NBER WORKING PAPER SERIES

HIGH MARGINAL TAX RATES ON THE TOP 1%? LESSONS FROM A LIFE CYCLE MODEL WITH IDIOSYNCRATIC INCOME RISK

Fabian Kindermann
Dirk Krueger

Working Paper 20601
http://www.nber.org/papers/w20601

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2014

We thank seminar participants at USC, the Wharton Macro Lunch, the 2014 Macro Tax conference in Montreal, the 4th SEEK Conference in Mannheim, the 2014 SED in Toronto and the 2014 NBER Summer Institute, as well as Juan Carlos Conesa and William Peterman for many useful comments. Krueger thanks the National Science Foundation for support under grant SES 1123547. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk

Fabian Kindermann and Dirk Krueger

NBER Working Paper No. 20601
October 2014
JEL No. E62,H21,H24

ABSTRACT

In this paper we argue that very high marginal labor income tax rates are an effective tool for social insurance even when households have preferences with high labor supply elasticity, make dynamic savings decisions, and policies have general equilibrium effects. To make this point we construct a large scale Overlapping Generations Model with uninsurable labor productivity risk, show that it has a wealth distribution that matches the data well, and then use it to characterize fiscal policies that achieve a desired degree of redistribution in society. We find that marginal tax rates on the top 1% of the earnings distribution of close to 90% are optimal. We document that this result is robust to plausible variation in the labor supply elasticity and holds regardless of whether social welfare is measured at the steady state only or includes transitional generations.

Fabian Kindermann
University of Bonn
Institute for Macroeconomics and Econometrics
Adenauerallee 24-42
53113 Bonn, Germany
fabian.kindermann@uni-bonn.de

Dirk Krueger
Economics Department
University of Pennsylvania
160 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104
and NBER
dkrueger@econ.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 1970's and exceeded 20% in 2007. At the same time the highest marginal tax rate declined from levels consistently above 60% to below 40%. This triggered popular and academic calls to 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 Occupy Wall Street movement.

However, reducing inequality is not necessarily an objective in and of itself for a benevolent government. In this paper we ask what is the welfare-maximizing labor income tax rate on the top 1% earners, where welfare is measured as the weighted sum of expected lifetime utility of households currently alive and born in the future.[1] We are especially interested in the question whether high marginal tax rates of the size advocated in the literature cited above can be rationalized on these normative grounds. To answer this question we construct a quantitative overlapping generations economy with ex-ante skill and thus earnings potential heterogeneity, idiosyncratic wage risk and endogenous labor supply and savings choices. We follow Castaneda et al. (2003) and assure, via an appropriate calibration of the labor productivity process, 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. Therefore 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 quantitatively determine the answer to the question above. To do so we compute, within a restricted class of income tax functions which has as one of the policy choice variables the marginal tax rate applying to the top 1%, the optimal one-time tax reform, which in turn induces an economic transition from the current status quo[2] towards a new stationary equilibrium. We find that the optimal marginal tax rates on the top 1% of earners is indeed very high, in excess of 90%, and thus consistent with the empirically observed levels after World War II. Note that since we explicitly consider the transition periods in our policy analysis, our results capture both short- and long-run consequences of the policy reforms we consider. Interestingly, even when including welfare of current and future top 1% earners in the social welfare function, and even when restricting attention only to the long-run consequences of the policy reform (by adopting a steady state welfare measure) we find very high optimal marginal tax rates, in the order of about 90%.

---

[1] We alternatively include or exclude households in the top 1% in our measure of social welfare; as it turns out, the differences in results is quantitatively small.

[2] Which we take to be a stylized version of the current U.S. personal income tax code.

We then show that these results are primarily driven by the social insurance benefits that these high taxes imply. Concretely, in order 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 (think of attractive entrepreneurial, entertainment or professional sports opportunities). The labor supply of these households is not prohibitively strongly affected even by very high marginal tax rates even with a utility function with high Frisch labor supply elasticity, since it is these periods of high labor productivity that households earn the majority of their lifetime income from. 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.

1.1 Related Literature

The basic 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 75%. On the empirical side the literature that motivates our analysis in the first place 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. Important examples include Domeij and Heathcote (2004), Conesa and Krueger (2006), Conesa et al. (2009), Bakis et al. (2013) and Fehr and Kindermann (2014). A subset of this literature (see e.g. Guner, Lopez-Daneri and Ventura, 2014, or Holter, Krueger and Stepanchuk, 2014) 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 smaller than the revenue-maximizing rate (from the top 1%), it is quantitatively close.[3]

Especially relevant for our work is the complementary paper by Badel and Huggett

[3] 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) suggests 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 (1999), Vogelgesang (2000) and Jacobs and Schindler (2012). Also relevant for our study is the theoretical literature on optimal taxation over the life cycle, e.g. Erosa and Gervais (2002).

(2014). These authors study a dynamic economy with endogenous human capital accumulation to quantify the effects of high marginal income tax rates 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 revenue-maximizing tax rates are significantly smaller than the one we find (and not far away from the currently prevalent ones in the U.S.). Also complementary to our work is Brüggemann and Yoo (2014) who study 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, report substantial adverse aggregate and large positive distributional consequences, resulting in net welfare gains from the policy reform they study.

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 model, especially Castaneda et al. (2003), but also Quadrini (1997), Krusell and Smith (1998) as well as Cagetti and DeNardi (2006).

2 The Model

We study a standard large-scale overlapping generations model in the spirit of Auerbach and Kotlikoff (1987), but 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.

In order to highlight the key ingredients of the model in its most transparent way for a given government policy 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.

2.1 Technology

The single good in this economy is produced by a continuum of representative, competitive firms that hire capital and labor on competitive spot markets to operate the constant returns to scale technology

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

where $\Omega \geq 0$ parametrizes 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 production technologies the number of operative firms as well as 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 (1).

2.2 Preferences and Endowments

Households in this economy 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 $1+n$ as large as the previous cohort, so that 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

$$U(c_j, n_j),$$

and the time discount factor β and are expected utility maximizers (with respect to longevity risk and with respect to idiosyncratic wage risk described below).

Households 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 the household, with $s = c$ denoting some college education and $s = n$ representing (less than or equal) high school education. The fraction of college educated households is exogenously given by ϕ_s. In addition, prior to labor market entry households draw a fixed effect[4] 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 $\pi_s(\eta'|\eta)$.

[4] Both education and the fixed effect will shift life cycle wage profiles in a deterministic fashion in the model, 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 of the model.

Idiosyncratic wage risk (determined by the process for η) and mortality risk (parameterized by the survival probabilities ψ_j) cannot be explicitly insured as markets are incomplete 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 $we(j,s,\alpha,\eta)n$, capital income $r_n a$ and transfers $b_j(s,\alpha,\eta)$[5] on consumption $(1+\tau_c)c$, including consumption taxes, and on paying labor income taxes $T(we(j,s,\alpha,\eta)n)$ as well as payroll taxes $T_{ss}(we(j,s,\alpha,\eta)n)$. Implicit in these formulations is that the consumption and capital income tax is assumed to be linear, whereas the labor earnings tax is given by the potentially nonlinear (but continuously differentiable) 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

$$v(j,s,\alpha,\eta,a) = \max_{c,n,a'} U(c,n) + \beta\psi_{j+1}\sum_{\eta'}\pi_s(\eta'|\eta)v(j+1,s,\alpha,\eta',a') \tag{2}$$

subject to

$$\begin{aligned} & (1+\tau_c)c + a' + T(we(j,s,\alpha,\eta)n) + T_{ss}(we(j,s,\alpha,\eta)n) \\ = \; & (1+r_n)a + b_j(s,\eta) + we(j,s,\alpha,\eta)n \end{aligned} \tag{3}$$

and subject to a tight borrowing limit $\alpha' \geq 0$. The result of this dynamic programming problem is a value function v and policy functions c, n, a' as functions of the state (j,s,α,η,a) of a household.

2.3 Government Policy

The government uses tax revenues from labor earnings, capital income and consumption taxes to finance an exogenously given 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 (and medicare program). 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 n in this economy (G,B) should be interpreted as per capita variables since these are constant in a stationary recursive competitive equilibrium.

[5] Transfers include social security for those that are retired as well as accidental bequests for all working households.

Letting by Φ denote the cross-sectional distribution[6] of households (constant in a stationary equilibrium), the budget constraint of the government in a stationary recursive competitive equilibrium with population growth reads as

$$\begin{aligned} & r\tau_k \int a'(j,s,\alpha,\eta,a)d\Phi + \tau_c \int c(j,s,\alpha,\eta,a)d\Phi + \int T(we(j,s,\alpha,\eta)n(j,s,\alpha,\eta,a))d\Phi \\ = \; & G + (r-n)B \end{aligned} \tag{4}$$

In addition, the PAYGO social security system is characterized by a payroll tax rate τ_{ss}, an earnings threshold $\bar{y}_{ss}$ only 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[7] as well as education s and the fixed effect α (which in turn determine expected wages over the life cycle). Thus $(\tau_{ss}, \bar{y}_{ss})$ completely determine the payroll tax function T_{ss}. The specific form of the function $p(s,\alpha,\eta)$ is discussed in the calibration section.

The budget constraint of the social security system then reads as

$$\int p(s,\alpha,\eta) \cdot \mathbf{1}_{\{j>j_r\}} d\Phi = \tau_{ss} \int \min\{\bar{y}_{ss}, we(j,s,\alpha,\eta)n(j,s,\alpha,\eta,a))\} d\Phi. \tag{5}$$

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

$$Tr = \frac{\int (1+r^n)(1-\psi_{j+1})a'(j,s,\alpha,\eta,a)d\Phi}{\int \mathbf{1}_{\{j\leq j_r\}} d\Phi}. \tag{6}$$

so that transfers received by households are given as

$$b(j,s,\alpha,\eta) = \begin{cases} Tr & \text{if } j \leq j_r \\ p(s,\alpha,\eta) & \text{if } j > j_r \end{cases} \tag{7}$$

2.4 Recursive Competitive Equilibrium (RCE)

Definition 1 *Given government expenditures G, government debt B, a tax system characterized by (τ_c, τ_k, T) and a social security system characterized by $(\tau_{ss}, \bar{y}_{ss})$, 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*

---

[6] Formally, and given our notation, Φ is a measure and the total mass of households of age $j = 1$ is normalized to 1.

[7] 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 increase labor supply in their last working period to boost pension payments.

1. *[Household maximization]: Given prices* (r, w)*, transfers* b_j *given by* (7) *and government policies* $(\tau_c, \tau_k, T, \tau_{ss}, \bar{y}_{ss})$*, the value function* v *satisfies the Bellman equation* (2)*, and* (c, n, a') *are the associated policy functions.*

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

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

3. *[Government Budget Constraints]: Government policies satisfy the government budget constraints* (4) *and* (5)*.*

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*

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

   (c) *The goods market clears*

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

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 B.*

2.5 Transition Paths

Our thought experiments will involve unexpected changes in government tax policy that will 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 of 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

$$G + (1+r_t)B_t = (1+n)B_{t+1} + r_t\tau_k(K_t + B_t) + \tau_c \int c_t(j, s, \alpha, \eta, a) d\Phi_t$$

$$+ \int T_t(w_t e(j,s,\alpha,\eta) n_t(j,s,\alpha,\eta,a)) d\Phi_t$$

and

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

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 t. 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).

