American Economic Journal: Macroeconomics 2022, 14(2): 319–366
https://doi.org/10.1257/mac.20150170

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

By Fabian Kindermann and Dirk Krueger*

This paper argues that high marginal labor income tax rates on top earners are an effective tool for social insurance even when households have high labor supply elasticity, households make dynamic savings decisions, and policies have general equilibrium effects. We construct a large-scale overlapping generations model with uninsurable labor productivity risk, show that it has a realistic wealth distribution, and numerically characterize the optimal top marginal rate. We find that marginal tax rates for top 1 percent earners of 79 percent are optimal as long as the model earnings and wealth distributions display a degree of concentration as observed in US data. (*JEL* D15, D31, H21, H24, H55, J22, J31)

In the last 40 years inequality in labor earnings, income, and wealth has increased substantially in the United States at the top of the distribution. Alvaredo et al. (2013) report that the share of household income accruing to the top 1 percent of income earners was about 10 percent in the early 1970s but now exceeds 20 percent in the United States.[1] At the same time, the highest marginal tax rate declined from levels consistently above 60 percent to below 40 percent. This triggered academic and popular calls to substantially raise top marginal income tax rates; see, e.g., Diamond and Saez (2011); Piketty and Goldhammer (2014); and Reich (2010); but also the Occupy Wall Street movement. In an influential paper, Diamond and Saez (2011) use a static, partial equilibrium labor supply model with households that differ in their labor productivity to argue that the revenue-maximizing tax rate on top 1 percent income earners is indeed high, at 73 percent. This is also the welfare-maximizing rate if the top 1 percent earners have zero weight in the social welfare function.

* Kindermann: University of Regensburg, CEPR, HCEO, and Netspar; Department of Economics, University of Regensburg (email: fabian.kindermann@ur.de); Krueger: University of Pennsylvania, CEPR, CFS, NBER, and Netspar; Department of Economics, University of Pennsylvania (email: dkrueger@upenn.edu). Richard Rogerson was coeditor for this article. We thank numerous seminar and conference participants as well as Juan Carlos Conesa, William Peterman, and Chad Jones for many useful comments. Krueger thanks the National Science Foundation for support under grants SES 1123547 and SES 1757084.

† Go to https://doi.org/10.1257/mac.20150170 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

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

The objective of this paper is to evaluate whether this recommendation of very high optimal marginal tax rates on top income earners is overturned in a dynamic general equilibrium macroeconomic model in which households are subject to uninsurable idiosyncratic labor productivity shocks, households make labor supply and intertemporal savings decisions, and the top-income households are valued in the social welfare function. An important result that emerges from our analysis is that such high marginal tax rates can be rationalized without appealing to redistributional concerns.

To insure that our model delivers an empirically plausible earnings and wealth distribution (relative to the evidence from the 2007 Survey of Consumer Finances), including at the very top end of the distribution, we calibrate a labor productivity process with superstar states directly to empirically observed top income and wealth shares, as in Castañeda, Díaz-Giménez and Ríos-Rull (2003). Consequently, in the model, the top 1 percent look exactly as in the data, at least with respect to their key economic characteristics. We find that the revenue-maximizing top rate is even higher than in the static model, at 87 percent, and the socially optimal rate—based on a consumption-equivalent variation welfare criterion—is smaller but quantitatively quite close, at 79 percent. This is consistent with the empirically observed levels after World War II. Using various decompositions of the welfare effects, we argue that this result is primarily driven by the social insurance against never making it into the top 1 percent earners that these high taxes imply, rather than due to purely redistributional motives of the government.

To be clear about our claims from the outset, we do not wish to argue that very high marginal tax rates on top earners are optimal in *all* dynamic models that generate empirically plausible earnings and wealth distributions. Rather, the objective of this paper is to show that when these distributions emerge from one prominent mechanism in the literature (large, persistent, but mean-reverting labor productivity shocks over the life cycle and associated precautionary savings) that enjoys some empirical support, as argued in Section VIIA, a strong normative argument for such high rates emerges naturally.

To make our argument, we first develop a simple, analytically tractable model in which we can theoretically characterize the revenue-maximizing and the welfare-maximizing tax rate on top income earners. We show that these rates depend on the elasticity of earnings with respect to the top marginal tax rate, the shape of the top income distribution, the use of the extra tax revenue, and the value of social insurance. Using this model, we also demonstrate that our welfare criterion effectively distinguishes between ex ante redistribution and ex post social insurance against idiosyncratic labor productivity risk.

