point are determined under the assumption that $a = a' = 0$, i.e. the policy function values solve the equation system

$$\begin{aligned}
\frac{c^{-\gamma}}{1+\tau_c} &= \lambda \\
\alpha n^{\chi} &= \lambda we(j,s,\alpha,\eta)\left[1 - T'(we(j,s,\alpha,\eta)n) - T'_{ss}(we(j,s,\alpha,\eta)n)\right] \\
(1+\tau_c)c &= b_j(s,\eta) + we(j,s,\alpha,\eta)n - T(we(j,s,\alpha,\eta)n) - T_{ss}(we(j,s,\alpha,\eta)n).
\end{aligned}$$

B.2 Determining Aggregate Quantities and Prices

Our algorithm to determine aggregate quantities and prices follows closely the Gauss-Seidel method already proposed in Auerbach and Kotlikoff (1987). Specifically, in order to determine an equilibrium path of the economy, we start with an initial guess of quantities $\{K_t, L_t\}_{t\geq 0}$ as well as tax rates $\{\tau_l, \tau_{ss,t}\}_{t\geq 0}$ and transfers $\{Tr_t\}_{t\geq 0}$. Our algorithm then iterates over the following steps:

1. Determine factor prices $\{r_t, w_t\}_{t\geq 0}$ that correspond to the quantities $\{K_t, L_t\}_{t\geq 0}$.
2. Solve the household optimization problem using these factor prices and the guesses for tax rates. Determine the measure of households.
3. Solve for the tax rate τ_l that balances the intertemporal budget constraint of the government by means of a quasi-Newton root finding method. Then calculate the path of government debt $\{B_t\}_{t\geq 0}$.
4. Determine the budget balancing payroll tax rates $\tau_{ss,t}$ using the social security system's sequential budget constraints.
5. Calculate lump-sum transfers Tr such that the sum of transfers equals the sum of bequests left by the non-surviving households.
6. Determine the new quantities $\{K_t^{new}, L_t^{new}\}_{t\geq 0}$ by aggregating individual decisions. Calculate updated quantities through

   $$K_t = (1-\omega)K_t + \omega K_t^{new} \quad \text{and} \quad L_t = (1-\omega)L_t + \omega L_t^{new}.$$

   ω thereby serves as a damping factor. Our preferred value for ω is 0.3.
7. Check whether the economy is in equilibrium, i.e.

   $$\max_{t\geq 0}\left|\frac{Y_t - C_t - I_t - G_t}{Y_t}\right| < \varepsilon.$$

   This means that the relative difference between aggregate demand and supply of goods should be smaller than a given tolerance level. If this is not the case, start with the updated guesses of quantities, tax rates and transfers at step 1. If this is the case, we have found an equilibrium path of the economy. To determine the initial equilibrium we use a tolerance level of $\varepsilon = 10^{-9}$ while for the transition path we set $\varepsilon = 10^{-6}$.

B.3 Calculation of Compensating Transfers

The calculation of compensating transfers is straightforward. In order to do so, we use a quasi-Newton root finding method that numerically determines the solutions to the equations

$$v_1(j,s,\alpha,\eta,a+\Psi_0(j,s,\alpha,\eta,a)) = v_0(j,s,\alpha,\eta,a)$$

and

$$Ev_t(j=1,s,\alpha,\bar{\eta},\Psi_t) = Ev_0(j=1,s,\alpha,\bar{\eta},0),$$

respectively. Note that in each iteration of the root finding method, we have to solve for the optimal household decisions.

C Appendix: Definition of Invariant Probability Measure

First we construct the share of the population in each age group. Let $\tilde{\mu}_1 = 1$, and for each $j \in \{2,\ldots,J\}$ define recursively

$$\tilde{\mu}_j = \frac{\psi_j \tilde{\mu}_{j-1}}{1+g_n}.$$

Then the share of the population in each age group is given by

$$\mu_j = \frac{\tilde{\mu}_j}{\sum_\iota \tilde{\mu}_\iota}.$$

Next, we construct the measure of households of age 1 across characteristics (s,α,η,a). By assumption (see the calibration section, Section 4 of the paper) newborn households enter the economy with zero assets, $a=0$ and at the mean idiosyncratic productivity shock $\bar{\eta}$. The share of college-educated households is exogenously given by ϕ_c and $\phi_n = 1-\phi_c$, and the fixed effect is drawn from a discrete pdf $\phi_s(\alpha)$. Thus

$$\Phi(\{j=1\},\{\alpha\},\{s\},\{\bar{\eta}\},\{0\}) = \mu_1 \phi_s \phi_s(\alpha)$$

for $s=\{n,c\}$ and zero else.