3 Mapping the Model into Data

Conceptually, we proceed in two steps when we 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. 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.

Most of the calibration is fairly standard for quantitative OLG models with idiosyncratic risk. However, given the purpose of the paper it is important that the model-generated cross-sectional earnings and wealth distribution is characterized by the same concentration as in the data, especially at the top of the earnings and wealth distribution. Broadly, 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. On the other hand, 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 that matches the data quite well.

3.1 Demographics

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

3.2 Technology

The production side of the model is characterized by the three parameters $(\Omega, \epsilon, \delta_k)$. We set the capital share in production 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 on capital δ_k is set such that the initial equilibrium interest rate in the economy is $r = 4\%$; this requires an annual depreciation rate of $\delta_k = 7.6\%$.

3.3 Endowments and Preferences

3.3.1 Labor Productivity

In every period a household is endowed with one unit of time which can be used for leisure and market work. One unit of work time yields a wage $we(j, s, \alpha, \eta)$, where $e(j, s, \alpha, \eta)$ is the idiosyncratic labor productivity (and thus the idiosyncratic component of the wage) of the household which depends on the age j education s and the fixed effect α of the household as well as its 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 observed at the top of the cross-sectional distribution, but not captured by any observations in the PSID. We then specify log-wages 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.[8]

Given these assumptions we need to specify the seven states of Markov chain $\{\eta_{s,1}, \ldots, \eta_{s,7}\}$ as well as the transition matrices π_s; in addition we need to determine the education-

[8] Conceptually, nothing prevents us to specify

$$e(j, s, \alpha, \eta) = \exp(\alpha + \varepsilon_{j,s} + \eta)$$

for $\eta = \eta_7$ but it turns out that our chosen specification provides a better fit to the earnings and wealth distributions.

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 direct 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 AR(1) processes as well as their persistences, $\{\sigma^2_{\alpha,s}, \sigma^2_{\eta,s}, \rho_s\}$, together with the share of households ϕ_s with a college education. 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 add to the Markov process described above two more states $\{\eta_{s,6}, \eta_{s,7}\}$. We augment the 5×5 Markov transition matrices $\pi_s = (\pi_{ij,s})$ as follows:

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

and 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. The transition matrix above reflects these assumptions which will permit us to match both the empirical earnings and wealth distribution (including at the top) very accurately.[9] In addition, we assume that $\eta_{n,7} = \eta_{c,7}$ and $\pi_{77,n} = \pi_{77,c}$. This leaves

[9] Recall that for the *highest* state wages are simply determined as $w \exp(\eta_7)$ and thus do not depend on the fixed effect α and the deterministic age profile; this formulation leads to a much better fit of the age-earnings and age-asset distributions.

us with ten additional parameters characterizing the labor productivity process which we summarize, including the empirical targets, in table 2. Appendix D gives the exact values of the transition probabilities and states of the Markov chains.

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

3.3.2 Preferences

We assume that the period utility function is given by

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

We exogenously set risk aversion to $\gamma = 2$ and $\chi = 1.67$ in order to obtain a Frisch elasticity of labor supply of $1/\chi = 0.6$. The disutility of labor parameter λ is chosen so that households spend, on average, one third of their time endowment on market work. Finally, the time discount factor β is chosen such that the capital-output ratio in the economy is equal to 2.9.

3.4 Government Policies

The two government policies we model explicitly are the tax system and the social security system.[10] The main focus of the paper is on the composition of the labor earnings and the capital income tax schedule, as well as the progressivity of the former, especially at the high end of the earnings distribution.

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

3.4.1 The Tax System

We assume that the labor earnings tax function is characterized by the marginal tax rate function $T'(y)$ depicted in figure 1. It is thus characterized by two tax rates τ_l, τ_h and two earnings thresholds $\bar{y}_l, \bar{y}_h$. Earnings below $\bar{y}_l$ are not taxed, earnings above $\bar{y}_h$ are taxed at the highest marginal rate τ_h, and for earnings in the interval $[\bar{y}_l, \bar{y}_h]$ 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 parameterized by few parameters and being continuously differentiable above the initial earnings threshold $\bar{y}_l$, 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 $\bar{y}_h$ 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.

For the initial equilibrium we choose the highest marginal tax rate $\tau_h = 39.6\%$, equal to the current highest marginal income tax rate of the federal income tax code.[11] That tax rate applies to labor earnings in excess of 4 times average household *income*, or $\bar{y}_2 = 4\bar{y}$. Households below 35% of median income do not pay any taxes, $\bar{y}_1 = 0.35y^{med}$ and we determine τ_l from budget balance in the initial stationary equilibrium, given the other government policies discussed below.[12] This requires $\tau_l = 12.2\%$, roughly the midpoint of the two lowest marginal tax rates of the current U.S. federal income tax code (10% and 15%). 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.

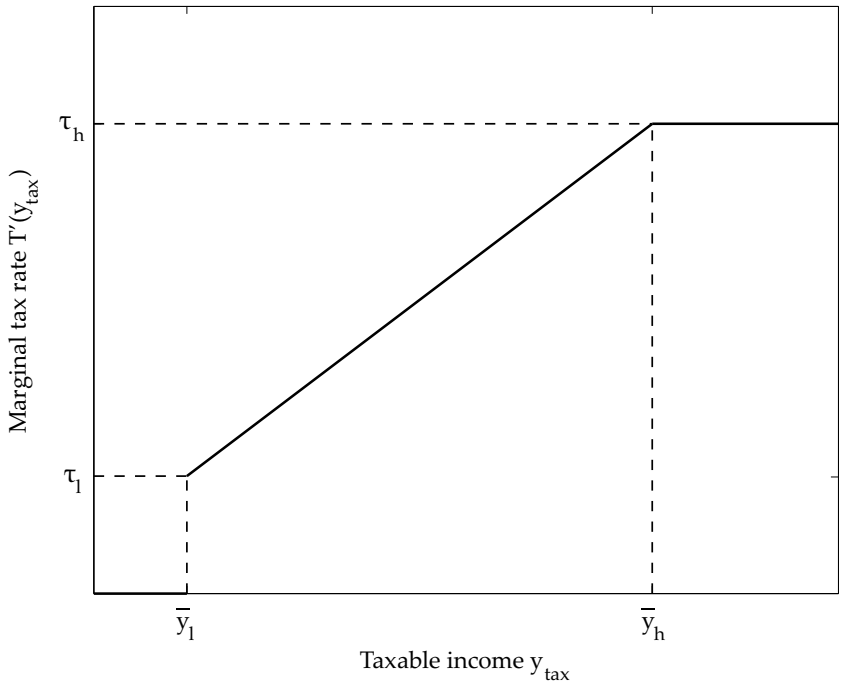
The initial proportional capital income tax rate is set to $\tau_k = 28.3\%$ and the consumption tax rate to $\tau_c = 5\%$. We choose exogenous government spending G such that it constitutes 17% of GDP; outstanding government debt B is set such that the debt-to-GDP ratio is 60% in the initial stationary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

---

[11] This value for the highest marginal tax rate is also close to the value assumed by Diamond and Saez (2011) once taxes for Medicare are abstracted from (we interpret Medicare as part of the social security system).

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

Figure 1: Marginal Labor Income Tax Function



3.4.2 The Social Security System

We model the social security system as a flat labor earnings tax τ_{ss} up to an earnings threshold $\bar{y}_{ss}$, 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, $\bar{y}(s, \alpha, \eta)$, and apply the actual progressive social security benefit formula $f(y)$ to $\bar{y}(s, \alpha, \eta)$. 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

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

We discuss the details of the benefit formula in appendix D.

3.5 Calibration Summary

The following 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 descibing current U.S. fiscal policy, as well as the capital share in production ϵ and the preference parameters (γ, χ). The choices for these paramters are standard relative to the literature, with the possible exception of the Frisch labor supply elasticity $1/\chi = 0.6$, 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.

Table 3: Exogenously Chosen Parameters

Parameter	Value	Target/Data
Survival probabilities $\{\psi_j\}$		HMD 2010
Population growth rate n	1.1%	
Capital share in production ϵ	33%	
Threshold positive taxation $\bar{y}_l$	35%	as fraction of y^{med}
Top tax bracket $\bar{y}_h$	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 $\bar{y}_{ss}$	200%	$\tau_p = 0.124$
Risk aversion γ	2	
Inverse of Frisch elasticity χ	1.67	

The set of parameters calibrated within the model include the technology parameters (δ_k, Ω), the preference parameters (β, 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 equilbrium is consistent with a capital-output ration 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 of the parameters.

4 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,

Table 4: Endogenously Calibrated Parameters

Parameter	Value	Target/Data
Technology level Ω	0.922	$w = 1$
Depreciation rate δ_k	7.6%	$r = 4\%$
Initial marginal tax rate τ_l	12.2%	Budget balance
Time discount factor β	0.977	$K/Y = 2.9$
Disutility from labor λ	36	$\bar{n} = 33\%$

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.

4.1 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

Variable	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	3.9%
Pension System	
Contribution rate (in %)	12.4%
Total pension payments	5.2%

All variables in % of GDP if not indicated otherwise

4.2 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	5.8	11.0	17.6	65.6	11.7	18.9	21.4	0.642
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 does the same for the wealth distribution. When comparing the model-implied earnings and wealth quintiles to the corresponding statistics from the data[13] we observe that the model fits the data very well, even at the top of the distribution. The same is true for the Gini coefficients of earnings and wealth.

Table 7: Wealth Distribution in Benchmark Economy

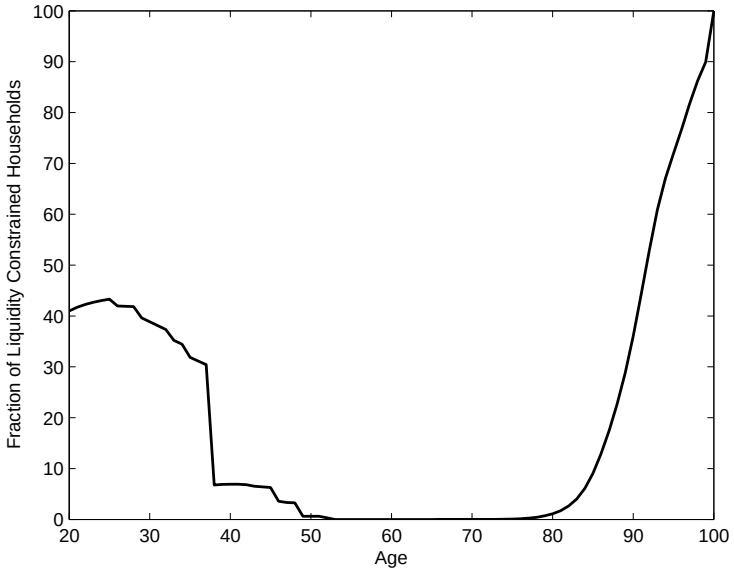
	Share of total sample (in %)								
	Quintiles					Top (%)			
	1st	2nd	3rd	4th	5th	90-95	95-99	99-100	Gini
Model	0.0	0.8	4.1	11.6	83.6	14.6	23.3	31.8	0.810
US Data	-0.2	1.1	4.5	11.2	83.4	11.1	26.7	33.6	0.816

An important aspect of the wealth data is that the bottom 40% of the population has essentially no net worth, something our model reproduces. Since a binding borrowing constraint significantly affects consumption, savings and labor supply choices, from the perspective of the model it is important to understand who are these households. In figure 2 we plot the share of each age cohort, in the initial stationary equilibrium, that has zero wealth. Not unexpectedly, binding liquidy constraints are mainly prevalent among young households and then are again observed among the very elderly who have outlived their resources and find it optimal to finance their consumption exclusively through social security benefits. We should note, however, that these elderly

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

households make up a fairly small fraction of the overall population since population growth as well as survival risk make cohort sizes above age 80 small relative to younger households.