To quantify the optimal marginal tax rate on the top 1 percent of earners, we then extend the simple model to a dynamic overlapping generations economy with ex ante skill heterogeneity, idiosyncratic wage risk, and endogenous labor supply and savings choices. The calibration of the model (as well as the sensitivity analysis) focuses on the parameters the simple model has identified as the key determinants of the revenue- and welfare-maximizing rates. We then use the calibrated version of the model to compute, within a restricted class of income tax functions, the optimal one-time tax reform, which in turn induces an economic transition from the current status quo (a stylized version of the current US income tax code) toward a new steady state. The key policy choice variable is the marginal tax rate applying to the top 1 percent. We find that this optimal top marginal tax rate along the transition path is indeed very high, at 79 percent. If we had ignored the transition path and the implied dynamics of the wealth distribution and had instead maximized steady-state welfare, the optimal marginal tax rate would be even higher, at 82 percent.

We then show that this result is primarily driven by the social insurance benefits that these high taxes imply. To match the very high concentration of labor earnings and wealth in the data, our model requires that households, with low probability, have the opportunity to work for very high wages. These high wages stand in for attractive entrepreneurial, entertainment, or sports opportunities in the real world. In the model, the labor supply of these households is not too strongly affected by very high marginal tax rates even with a utility function with high Frisch labor supply elasticity, as long as these households have not yet accumulated massive amounts of wealth, i.e., as long as they have not been superstars for too long. A strong negative income effect on leisure makes these households maintain their labor effort even as marginal tax rates rise. From the perspective of implementing social insurance against idiosyncratic labor productivity risk via the income tax code, it is then optimal to tax these incomes at a high rate.

After reviewing the literature, in Section I, we construct a simple version of our model to develop the economic intuition for our quantitative results. Sections II, III, and IV set out the quantitative model and discuss its calibration and its fit to the microeconomic data. In Section V we present and decompose our optimal tax results, and Section VI contains sensitivity analyses. Section VII concludes, and the online Appendix contains all theoretical derivations as well as details of the calibration and the computational algorithm.

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

Methodologically, our paper is most closely related to the quantitative dynamic optimal taxation literature.[2] Examples include Domeij and Heathcote (2004); Conesa and Krueger (2006); Conesa, Kitao and Krueger (2009); Bakış, Kaymak and Poschke (2015); Fehr and Kindermann (2015); and Hubmer, Krusell and Smith

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

(2020). A subset of this literature (see, e.g., Guner, Lopez-Daneri and Ventura 2016; Holter, Krueger and Stepanchuk 2019; or Badel, Huggett and Luo 2020) characterizes the relationship between tax rates and tax revenues (that is, the Laffer curve). In this paper we show that although the welfare-optimal top marginal tax rate is quantitatively smaller than the revenue-maximizing rate (from the top 1 percent), it is fairly close.[3]

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

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

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

I. Building Intuition: A Static Model of Labor Supply

In this section, we build some intuition for our results using a simplified static version of the full model employed in the quantitative analysis. We first set out

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

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

A. *A Simple Static Model*

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

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

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

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

$$T(z) = \tau_h \max(z - \bar{z}, 0) - R. \tag{2}$$

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

$$c_l = e_l n_l + R \qquad \text{and} \qquad c_h = \bar{z} + (1-\tau_h)(e_h n_h - \bar{z}) + R. \tag{3}$$

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

B. *The Top Laffer Curve*

We first focus on the tax revenue generated from the top income earners as a function of the top tax rate τ_h. We call this relation the Top Laffer curve. Evaluating this fiscal effect is of first-order importance for a welfare analysis, as it defines the absolute upper limit for the top rate.[4] Taxes paid by an individual with earnings $z_h \geq \bar{z}$ can be written as

$$T(z_h) = \tau_h(z_h - \bar{z}) - R. \tag{4}$$

Our objective is to characterize the top marginal tax rate τ_h that maximizes tax revenue from the top earners $T(z_h)$. The next proposition characterizes this rate.[5]