Finally we construct the probability measure for all ages $j>1$. For all Borel sets of assets $\mathcal{A}$ we have

$$\Phi(\{j+1\},\{\alpha\},\{s\},\{\eta'\},\mathcal{A}) = \frac{\psi_{j+1}\pi_s(\eta'|\eta)}{1+g_n}\int \mathbf{1}_{\{a'(j,s,\alpha,\eta,a)\in\mathcal{A}\}}\Phi(\{j\},\{\alpha\},\{s\},\{\eta\},da)$$

where

$$\int \mathbf{1}_{\{a'(j,s,\alpha,\eta,a)\in\mathcal{A}\}}\Phi(\{j\},\{\alpha\},\{s\},\{\eta\},da)$$

is the measure of assets a today such that, for fixed (j,s,α,η), the optimal choice today of assets for tomorrow, $a'(j,s,\alpha,\eta,a)$ lies in $\mathcal{A}$.

D Appendix: Details of the Calibration

D.1 Markov Chain for Labor Productivity

The Markov chain governing idiosyncratic labor productivity for both education groups is given by

$s = n$

i, j	1	2	3	4	5	6	7
1	0.969909	0.029317	0.000332	0.000002	0.000000	0.000440	0.000000
2	0.007329	0.970075	0.021989	0.000166	0.000000	0.000440	0.000000
3	0.000055	0.014659	0.970130	0.014659	0.000055	0.000440	0.000000
4	0.000000	0.000166	0.021989	0.970075	0.007329	0.000440	0.000000
5	0.000000	0.000002	0.000332	0.029317	0.969909	0.000440	0.000000
6	0.000000	0.000000	0.002266	0.000000	0.000000	0.970000	0.027734
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.288746	0.711254
$\exp(\eta_{n,i})$	0.1354	0.3680	1.0000	2.7176	7.3853	19.7204	654.0124

and

$s = c$

i, j	1	2	3	4	5	6	7
1	0.960937	0.029046	0.000329	0.000002	0.000000	0.009686	0.000000
2	0.007261	0.961102	0.021786	0.000165	0.000000	0.009686	0.000000
3	0.000055	0.014524	0.961157	0.014524	0.000055	0.009686	0.000000
4	0.000000	0.000165	0.021786	0.961102	0.007261	0.009686	0.000000
5	0.000000	0.000002	0.000329	0.029046	0.960937	0.009686	0.000000
6	0.000000	0.000000	0.047247	0.000000	0.000000	0.949922	0.002831
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.288746	0.711254
$\exp(\eta_{c,i})$	0.2362	0.4860	1.0000	2.0575	4.2334	8.3134	654.0124

D.2 Numerical Computation of Policy Elasticities

In order to be able to apply the formula for the Laffer tax rate proposed in Proposition 1 in our full quantitative simulation model, we have to calculate the policy elasticities $\epsilon(z_h)$ and $\epsilon(\tau_a(\bar{z}))$. To this end, we proceed in the following steps:

1. We start from the initial equilibrium described in Section 5 and compute a transition path that results from keeping the top tax rate at its initial equilibrium level of $\tau_h = 0.396$, but setting the top tax threshold $\bar{z}_h$ such that exactly the top 1% earners are hit by the top rate. For each period $t \geq 1$ of the transition, we can then calculate average top 1% earnings as

$$z_{h,t}^0 = \int \mathbb{1}_{z_t(j,s,\alpha,\eta,a)\geq \bar{z}_t}\; z_t(j,s,\alpha,\eta,a)\; d\Phi_t$$
$$\text{with} \quad z_t(j,s,\alpha,\eta,a) = w_t e(j,s,\alpha,\eta) n_t(j,s,\alpha,\eta,a).$$

The average tax rate at the top earnings threshold then is

$$\tau_a^0(\bar{z}_t) = \frac{T_t(\bar{z}_t)}{\bar{z}_t}.$$

2. We now increase the top marginal net-of-tax rate by an amount δ and calculate a new equilibrium path. From this, we obtain a new value for average top 1% earnings $z_{h,t}^1$ and a new value for the average tax rate $\tau_a^1(\bar{z}_t)$. In our numerical calculations, we use $\delta = 0.01$.