Figure 2: Fraction of Households with $a = 0$, by Age



Overall, 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 in nature.

5 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 tax revenues 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 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 argue 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.

5.1 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 given the change in τ_h the government in addition (and again permanently) adjusts the entry marginal tax rate τ_l (but not the threshold $\bar{y}_l$) as well as $\bar{y}_h$ to assure both that the intertemporal budget constraint holds and that the top 1% earners are defined by the threshold $\bar{y}_h$ (in the first period of the policy-induced transition path). An appropriate sequence of government debt along the transition path insures that the sequential government budget constraints hold for very period t along the transition.

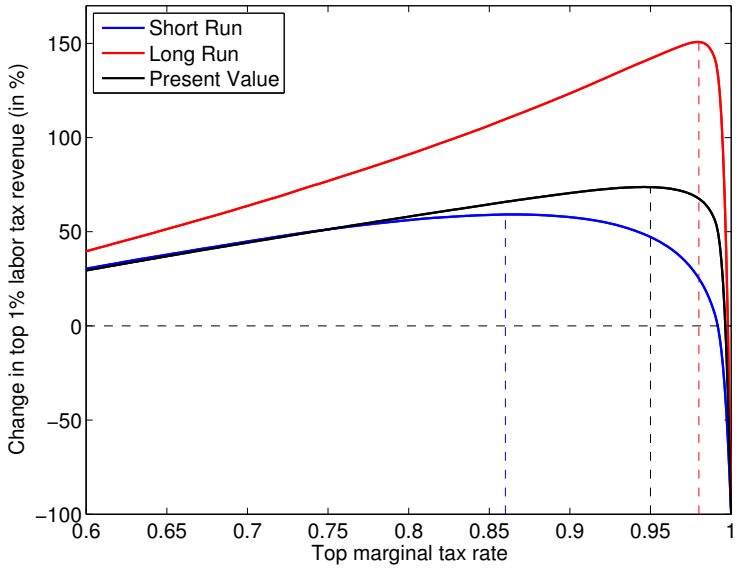
In the aggregate, a transition path is thus characterized by deterministic sequences of interest rates, wages and government debt $\{r_t, w_t, B_{t+1}\}_{t=1}^{T}$ converging to the new stationary equilibrium indexed by a new policy $(\tau_l, \tau_h, \bar{y}_l, \bar{y}_h)$. For every period $t \geq 1$ 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 insure that the intertemporal government budget constraint is satisfied.

5.2 Top Marginal Tax Rates and Tax Revenues

5.2.1 The Top 1% Laffer Curve in Our Economy

In figure 3 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.[14] 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 value discounted of all 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.

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



From this figure we observe that the revenue maximizing top marginal tax rate, independent of the time horizon used, is very high, 86% in the short run and 98% in the long run (and not surprisingly, in the middle at 95% when based on the present value of tax revenues). As we will see when displaying transition paths induced by the policy reform, in the short run top 1% earners have more wealth, and thus their labor supply responds more negatively to changes in the top marginal tax rate, lowering the revenue maximizing rate relative to the long run in which top earners are wealth-poorer and thus more inelastic towards tax changes.

Revenue-maximizing tax rates of course 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

[14] By construction this plot would be zero at a top marginal rate of 39.6%, but since the peek of this "Top 1% Laffer curve" is substantially to the right of the status quo we truncate the x-axis at 60%.

we first want to explore why the revenue-maximizing tax rates we find in our dynamic general equilibrium model are quite higher still than the 73% rate Diamond and Saez have advocated for.

5.2.2 Connecting Our Results to the Static Optimal Taxation Literature

Diamond and Saez' recommendation are based on the seminal paper by Saez (2001) who derives a concise formula for the revenue-maximizing top marginal tax rate in a static model of household labor supply that reads as:[15]

$$\tau_h = \frac{1}{1 \quad + \quad \underbrace{a \cdot e_c}_{\text{Subst. effect}} \quad - \quad \underbrace{(e_c - e_u)}_{\text{Inc. effect}}}$$

The parameter a governs the relationship between the top earnings threshold and mean labor earnings above this threshold.[16] The entities e_u, e_c are, respectively, the average (within the top 1% of earners) uncompensated and compensated elasticity of earnings with respect to 1 minus the constant marginal tax rate τ. Diamond and Saez (2011) assume that $e_u = e_c$ (that is, the absence of income effects), based on empirical studies argue for values of $a = 1.5$ and $e_u = e_c = 0.25$ for the top 1% of earners and thus end up with a revenue maximizing (and thus optimal) top marginal earnings tax rate of $\tau_h = 73\%$. We can compute the values of a, e_u, e_c implied by our model as well,[17] and if we do so we obtain values of $a = 1.59, e_c = 0.34, e_c - e_u = 0.34$, which, using Saez' formula, implies an optimal tax rate at the top of $\tau_h = 83.5\%$, close to our revenue maximization tax rate of $\tau_h = 86\%$ in the short run.[18] The slight difference between our result and that from Saez' formula is likely mainly due to the earnings elasticities not being completely policy invariant in our model.

The reason we find a still larger revenue-maximizing top marginal tax rate than Diamond and Saez (2011) recommend primarily stems from the fact that they implicitly assume that the income effect of tax changes on labor supply and thus earnings is zero, whereas it is significantly positive in our model for high income earners.

To summarize this discussion, the main reason that very high marginal tax rates are

[15] As long as the social welfare weight of top earners is negligible, this is also the welfare-maximizing top marginal tax rate.

[16] When earnings above the top earnings threshold follow a Pareto distribution then a is exactly the Pareto parameter of this distribution. Yet Saez (2001) formula doesn't rely on a Pareto distribution, but only on the relation between the top earners threshold $y^{1\%}$ and mean income above this threshold $y_m^{1\%}$ whereas a is defined as $\frac{y_m^{1\%}}{y_m^{1\%} - y^{1\%}}$.

[17] Details on the computation of these elasticities can be found in the appendix.

[18] The peak of the *short-run* Laffer curve is most comparable to the revenue-maximizing rate in the static models of Saez (2001) and Diamond and Saez (2011).

revenue (from the top earners) maximizing in our model is that a) the very top earners (the η_7 households) pay a dominant share of labor income taxes and b) their labor supply does not collapse even under very high marginal tax rates because of a strong income effect. Being in the rare (and fairly mean-reverting) high productivity state constitutes the chance of earning a large part of lifetime earnings for these households, and continues to be so even if marginal taxes rise to very high levels for these households.

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 have to first describe in detail how we measure social welfare, a task we tackle next.

5.3 Welfare-Maximizing Tax Rates

5.3.1 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 compensate every individual back to their initial equilibrium utility level, ex post for the currently living and ex ante for future generations.[19] That is, for each household currently alive we find the transfer $\Psi_0(j, s, \alpha, \eta, a)$ that satisfies

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

where v_0 denotes the value function in the initial steady state and 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 positive Ψ's constitute welfare *losses* from a given reform, relative to the status quo.

The total present discounted value of all transfers is then given by

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

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.

---

[19] These wealth transfers induce behavioral responses which we capture when computing the transfers necessary to make a household indifferent. We do however, abstract from the general equilibrium effects these hypothetical transfers would induce.

For future cohort the transfer is one number per cohort, for currently alive households the required transfers differ by characteristics (j, s, α, η, a). Future transfers are discounted to the present 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, with positive numbers indicating welfare gains.

In order to turn the welfare measure into easily interpretable numbers we turn the present value of the transfers into an annuity that pays out over the whole transition path and in the new long-run equilibrium and express the size of this annuity as a percent of initial aggregate consumption. That is, we calculate

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

where we again recall that positive required transfers W signal welfare losses from reform, and thus negative W constitute welfare gains. Expressing welfare gains in percent of consumption,

$$CEV = 100 * \frac{C}{C_0}$$

This idea of calculating the welfare consequences of policy reforms follows closely that of Huang et al. (1997), and more generally, the hypothetical lump-sum redistribution authority originally envisioned by Auerbach and Kotlikoff (1987) which would implement the transfer scheme described above. The advantage of this way of measuring aggregate welfare over alternative approaches is that it separates aggregate efficiency considerations from the potential desire of the policy maker to engage in intertemporal or intragenerational redistribution.[20]

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 (that is, prior to the education level s and fixed effect α being realized), would need to receive to be indifferent to being born into the steady state associated with a new policy configuration.[21]

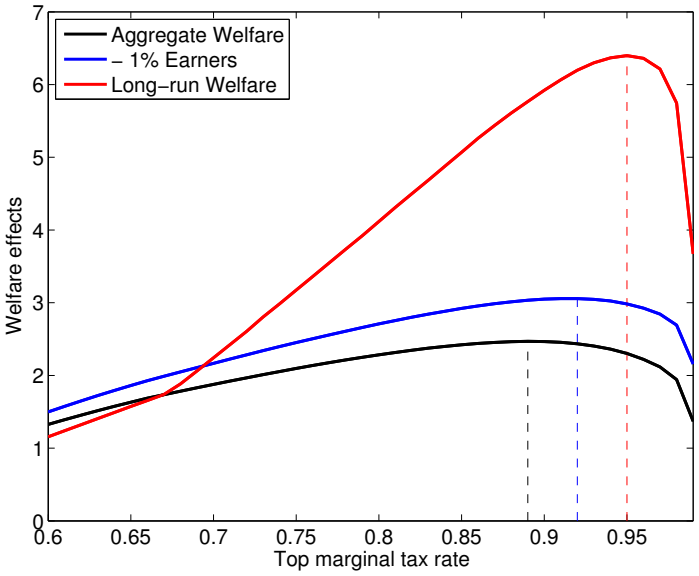
5.3.2 Optimal Size of the Top Marginal Earnings Tax Rate

In this section we document the optimal top marginal labor earnings tax rate. In figure 4 we plot three welfare measures against the top marginal tax rate τ_h. The black line plots the aggregate welfare measure, derived by compensating all current and future generations so that their (remaining) lifetime utility is the same as it would have been under the status quo fiscal policy. The blue line shows exactly the same plot, but ignoring lifetime welfare of the top 1% earnings households in the calculation of aggregate welfare. Finally the red line displays "long run welfare" measured as expected lifetime utility of households born into the steady state associated with a particular top marginal tax rate τ_h.