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

$$\tau_{Laffer} = \frac{1 - (a-1) \cdot \tau_a(\bar{z}) \cdot \epsilon(\tau_a(\bar{z}))}{1 + a \cdot \epsilon(z_h)} \tag{5}$$

where

$$\frac{a}{a-1} = \frac{z_h}{\bar{z}} \qquad and \qquad \tau_a(\bar{z}) = \frac{-R}{\bar{z}} \tag{6}$$

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

$$\epsilon(z_h) = \frac{dz_h}{d(1-\tau_h)} \cdot \frac{1-\tau_h}{z_h} \qquad and \qquad \epsilon(\tau_a(\bar{z})) = \frac{d\tau_a(\bar{z})}{d(1-\tau_h)} \cdot \frac{1-\tau_h}{\tau_a(\bar{z})} \tag{7}$$

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

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

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

PROOF:

See online Appendix A.

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

$$\tau_{Laffer} = \frac{1}{1 + a \cdot \epsilon(z_h)}. \tag{8}$$

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

In order to interpret the components of the revenue-maximizing tax rate in (5), consider first the case in Corollary 2. Formula (8), first derived by Saez (2001), is used by Diamond and Saez (2011) to make explicit policy recommendations about actual income taxes at the top of the distribution. It shows the trade-off between a *mechanical* increase in tax revenue of $d\tau_h(z_h - \bar{z})$ and a negative *behavioral* response $\tau_h dz_h$ stemming from the adjustment of labor supply and thus earnings z_h by top earners to changes in the top marginal tax rate. The less elastic earnings are to the tax rate (the lower is $\epsilon(z_h)$) and the fatter the right tail of the earnings distribution (the lower is a), the higher the revenue-maximizing rate τ_{Laffer}. Note that the elasticity $\epsilon(z_h)$ is a *policy elasticity* in the sense of Hendren (2016); that is, $\epsilon(z_h)$ summarizes the earnings reaction of top earners to *the specific tax experiment* considered here.

The full formula in (5) includes an additional *mechanical* effect that emerges when a change in the top rate is associated with a change of other elements of the tax schedule: in our simple model, an increase in the transfer R; in our quantitative analysis below, a reduction of marginal tax rates at lower incomes. In this case $\epsilon(\tau_a(\bar{z})) \neq 0$, and an increase in τ_h changes the average tax rate at the threshold income level $\bar{z}$. It thus also changes tax revenues from top earners on their earnings below the threshold $\bar{z}$.

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

C. *The Policy Elasticity*

Although the formula in (5) is general, its ingredients—especially the policy elasticity $\epsilon(z_h)$, but also the Pareto coefficient a—depend on the specific model and are typically not invariant to τ_h. To clarify which features and parameters of the model these statistics depend on, we now study the household optimization problem. In the context of this simple model we can analytically characterize the policy elasticity $\epsilon(z_h)$ that determines the peak of the top 1 percent Laffer curve in Proposition 1.

PROPOSITION 3: *The policy elasticity is*

$$\epsilon(z_h) = \epsilon_h^u - \eta_h \cdot \frac{\bar{z}}{z_h} \cdot \left[1 + \frac{\tau_a(\bar{z})}{1 - \tau_h} \cdot \epsilon(\tau_a(\bar{z}))\right], \tag{9}$$

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

$$\epsilon_h^u = \frac{(1-\gamma)(1-\tau_h) z_h + \tau_h \bar{z} + R}{\left(\gamma + \frac{1}{\chi}\right)(1-\tau_h) z_h + \frac{\tau_h \bar{z} + R}{\chi}} \tag{10}$$

and

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

PROOF:

See online Appendix A.

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

$$\epsilon(z_h) = \frac{1-\gamma}{\gamma + \frac{1}{\chi}}. \tag{11}$$

Proposition 3 and Corollary 4 characterize the determinants of the labor supply reaction to the tax policy experiment. They show that the elasticities of labor supply with respect to wages and exogenous income depend on the structural parameters of the model, including the preference parameters controlling income and substitution effects γ and χ, as well as those governing the income distribution as summarized by a. This is most transparent for a purely proportional tax system, as in Corollary 4, above.[6] In this case the policy elasticity $\epsilon(z_h) = \epsilon_h^u$ is a function solely of the structural preference parameters capturing the standard substitution effect χ and income effect γ of a change in net wages on labor supply. If $\gamma = 0$, the income effect is absent and the policy elasticity is $\epsilon(z_h) = \epsilon_h^u = \chi$. If $\gamma = 1$, then the policy elasticity is 0 as income and substitution effect cancel.