3. The relevant elasticities for the Laffer tax rate formula then are

$$\epsilon_t(z_{h,t}) = \frac{z_{h,t}^1 - z_{h,t}^0}{\delta} \cdot \frac{0.604}{z_{h,t}^0} \quad \text{and} \quad \epsilon_t(\tau_a(\bar{z}_t)) = \frac{\tau_a^1(\bar{z}_t) - \tau_a^0(\bar{z}_t)}{\delta} \cdot \frac{0.604}{\tau_a^0(\bar{z}_t)}.$$

The elasticities we derive from this procedure are listed in the columns *initial* within Table 8. We then repeat the above exercise, but instead of starting from the initial equilibrium tax rate $\tau_h = 0.396$, we start from the actual Laffer tax rate. The resulting elasticities are then shown in the columns *final*.

D.3 The Social Security System

We use the US pension formula to calculate pension payments. Specifically, for a given average labor earnings $\tilde{z}$ we set

$$p(s,\alpha,\eta) = f(\tilde{z}) = \begin{cases} r_1\tilde{z} & \text{if } \tilde{z} < b_1 y^{med} \\ r_1 b_1 y^{med} + r_2(\tilde{z} - b_1 y^{med}) & \text{if } \tilde{z} < b_2 y^{med} \\ r_1 b_1 y^{med} + r_2(b_2 - b_1) y^{med} + r_3(\tilde{z} - b_2 y^{med}) & \text{otherwise} \end{cases}$$

Here r_1, r_2, r_3 are the respective replacement rates and b_1 and b_2 the bend points. We express these points in terms of median household *income* y^{med} which is the median of income from labor and assets (including bequests and pension payments). We use $y^{med} = 50,000$ as a reference value for this (see US Census Bureau for 2009). Consequently, the bend points are $b_1 = 0.184$ and $b_2 = 1.144$ and the respective replacement rates are $r_1 = 0.90$, $r_2 = 0.32$ and $r_3 = 0.15$. The maximum amount of pension benefit a household can receive is therefore 30,396, or 0.608 times the median income. All data is taken from the information site of the social security system for 2012. Finally, we calibrate the contribution cap of the pension system $\bar{z}_{ss}$ in order to obtain a contribution rate of 12.4 percent.

E Appendix: Additional Figures

Figure 22: Variance of Consumption and Hours over the Life Cycles, Entire Population

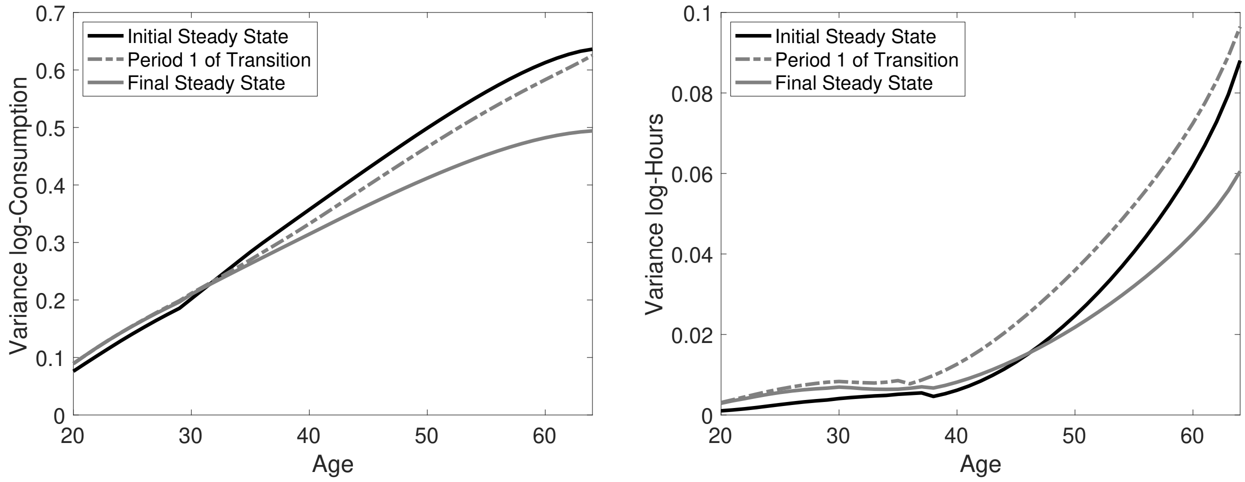
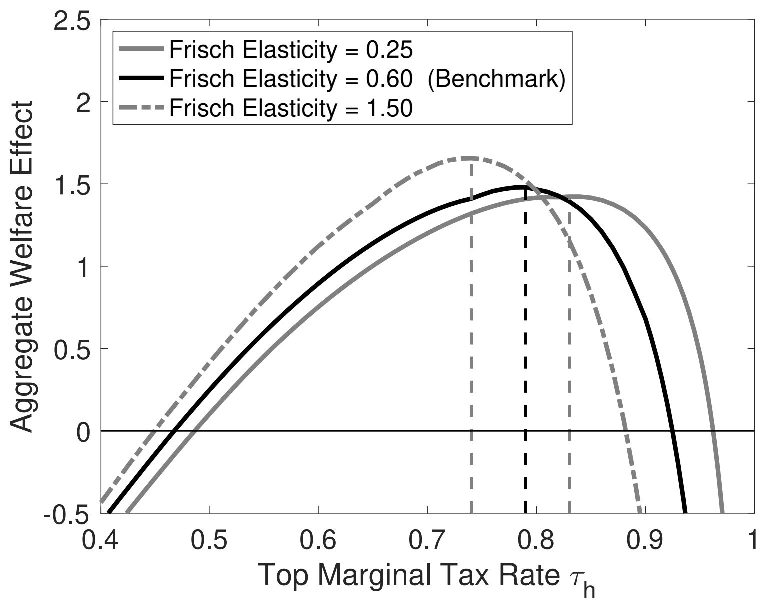