[20] Fehr and Kindermann (2014) 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.

[21] This is the same long-run welfare measure as employed by Conesa et al. (2009) in their study of optimal capital taxation.

Figure 4: Three Aggregate Welfare Measures as Functions of τ_h



The optimal top marginal tax rate is indeed very high, around 90%, under all three welfare measures. Welfare (as defined above) including the top 1% households, and including the transition effects is hump-shaped and maximized at $\tau_h = 89\%$. Recall that the top marginal tax rate that maximizes the present discounted value of tax revenues from the top 1% earners is 95%, somewhat higher than this welfare-optimal rate, but quantitatively close.[22]

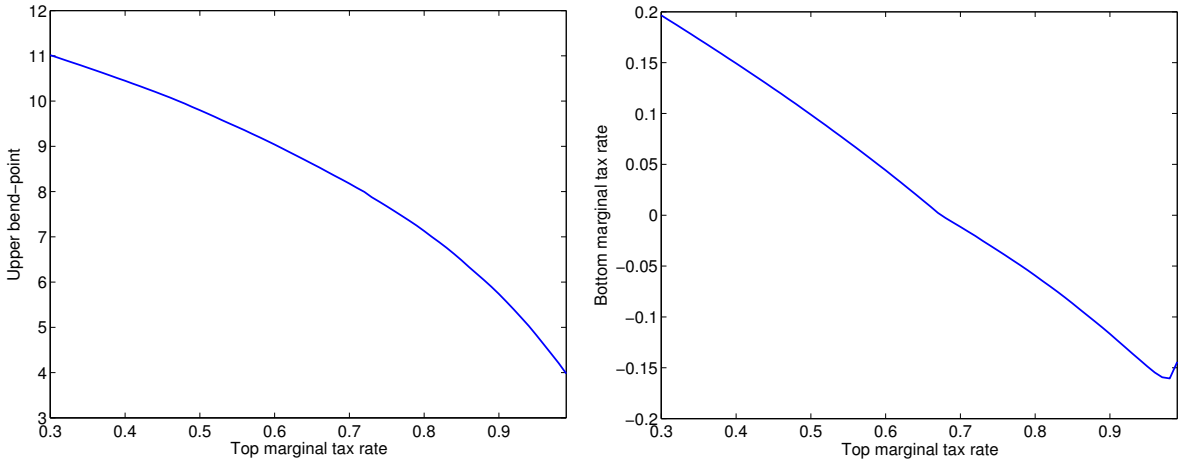
Excluding the top earnings group on which the top marginal tax rates apply leaves the figure broadly unchanged, but, not surprisingly raises the optimal top rate further, to $\tau_h = 92\%$. Focusing exclusively on welfare in the long run the optimal top marginal rate is even larger, at $\tau_h = 95\%$. Below we explore the sources of these welfare gains in greater detail. It is also noteworthy that the welfare gains induced by the very high marginal tax rates are very substantial, in the order of 2% of consumption when including the transition and in excess of 6% when comparing steady states.

As discussed above, in these thought experiments, as we vary τ_h we adjust the upper bend point $\bar{y}_h$ above which the highest marginal tax rate applies so that (in the first period of the transition) the top 1% of earners face this marginal tax rate. The intertemporal budget constraint of the government is balanced by adjusting the entry marginal

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

tax rate τ_l (and holding fixed the lower bend point $\bar{y}_l$ at which this rate applies).[23]

Figure 5: Upper Bend Point $\bar{y}_h$ as Fraction of Median Income, Bottom Marginal Tax Rate τ_h



In figure 5 we plot the required upper bend point $\bar{y}_h$ (left panel, as fraction of initial equilibrium median income) and the entry marginal tax (or subsidy) rate (right panel) against the top marginal tax rate τ_h. We observe that as the top marginal tax rate τ_h increases, due to endogenous responses in labor supply the earnings distribution compresses and the top 1% of earnings are obtained at a lower fraction of median earnings: the plot in the left panel is uniformly downward sloping. Second, since an increase in the highest marginal tax rate generates additional revenues for the government, the entry tax rate can fall and households pay lower taxes at the bottom of the distribution, and for sufficiently large τ_h even receive wage subsidies. These subsidies are initially increasing in τ_h (τ_l becomes more negative), but as the top marginal tax rate exceeds its 'peak of the Laffer curve' level at around $\tau_h = 95\%$ wage subsidies start to fall.

5.4 Understanding the Welfare Gains

In order to interpret the reported welfare gains from the optimal tax reform (and to understand why it is optimal in the first place) we now 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 output, *aggregate* consumption and the capital stock in the economy. Second, we quantify the redistributive

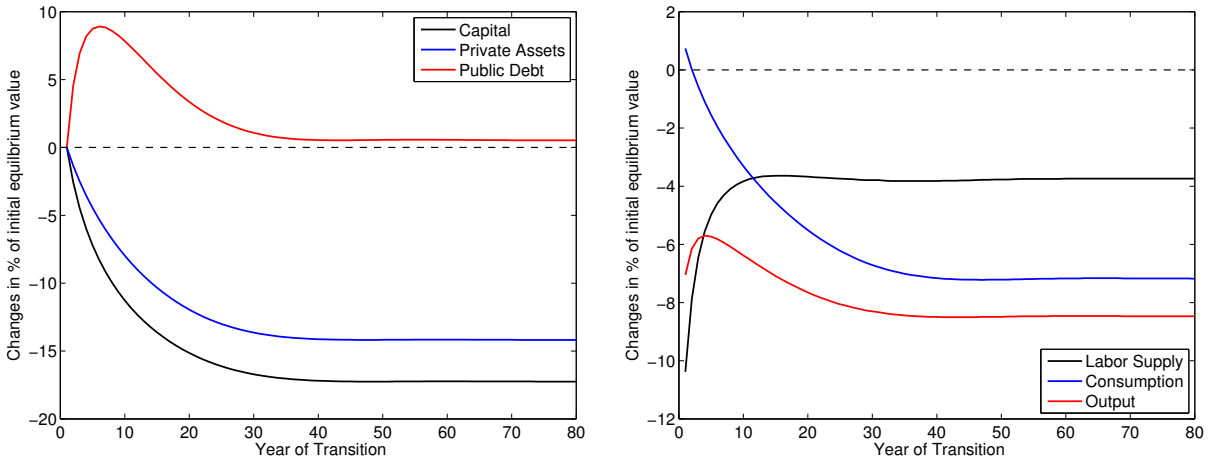
[23] If the required τ_l is non-negative, all households with earnings below $\bar{y}_l$ pay zero taxes, if τ_l is negative, all households with earnings below $\bar{y}_l$ receive a subsidy of τ_l per dollar earned, akin to the Earned Income Tax Credit in the U.S. Note that this slight asymmetry about how income below $\bar{y}_l$ is treated induces a small kink both in the welfare plot, figure 4 as well as in figure 5 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.

and insurance benefits of the reform, arguing that the latter are crucial for understanding our overall welfare results.

5.4.1 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. The path for 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.

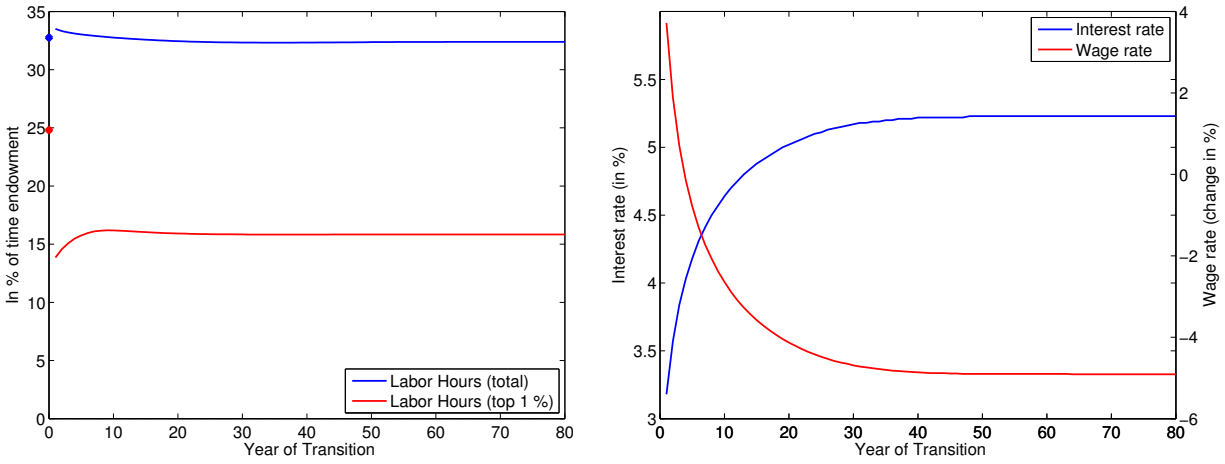
Figure 6: Capital, Assets, Government Debt; Labor Supply, Consumption and Output along Transition



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 10% and a corresponding fall of output by close to 7% (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 40%. 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 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

Figure 7: Hours and Prices along Transition

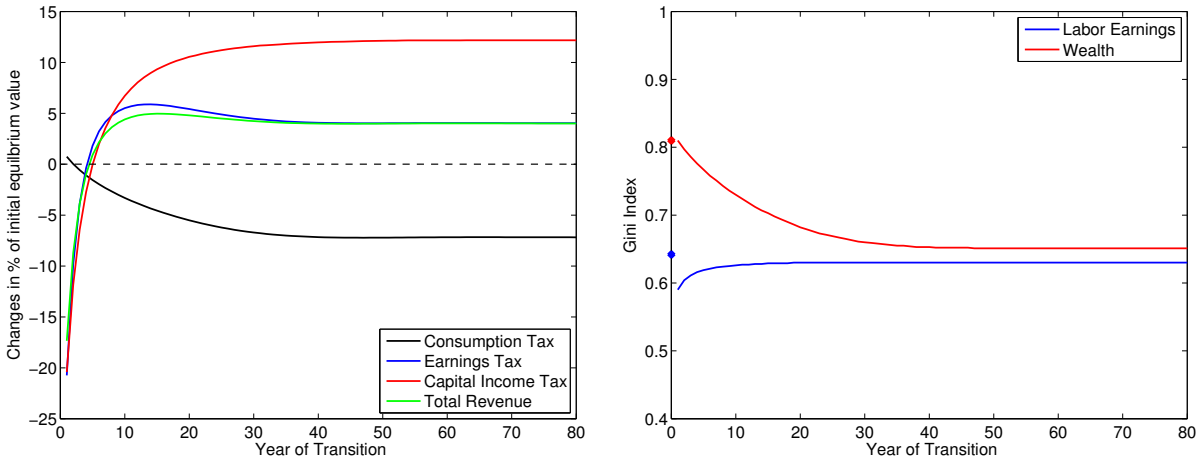


higher public debt used to finance the tax transition, but mainly driven by the decline in private saving of the high earners that are now subject to much higher marginal (and significantly higher average) labor earnings tax rates under the new tax system. Thus whereas in the short run most of the loss in output is absorbed by lower investment, in the long run aggregate consumption declines significantly as well, by about 7% (see again the 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 the consumption tax. On the other hand, once hours of the high productivity, high earnings top 1% 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 90% of all revenue from the labor earnings tax. Interestingly, revenues from capital income taxes also rise due to the higher return a reduced capital-labor ratio implies, despite the decline in the tax base for this tax.