Thus, the policy elasticity of top labor earnings with respect to a change in the top tax rate—and therefore, the peak of the Laffer curve—depends crucially on the Frisch elasticity χ parameterizing the substitution effect, the Pareto coefficient a summarizing the top of the earnings distribution, and the parameter γ determining the size of the income effect. Consequently, we will place special focus on calibrating these parameters in Section III.

[6]However, Proposition 3 also clarifies that the precise tax reform matters for the income effect and thus the policy elasticity, including the level of $\bar{z} > 0$, as well as the changes in other parts of the tax schedule, since then $\epsilon(\tau_a(\bar{z})) \neq 0$. We discuss these effects in detail in online Appendix A.

Table 1—Household Types

Type	Mass	Risk	Productivity	CEV
l	$(1-\Psi)\,\Phi_l$	No	e_l	T_l
h	$(1-\Psi)(1-\Phi_l)$	No	e_h	T_h
u	Ψ	Yes	e_l with prob. Φ_l e_h with prob. $1-\Phi_l$	T_u

D. *From the Laffer Curve to Welfare*

Maximizing tax revenues and thus the size of transfers R is welfare maximizing if and only if top earners receive no weight in the social welfare function. We now characterize welfare-maximizing top marginal rates. To distinguish between the benefits of high top marginal rates due to redistribution between ex ante different individuals and the benefits of social insurance against the risk of not becoming a top earner, we now augment the model. As before, ex post a share Φ_l of the population has productivity e_l and a share $1-\Phi_l$ has productivity e_h. However, now a share $(1-\Psi)$ of households know their productivity ex ante, whereas a fraction Ψ of individuals faces productivity uncertainty. Financial markets are incomplete as in the quantitative model and thus, by assumption, no explicit insurance contracts against this idiosyncratic risk can be traded. The three types of households are summarized in Table 1.

By construction, the ex post productivity distribution has Φ_l individuals with e_l and $1-\Phi_l$ individuals with e_h. The parameter Ψ measures the share of individuals with productivity risk and thus income risk and therefore governs the degree to which social insurance can be beneficial. With $\Psi = 0$, this is the model analyzed so far. Furthermore, since ex post, after productivity has been realized, the labor supply of individuals without and with income risk is identical, the Laffer curve analysis in the preceding sections goes through completely unchanged.

In order to make our welfare points as transparent as possible, we focus on preferences without income effects. However, in order to capture potential benefits from social insurance, we require households to be risk-averse; therefore, we assume that individuals have Greenwood, Hercovitz, and Huffman (1988), or GHH, -style preferences:

$$U(c,n) = \frac{\left(c - \dfrac{n^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right)^{1-\gamma}}{1-\gamma}, \tag{12}$$

where $\gamma = 1$ is understood to imply logarithmic utility. With these preferences, optimal labor supplies and consumption are given by

$$n_l^* = [e_l]^{\chi} \quad \text{and} \quad c_l^* = [e_l]^{1+\chi} + R, \tag{13}$$

$$n_h^* = \big[(1-\tau_h)\,e_h\big]^{\chi} \quad \text{and} \quad c_h^* = \big[(1-\tau_h)\,e_h\big]^{1+\chi} + \tau_h\bar{z} + R, \tag{14}$$

and we can express the expected utility of each type as functions of the tax rate and the lump-sum transfer. Realized utilities of the types with certain productivity are given as

$$(15)\qquad V_l(\tau_h) \;=\; \frac{\left[\frac{[e_l]^{1+\chi}}{1+\chi} + R\right]^{1-\gamma}}{1-\gamma} \text{ and } V_h(\tau_h) \;=\; \frac{\left[\frac{[(1-\tau_h)\,e_h]^{1+\chi}}{1+\chi} + \tau_h\bar{z} + R\right]^{1-\gamma}}{1-\gamma}.$$

Expected utility of individuals with ex ante uncertain productivity is determined as

$$(16)\qquad V_u(\tau_h) \;=\; \Phi_l V_l(\tau_h) + (1-\Phi_l)\,V_h(\tau_h).$$

Finally, the government budget constraint relates the lump-sum transfer R received by everyone to the taxes paid by high-income individuals:

$$(17)\qquad R \;=\; [1-\Phi_l]\,\tau_h\Big[(e_h)^{1+\chi}(1-\tau_h)^{\chi} - \bar{z}\Big].$$