Figure 23: Aggregate Welfare as Function of τ_h, Different Frisch Elasticities



F Appendix: Sensitivity analysis

When doing sensitivity analysis, we have to partly recalibrate the model in order to make results comparable. For each different specification of the model we therefore recalibrate the technology level Ω such that the wage rate for effective labor is again equal to $w = 1$ as well as the depreciation rate δ_k such that the interest rate remains at 4%. The former ensures stability of our computational algorithm, the latter is necessary to guarantee equal weights of generations in the social welfare function. Finally we recalibrate the taste parameter for the disutility of labor λ so that average hours worked remain at 33% of the time endowment. We furthermore do some specific adjustments for different sensitivity scenarios which we outline in the following.

F.1 Size of the Income Effect

When we impose log preferences the relationship between hours worked and individual labor productivity changes dramatically. As a consequence we have to completely recalibrate the total income process. The following table shows which probabilities and productivity levels we have to choose in this case to obtain the same fit for the earnings and wealth distribution in our model:

$s = n$

1	0.969945	0.029318	0.000332	0.000002	0.000000	0.000403	0.000000
2	0.007330	0.970111	0.021990	0.000166	0.000000	0.000403	0.000000
3	0.000055	0.014660	0.970166	0.014660	0.000055	0.000403	0.000000
4	0.000000	0.000166	0.021990	0.970111	0.007330	0.000403	0.000000
5	0.000000	0.000002	0.000332	0.029318	0.969945	0.000403	0.000000
6	0.000000	0.000000	0.012043	0.000000	0.000000	0.969903	0.018054
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.269999	0.730001
$\exp(\eta_{n,i})$	0.1722	0.4149	1.0000	2.4101	5.8085	18.0227	374.1023

and

$s = c$

i, j	1	2	3	4	5	6	7
1	0.960202	0.029024	0.000329	0.000002	0.000000	0.010444	0.000000
2	0.007256	0.960366	0.021769	0.000164	0.000000	0.010444	0.000000
3	0.000055	0.014513	0.960421	0.014513	0.000055	0.010444	0.000000
4	0.000000	0.000164	0.021769	0.960366	0.007256	0.010444	0.000000
5	0.000000	0.000002	0.000329	0.029024	0.960202	0.010444	0.000000
6	0.000000	0.000000	0.068922	0.000000	0.000000	0.928130	0.002948
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.269999	0.730001
$\exp(\eta_{c,i})$	0.2809	0.5300	1.0000	1.8867	3.5597	6.3118	374.1023

F.2 Persistence of High Productivity States

To make the highest productivity state completely permanent we again have to adjust the transition probabilities in our model. This time we assume that only at age 30 there is a certain probability that individuals can climb up to the highest productivity region.

This probability is the same for each individual of an education level. In order to determine this probability we calculate the fraction of individuals in the highest productivity region between the ages 30 and j_r for each education level in the benchmark model. We then choose the probability to get a permanent very high income shock in the sensitivity model such that the fraction of households in the highest income region is exactly the same as in the benchmark model.