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 (due to a negative wealth effect on leisure in light of falling net worth for this group), so does earnings inequality, without reaching its pre-reform level. Wealth inequality, on the other hand, is monotonically 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.

Figure 8: Tax Revenues and Inequality along Transition



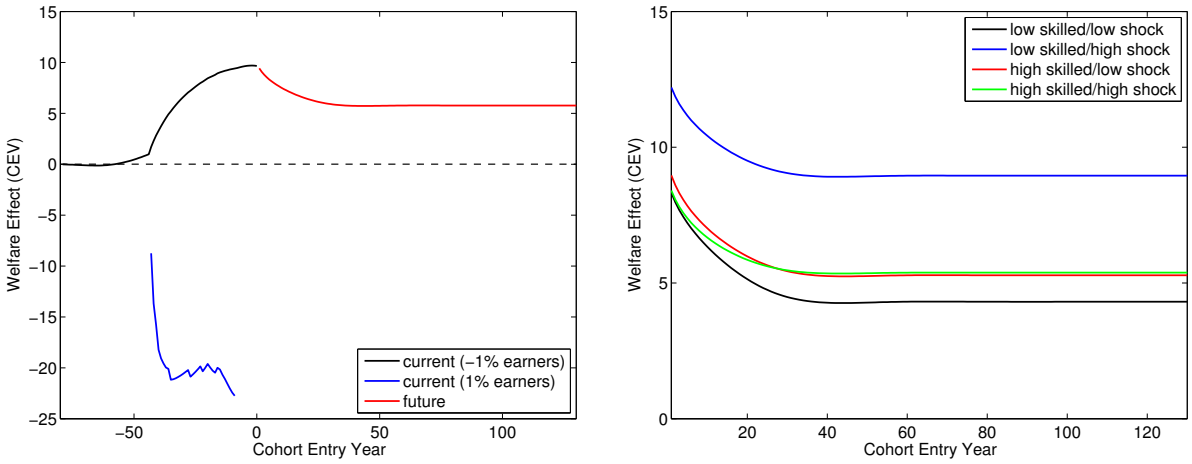
To summarize, the aggregate statistics indicate a massive decline in aggregate output and a (somewhat delayed) fall in aggregate consumption, coupled with a reduction of hours worked (mainly at the top of the earnings distribution). Furthermore, earnings and wealth inequality are significantly lower 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 5.3.2 come from enhanced social insurance and/or redistribution rather than from stimulating aggregate economic activity. In the next section we will provide a series of decompositions 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 (by a significant margin) the aggregate consumption losses, since these losses accrue exclusively to those few households that happen to rise to the very top of the earnings distribution.

5.4.2 Ex-Ante Redistribution or Ex-Post Insurance?

In order to understand why the tax system we characterized as optimal implies such substantial welfare gains despite its adverse impact on macroeconomic aggregates we start our decomposition analysis with a display of 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 5.1.

Figure 9: Welfare Effects of Policy Reform by Age Cohort



Not surprisingly, the welfare impact of the reform on the top 1% earners currently alive (which are evidently of working age) 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 the cohort is, since younger workers spend a larger share of their working life under the new tax regime. Finally, the welfare impact of future generations is positive and large, in excess of 6% of lifetime consumption. It is slightly declining along the transition as the economy consumes part of its capital stock, however.[24]

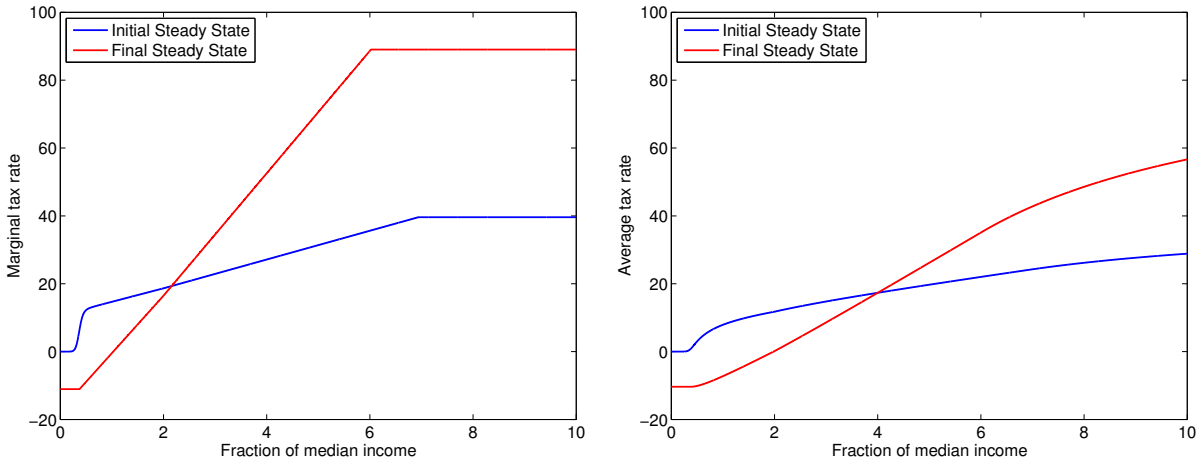
The right panel of figure 9 focuses on generations born after the implementation of the reform, but takes an ex-post (that is, after household type has been realized) perspective by disaggregating welfare gains from the tax reform by household type. Recall that our economy is populated by households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

[24] The aggregate welfare measures in section 5.3.2 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.

effect is significantly larger than the gains the other groups realize.

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 both marginal (left panel) and average (right panel) tax rates against labor earnings in the initial and the final steady state.

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



It shows that households with up to about four times median (initial steady state) earnings face lower average taxes whereas (not surprisingly, given the very high marginal taxes at the very top) high earners face massively higher marginal and average taxes. In figures 11 and 12 we display the differences in marginal and average tax rates between the two tax codes and insert box plots to summarize the earnings distribution in the model, both in the initial (figure 11) and final (figure 12) steady state. As our model is populated by four ex-ante heterogeneous household types who differ by their level of education and earnings fixed effect each panel of these figures 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 signified 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.

We want to draw attention to three main findings evident in these figures. First, the overwhelming majority of households is located in parts of the earnings distribution that faces lower average (but also lower 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 and especially average tax rates occurs among the middle class, households with earnings between 50% and 150% 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

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

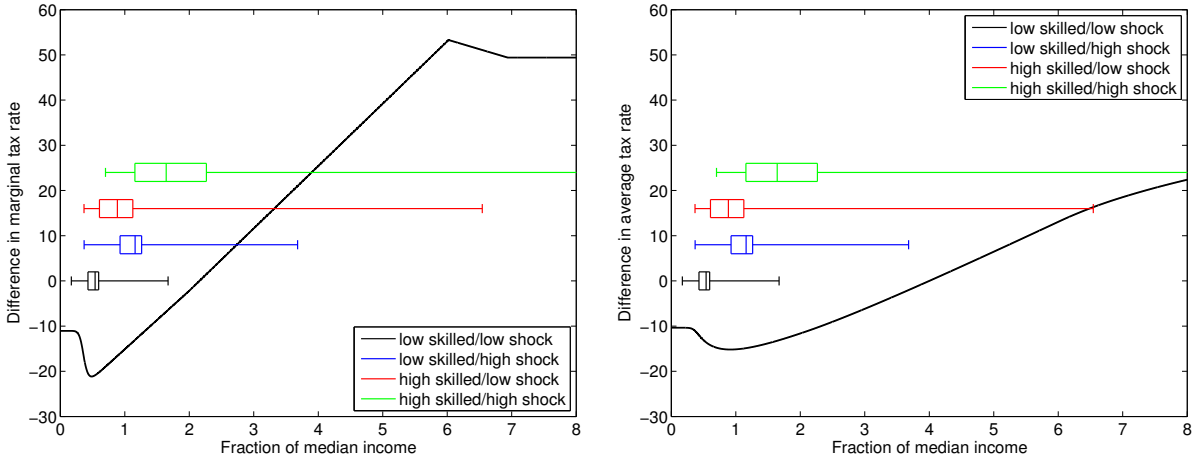
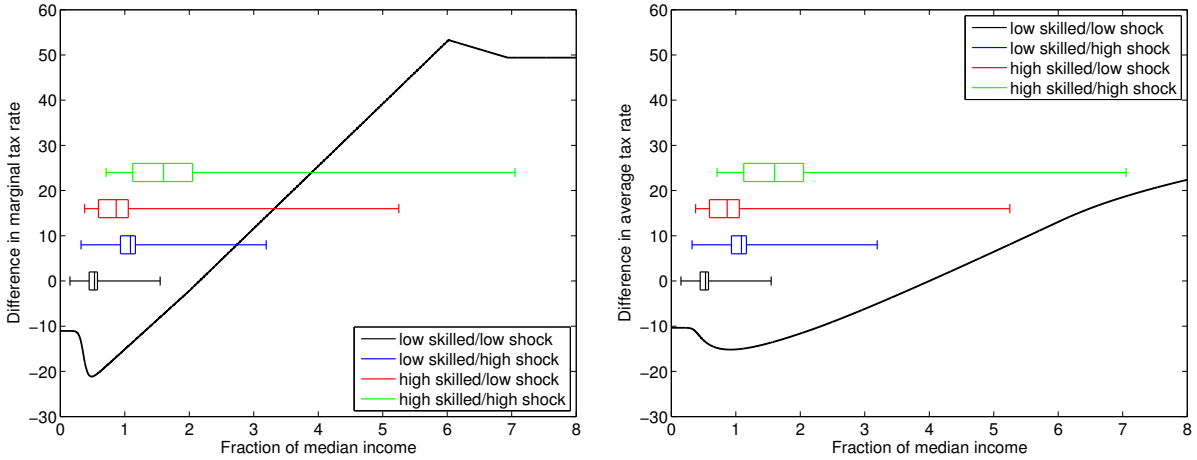


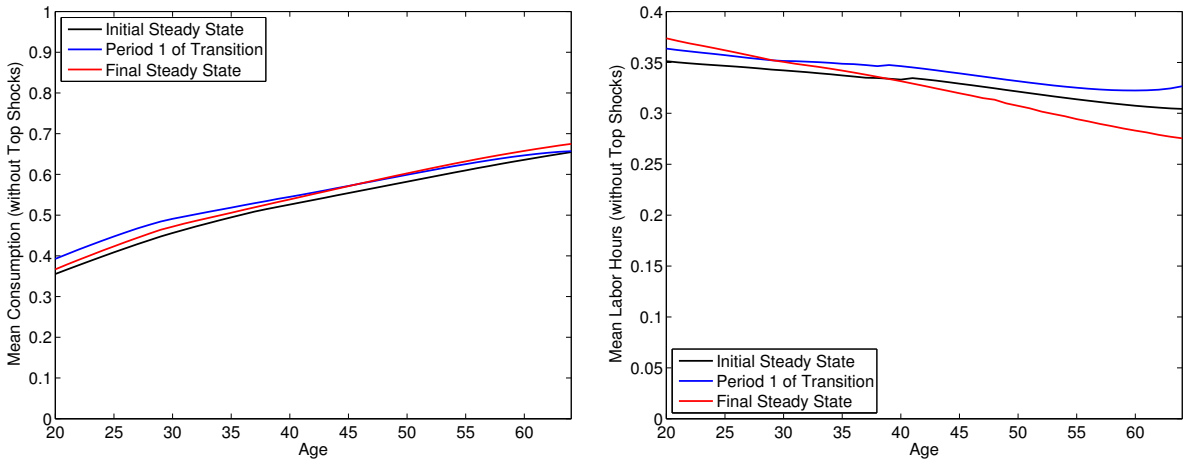
Figure 12: Difference in Tax Schedules and Earnings Distribution (by Type) in Final Steady State



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 if experiencing the same fate. 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.

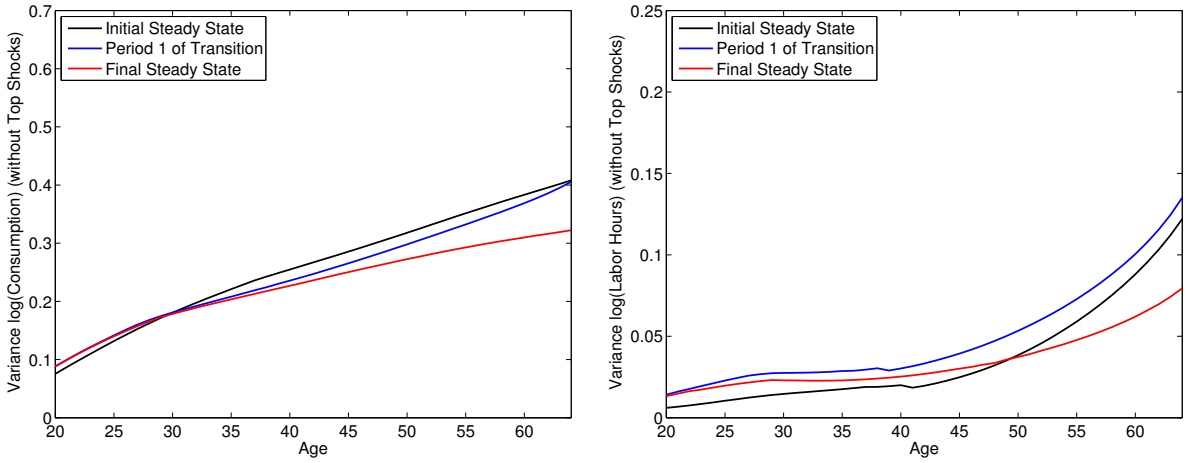
The previous discussion, however, does not clarify what are the *common* sources of the welfare gains of each of these four groups. To make progress along this front, in figure 13 we plot average consumption and hours worked over the life cycle, not counting consumption and hours occurring when households 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 cohort variance of consumption of hours, and

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



figures 15 and 16 repeat the same for the entire population, that is, they include the top productivity states in the calculation of the means and the 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 actually uniformly larger under the new, relative to the old tax system (comparing steady states), despite the fact that aggregate consumption is 7% 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 (and decline significantly at older ages in the new, relative to the old steady state). Couple this with a very sizable reduction of lifetime consumption risk (approximated by the within-cohort consumption variance), see the left panel of either figure 14 or 16, and the 7% steady state welfare gains emerge.

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

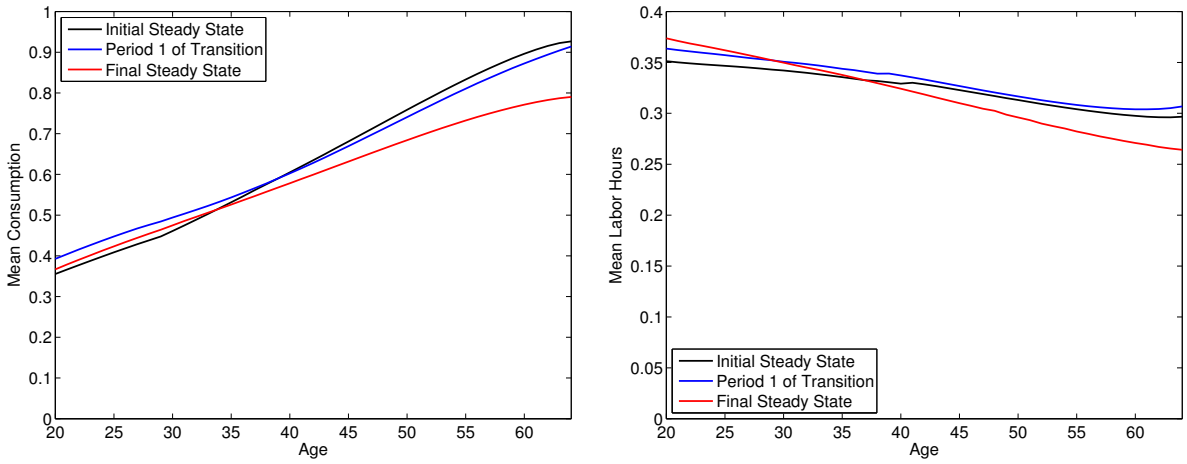
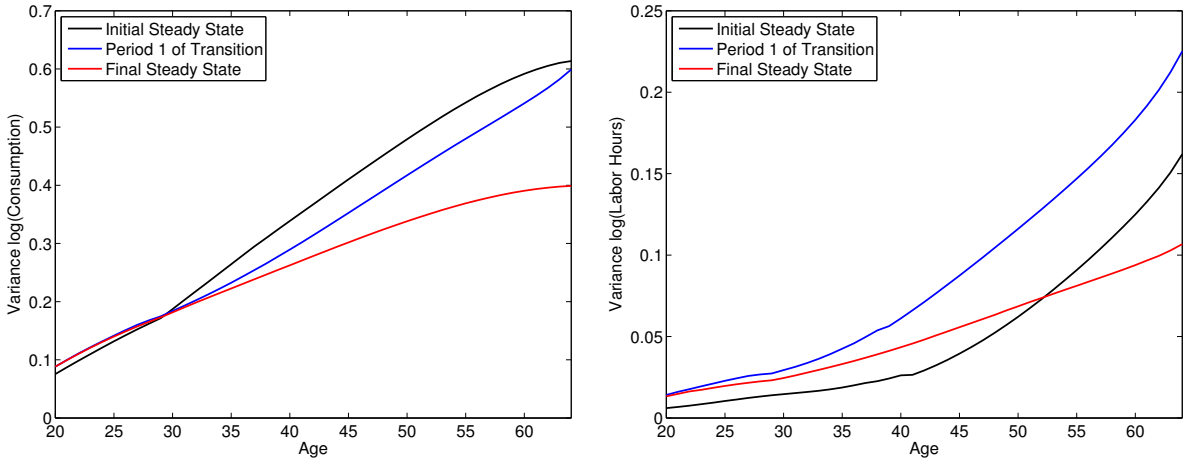


Figure 16: Variance of Consumption and Hours over the Life Cycles, Entire Population



5.5 Sensitivity Analysis

In this section we want to discuss the sensitivity of our results to the key modeling and parametric assumptions we have made so far.[25]

[25] Details on how we adjust the model to produce these results can be found in the appendix.

5.5.1 Realistic Income Inequality is Key for the Results

Suppose instead households face a labor productivity process that does *not* contain the small chance of very high wage and thus earnings realizations.[26] By implication, in this version of the model the earnings, income and wealth distributions will not display the degree of concentration observed in U.S. data, and thus it won't paint an accurate picture of who the top 1% are and what are their economic circumstances. This economy serves, however, a useful role for understanding what drives our results of desirable high marginal income tax rates for the top earners in society.

Figure 17: Laffer Curves and Welfare as Function of τ_h, Absent Top Productivity Shocks

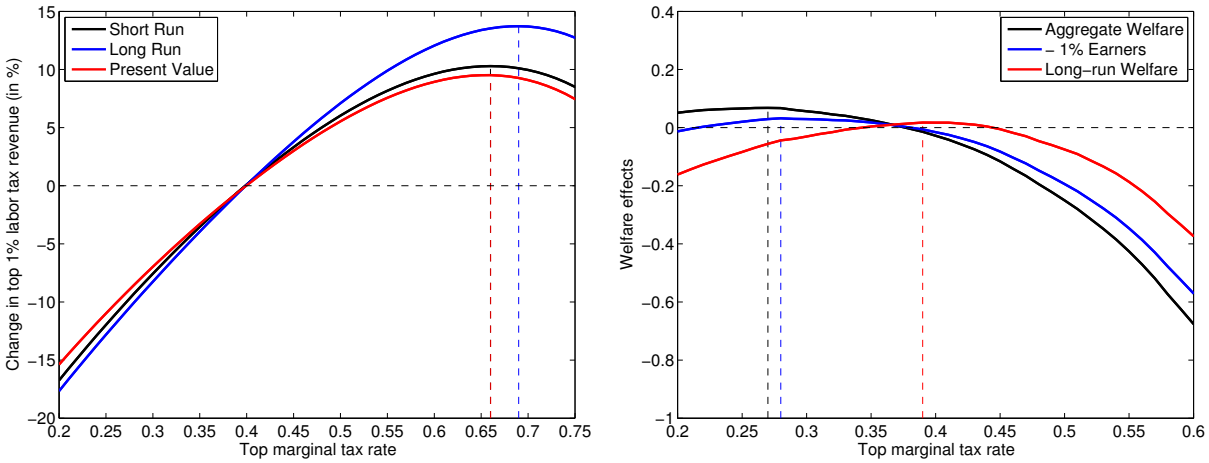


Figure 17 displays the top 1% Laffer curves (left panel) and welfare (right panel). As the figure shows, in the absence of the top two productivity shocks, and thus in the absence of a realistic degree of earnings and wealth dispersion, the optimal top marginal labor earnings tax rate falls, and independent of the welfare metric applied, is fairly close to the current rate of 39.6%. This happens for two reasons. First, the revenue-maximizing top marginal tax rate falls, to between 65% and 70% (rather than above 85%, as in the benchmark economy), on account of a smaller income effect of the now less-earnings rich top 1%. However, now the divergence between revenue-maximizing (from the top 1%) top tax rates (still above 65%) and welfare maximizing top tax rates (below 40%) is much more significant. Since the largest productivity realizations are now much less severe, the large social insurance benefits of high tax progressivity vanish.[27]

Overall, we conclude from this section that our main result of very high marginal tax

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

[27] Taking a steady state perspective the optimal labor income tax system is still substantially progressive, consistent with the findings in Conesa and Krueger (2006) and Fehr and Kindermann (2014).

rates for top earners depends crucially on a productivity process capable of producing earnings- and wealth rich household to the same extent as they are present in the data.

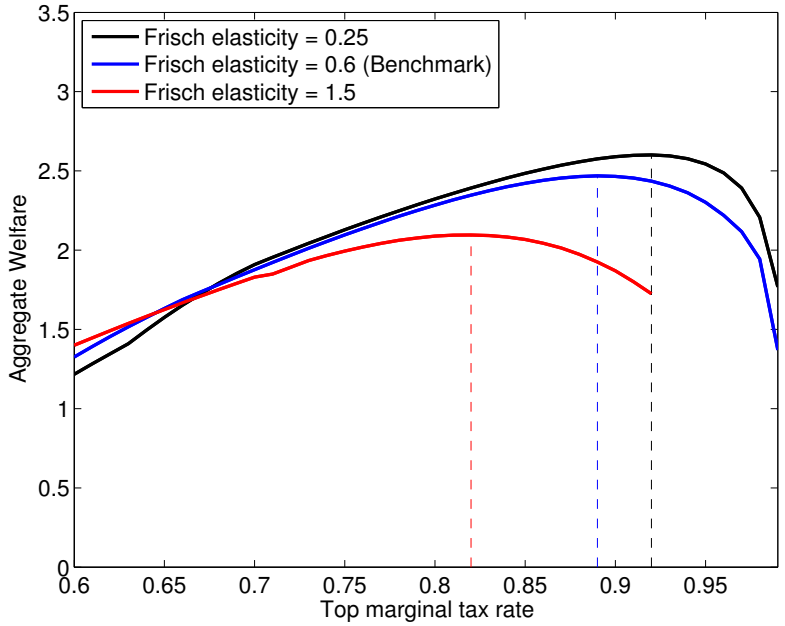
5.5.2 Labor Supply Elasticity

As most clearly seen in the simple formula by Diamond and Saez (2011) based on their static model, the optimal top marginal tax rate depends on the parameters governing elasticity of labor supply with respect to tax rates. Therefore we now conduct sensitivity analysis with respect to the Frisch elasticity parameter that governs the (intertemporal) labor supply elasticity in the dynamic model. In table 8 and figure 18 we document how our optimal tax and welfare results depend on the Frisch labor supply elasticity.[28]

Table 8: Sensitivity with Respect to Frisch Labor Supply Elasticity

Scenario	$\overline{\tau}$	$\underline{\tau}$	K	L	LR Wel.	Agg Wel.
Frisch elasticity = 0.25	92.00	-13.33	-15.82	-2.33	6.01	2.60
Benchmark =0.60	89.00	-11.05	-17.24	-3.74	5.76	2.47
Frisch elasticity = 1.50	82.00	-7.97	-17.19	-4.82	4.67	2.10

Figure 18: Aggregate Welfare as Function of τ_h, Different Frisch Elasticities



[28] For a high Frisch elasticity and tax rates above 92% we experienced convergence problems with our algorithm.

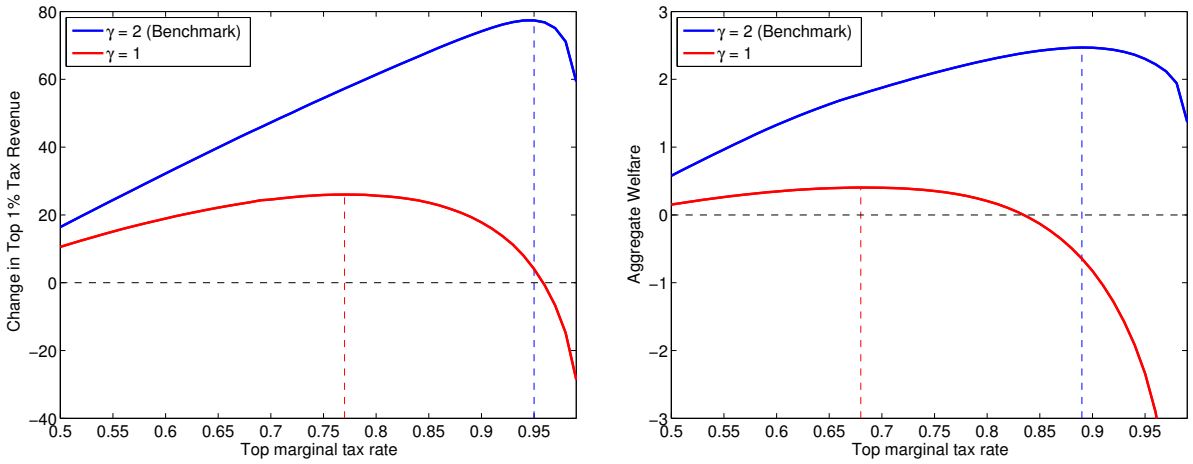
The key finding from the previous figure and table is that although the positive and normative results change in the expected direction (a larger elasticity reduces the size of the top marginal tax rate and the associated welfare gains from the policy reform), the differences are quantitatively fairly 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 80%.

As we have shown above in section 5.2.2 the optimal tax rate in our model 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 which remains broadly unaffected by the change in the parameter governing the Frisch elasticity. We conclude that our results are qualitatively, and to a very large extent quantitatively robust to plausible changes in the Frisch labor supply elasticity.

5.5.3 Size of the Income Effect

Given the importance of the income effect we now document how changes in its magnitude affect our results. To this end we now vary γ from 2 to 1, thereby assuming log-utility in consumption and making our preference specification consistent with balanced growth. In our model γ controls the size of the income effect, with smaller values implying smaller income effects and thus overall stronger responses of labor supply at the top to changes in marginal tax rates.

Figure 19: Laffer Curve and Aggregate Welfare, $\gamma = 2$ vs. $\gamma = 1$



In figure 19 we plot the top earner Laffer curve (the present discounted value version), both for risk aversion of $\gamma = 1$ and $\gamma = 2$, whereas the right panel does the same for aggregate welfare. We observe that the magnitude of the income effect is quantitatively very 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 77% and the welfare maximizing rate is 68%. Turning back to

the interpretation of the results through the lens of the compensated earnings elasticity and the income effect, the former rises (on average, for the top 1%) from $e_c = 0.34$ in the benchmark economy to $e_c = 0.45$ with log-utility, and the income effect falls from $e_c - e_u = 0.34$ to $e_c - e_u = 0.23$. Thus the income effect does not offset the compensated elasticity and the revenue-maximizing tax rate falls from 0.95 to 0.77, rather close to the original recommendation by Diamond and Saez (2011).

Second, as in the previous subsection (and in contrast to the benchmark model) the divergence between revenue maximization and welfare maximization becomes more important now as lower risk aversion shrinks the insurance benefits of highly progressive labor income taxes. Thus the socially optimal top rate is even lower now, but still at a very sizable 68%, 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. [29]

5.5.4 Persistence of High Productivity States

As final robustness check we assess to what extent 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 model permanent superstars, but be consistent with our benchmark model in which the probability of becoming very productive is essentially zero before age 30 we proceed as follows. We now assume that there is one shock at age 30 which transports an (education-specific) fraction of the population to the η_7 shock.[30] When a household climbs up to η_7 she will stay there forever. Reversely, after age 30 there is no chance of climbing up to this shock anymore.[31]

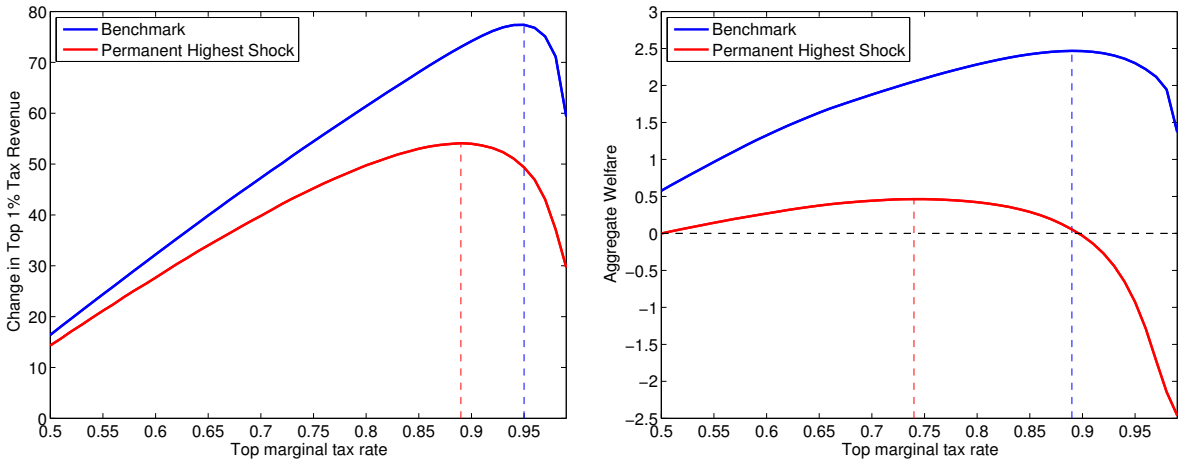
Figure 20 again plots the Laffer curve and aggregate welfare, both for the benchmark economy and the economy with permanent superstars. We observe that both the revenue maximizing tax rate (moderately) and the welfare maximizing tax rate (more significantly) falls, relative to the benchmark. The revenue-maximizing tax rate falls since the upper tail of the earnings distribution becomes significantly thinner (since permanently highly productive households strongly reduce their labor supply), and thus av-

[29] As an important additional distinction, with log preferences hours worked increase with productivity, whereas in the benchmark the correlation between productivity and hours was slightly negative.

[30] The persistence and probabilities for the η_6 shock remains untouched.

[31] This change in the model affects its ability to reproduce the empirically observed earnings and especially the wealth distribution. Whereas the earnings Gini remains close to its empirical counterpart, the wealth Gini falls from 81% to 77% and the wealth share of the top 1% decreases from 32% to 19%.

Figure 20: Laffer Curve and Aggregate Welfare, Persistent vs. Permanent Highest Wage Shocks



erage income within the top 1% earners falls. Second, again the difference between revenue maximization and welfare maximization becomes more sizable because, relative to the benchmark, in the economy with permanent superstars the insurance benefit of progressive taxes from falling back into the normal range of productivity disappears. However, even with these changes our main message remains robust: marginal tax rates on the top 1% earners in excess of 70% remain optimal from a social welfare point of view.

6 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 90%, fairly independently of whether the top 1% is included or excluded in the social welfare function, and independently of whether transitional or long run welfare is considered. 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 somewhat persistent) earnings opportunity, 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 persistent but mean reverting. 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 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 and Huggett (2014).

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] Badel, A. and M. Huggett (2013), "Taxing Top Earners: A Human Capital Perspective," *Federal Reserve Bank of St. Louis Working Paper*, 2014-017A.

[5] Bakis, O., B. Kaymak and M. Poschke (2013), "On the Optimality of Progressive Income Redistribution," Working paper.

[6] Bassetto, M. (1999), "Optimal Fiscal Policy with Heterogeneous Agents," *Working paper*.

[7] 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 Gérard Debreu*, North-Holland, Amsterdam, 79-102.

[8] Brüggemann, B. and J. Yoo (2014), "Aggregate and Distributional Effects of Increasing Taxes on Top Income Earners," *Mimeo*.

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

[10] 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

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

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

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

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

[15] 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.

[16] 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 No. 1.

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

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

[19] Fehr, H. and F. Kindermann (2014), "Optimal Taxation with Current and Future Cohorts," *forthcoming: Journal of Economicy Dynamics & Control*.

[20] Guner, N., M. Lopez-Daneri and G. Ventura (2014), "Heterogeneity and Government Revenues: Higher Taxes at the Top?," *Working Paper*.

[21] Holter, H., D. Krueger and S. Stepanchuk (2014), "How Does Tax Progressivity and Household Heterogeneity Affect Laffer Curves?," *Working Paper*.

[22] 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.

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

[24] Huggett, M. (1993), "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics & Control*, 17, 953-969.

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

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

[27] 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.

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

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

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

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

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

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

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

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

A 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.

A.1 Computation of Policy and Value Functions

We solve for policy and value functions using the method of endogenous gridpoints. 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

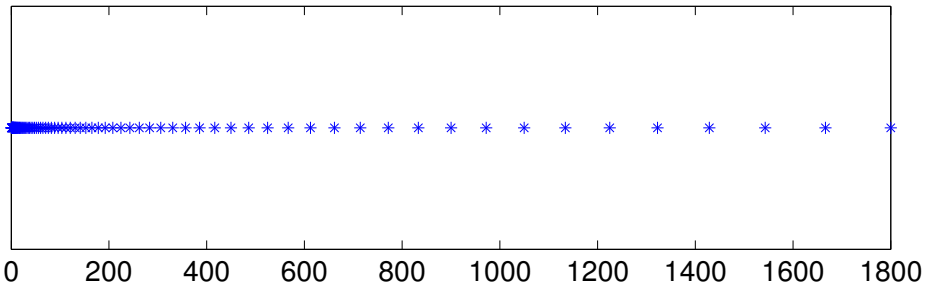
$$(j, s, \alpha, \eta, a) \in \{1, \ldots, J\} \times \{n, c\} \times \{-\sigma_\alpha, +\sigma_\alpha\} \times \{\eta_{s,1}, \ldots, \eta_{s,7}\} \times \{\hat{a}^1 \ldots, \hat{a}^{100}\}.$$

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.[32] A g_a of 0 would result in equidistantly spaced gridpoints. Setting $g_a > 0$ the distance between two successive gridpoints $\hat{a}^i$ 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

Figure 21: Discretized asset state space



borrowing constraints may occur and therefore policy functions may have kinks or be

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

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,

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

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

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

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

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

We now apply the method of endogenous gridpoints as follows: We assume that savings for tomorrow would amount to $a' = \hat{a}^i$ for all $i = 1,\ldots,100$. 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 $n^e(j,s,\alpha,\eta,\hat{a}^i)$.[33] Plugging these into the budget constraint, we can determine the endogenous gridpoint 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

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

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

Using the interpolation data

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

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 $\hat{a}^i, i = 1,\ldots,100$ by piecewise linear interpolation.[34]

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 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

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

[33] 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 rootfinding 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.

[34] We do not interpolate v_c^e and v_n^e directly, but rather $[(1-\gamma)v_c^e]^{1/(1-\gamma)}$ and $[(1+\chi)v_n^e]^{1/(1+\chi)}$ 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.

A.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 rootfinding 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 $\varepsilon = 10^{-6}$.

A.3 Calculation of Compensating Transfers

The calculation of compensating transfers is straightforward. In order to do so, we use a quasi-Newton rootfinding 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 rootfinding method, we have to solve for the optimal household decisions.

B Appendix: Computation of Elasticities

In order to apply Saez' formula

$$\tau_h = \frac{1}{1 \quad + \quad \underbrace{a \cdot e_c}_{\text{Subst. effect}} \quad - \quad \underbrace{(e_c - e_u)}_{\text{Inc. effect}}}$$

to all individuals above the 1% earnings threshold (henceforth denoted by $y^{1\%}$), we have to estimate the parameter a as well as the compensated and uncompensated elasticities of earnings with respect to the net of tax rate $1-\tau$ in our model.

The parameter a can be easily calculated from

$$a = \frac{y_m^{1\%}}{y_m^{1\%} - y^{1\%}} \quad \text{with} \quad y_m^{1\%} = \frac{\int y \cdot \mathbf{1}_{y \geq y^{1\%}} \, d\Phi_0}{\int \mathbf{1}_{y \geq y^{1\%}} \, d\Phi_0}.$$

$y_m^{1\%}$ denotes mean labor earnings of households above the 1% earnings threshold $y^{1\%}$. Note that Saez' formula also works when incomes are not Pareto distributed at the top, meaning it just relies on the relation between the threshold above which individuals should be taxed and the mean earnings above this threshold. Yet, if incomes are Pareto distributed at the top then a is the parameter of the Pareto distribution.

Determining the (Hicksian) compensated elasticity is complicated. Yet, we can exploit the Slutzky equation

$$e_c = e_u - i,$$

where $i = (1-\tau)\frac{\partial y}{\partial T}$ is the income effect associated to a marginal change in the net of tax rate, see Saez (2001). In order to determine the income effect i we proceed as follows: Starting from the initial equilibrium household decisions $n(j,s,\alpha,\eta,a)$ with a distribution over the state space of Φ_0, we adjust the tax schedule $T(y)$, such that each household has to pay a lump-sum transfer equal to 1% of his initial equilibrium income, i.e. we set

$$\tilde{T}(y) = T(y) + tr \quad \text{with} \quad tr = 0.01 \cdot w \cdot e(j,s,\alpha,\eta) \cdot n(j,s,\alpha,\eta,a).$$

We then resolve the household optimization problem and determine new labor supply decisions $\tilde{n}(j,s,\alpha,\eta,a)$. We can now determine the individual income effect of a household as

$$i(j,s,\alpha,\eta,a) = (1-T'(y)) \cdot \frac{w(j,s,\alpha,\eta) \cdot [\tilde{n}(j,s,\alpha,\eta,a) - n(j,s,\alpha,\eta,a)]}{-0.01 \cdot w(j,s,\alpha,\eta) \cdot n(j,s,\alpha,\eta,a)}$$

The total income effect then is

$$i = \frac{\int i(j,s,\alpha,\eta,a) \cdot \mathbf{1}_{y \geq y^{1\%}} \, d\Phi_0}{\int \mathbf{1}_{y \geq y^{1\%}} \, d\Phi_0}.$$

The derivation of the uncompensated elasticity of earnings with respect to the net of tax rate as well as the uncompensated elasticity $e_u = \frac{\partial y}{\partial(1-\tau)} \cdot \frac{1-\tau}{y}$ is equally straightforward. We therefore adjust the tax schedule so that marginal tax rates increase by 1%, i.e.

$$\hat{T}'(y) = T'(y) + 0.01.$$

We then again solve the household optimization problem to determine the new labor supply decisions $\hat{n}(j,s,\alpha,\eta,a)$. The individual uncompensated elasticity of earnings then can be calculated from

$$e_u(j,s,\alpha,\eta,a) = \frac{w(j,s,\alpha,\eta) \cdot [\hat{n}(j,s,\alpha,\eta,a) - n(j,s,\alpha,\eta,a)]}{-0.01} \cdot \frac{1 - T'(y) \cdot}{w(j,s,\alpha,\eta) \cdot n(j,s,\alpha,\eta,a)}.$$

According to Saez (2001) the correct uncompensated elasticity to use for his formula is income weighted, i.e.

$$e_u = \frac{\int e_u(j,s,\alpha,\eta,a) \cdot y \cdot \mathbf{1}_{y \geq y^{1\%}} \, d\Phi_0}{\int y \cdot \mathbf{1}_{y \geq y^{1\%}} \, d\Phi_0}.$$

Finally using the uncompensated elasticity and the income effect we can calculate the compensated elasticity from the formula given above.

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

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

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 3 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

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

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	0.969680	0.029310	0.000332	0.000002	0.000000	0.000676	0.000000
2	0.007328	0.969846	0.021984	0.000166	0.000000	0.000676	0.000000
3	0.000055	0.014656	0.969901	0.014656	0.000055	0.000676	0.000000
4	0.000000	0.000166	0.021984	0.969846	0.007328	0.000676	0.000000
5	0.000000	0.000002	0.000332	0.029310	0.969680	0.000676	0.000000
6	0.000000	0.000000	0.024837	0.000000	0.000000	0.949980	0.025183
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.267852	0.732148
$\exp(\eta_{n,i})$	0.1159	0.3405	1.0000	2.9369	8.6255	15.8180	1284.3139

and

$s = c$							
i, j	1	2	3	4	5	6	7
1	0.957899	0.028954	0.000328	0.000002	0.000000	0.012817	0.000000
2	0.007239	0.958063	0.021717	0.000164	0.000000	0.012817	0.000000
3	0.000055	0.014478	0.958118	0.014478	0.000055	0.012817	0.000000
4	0.000000	0.000164	0.021717	0.958063	0.007239	0.012817	0.000000
5	0.000000	0.000002	0.000328	0.028954	0.957899	0.012817	0.000000
6	0.000000	0.000000	0.028087	0.000000	0.000000	0.969688	0.002225
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.267852	0.732148
$\exp(\eta_{c,i})$	0.2112	0.4595	1.0000	2.1761	4.7353	7.3949	1284.3139

D.2 The Social Security System

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

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

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 $\bar{y}_{ss}$ in order to obtain a contribution rate of 12.4 percent.

E 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 insures 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 α 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.

E.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	0.969813	0.029314	0.000332	0.000002	0.000000	0.000539	0.000000
2	0.007329	0.969979	0.021987	0.000166	0.000000	0.000539	0.000000
3	0.000055	0.014658	0.970034	0.014658	0.000055	0.000539	0.000000
4	0.000000	0.000166	0.021987	0.969979	0.007329	0.000539	0.000000
5	0.000000	0.000002	0.000332	0.029314	0.969813	0.000539	0.000000
6	0.000000	0.000000	0.029262	0.000000	0.000000	0.949874	0.020864
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.181122	0.818878
$\exp(\eta_{n,i})$	0.1722	0.4149	1.0000	2.4101	5.8085	8.6515	276.8624

and

			$s = c$				
i, j	1	2	3	4	5	6	7
1	0.958451	0.028971	0.000328	0.000002	0.000000	0.012248	0.000000
2	0.007243	0.958615	0.021729	0.000164	0.000000	0.012248	0.000000
3	0.000055	0.014486	0.958670	0.014486	0.000055	0.012248	0.000000
4	0.000000	0.000164	0.021729	0.958615	0.007243	0.012248	0.000000
5	0.000000	0.000002	0.000328	0.028971	0.958451	0.012248	0.000000
6	0.000000	0.000000	0.070700	0.000000	0.000000	0.927047	0.002253
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.181122	0.818878
$\exp(\eta_{c,i})$	0.2809	0.5300	1.0000	1.8867	3.5597	4.3846	276.8624

E.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.