Measuring Social Welfare.—We measure social welfare as consumption-equivalent variation. Specifically, we ask how much of a transfer T_i we would have to give to an agent of type $i = l,h,u$ in a situation with a zero top marginal tax rate to make this very agent as well off as in a situation with tax rate τ_h. For types $i \in \{l,h\}$, these transfers are determined as

$$(18)\qquad \frac{\left[\frac{[e_l]^{1+\chi}}{1+\chi} + T_l\right]^{1-\gamma}}{1-\gamma} \overset{!}{=} V_l(\tau_h) \qquad \text{and} \qquad \frac{\left[\frac{[e_h]^{1+\chi}}{1+\chi} + T_h\right]^{1-\gamma}}{1-\gamma} \overset{!}{=} V_h(\tau_h)$$

and thus $T_l = R$ and $T_h = [(1-\tau_h)^{1+\chi} - 1]\big((e_h)^{1+\chi}/(1+\chi)\big) + \tau_h\bar{z} + R$. For individuals with ex ante income risk, this transfer is determined as the solution to

$$(19)\qquad \Phi_l\frac{\left[\frac{[e_l]^{1+\chi}}{1+\chi} + T_u\right]^{1-\gamma}}{1-\gamma} + (1-\Phi_l)\frac{\left[\frac{[e_h]^{1+\chi}}{1+\chi} + T_u\right]^{1-\gamma}}{1-\gamma} \overset{!}{=} V_u(\tau_h).$$

The equivalent variation-based welfare measure is then given by

$$(20)\qquad \mathcal{V}(\tau_h;\Psi) \;=\; (1-\Psi)\big[\Phi_l T_l(\tau_h) + (1-\Phi_l)\,T_h(\tau_h)\big] + \Psi T_u(\tau_h)$$

and the government seeks to maximize $\mathcal{V}(\tau_h;\Psi)$ by choice of the top marginal rate τ_h, with transfers implied by the government budget constraint (17). To characterize this rate, and also as a point of contrast, it is also useful to define utilitarian social welfare as

$$(21)\qquad \mathcal{W}(\tau_h) \;=\; (1-\Psi)\,\Phi_l V_l(\tau_h) + (1-\Psi)\,\Phi_h V_h(\tau_h) + \Psi\,V_u(\tau_h)$$
$$= \Phi_l V_l(\tau_h) + (1-\Phi_l)\,V_h(\tau_h).$$

Theoretical Characterization of the Optimal Top Marginal Rate.—We now theoretically analyze the welfare-maximizing top marginal income tax rate. First, the next proposition characterizes the top rate a utilitarian planner would choose.

PROPOSITION 5: *Assume that* $e_l \leq \min\{\bar{z}^{\frac{1}{1+\chi}}, (\chi/(1+\chi))e_h\}$.[7] *Then the unique top marginal tax rate* τ_h^U *maximizing utilitarian social welfare* $\mathcal{W}(\tau_h)$ *is:*

(*i*) *independent of* Ψ *and thus independent of the benefit of social insurance*;

(*ii*) *equal to zero if households are risk-neutral*, $\tau_h^U = 0$ *if* $\gamma = 0$; *and strictly positive if households are risk averse—if* $\gamma > 0$, *then* $0 < \tau_h^U < \tau_{Laffer}$;

(*iii*) *strictly increasing in* γ *and productivity dispersion* e_h/e_l *and income dispersion* z_h/z_l.[8]

PROOF:

See online Appendix A.

The fact that utilitarian social welfare $\mathcal{W}(\tau_h)$ is *independent* of Ψ suggests that it is not a good measure for disentangling welfare gains from redistribution between ex ante heterogeneous households $i \in \{l, h\}$ and from social insurance for ex ante identical households $i = u$. This leads us to adopt the equivalent variation-based welfare criterion $\mathcal{V}(\tau_h; \Psi)$ for the remainder of this paper. The next proposition characterizes τ_h^* maximizing $\mathcal{V}(\tau_h; \Psi)$.

PROPOSITION 6: *Maintain the same assumptions as in Proposition 5. Then, the welfare-maximizing top marginal rate* τ_h^* *satisfies the following properties:*

(*i*) *If* $\Psi = 0$, *then* $\tau_h^* = 0$; *and if* $\chi > 0$, *then* $\mathcal{V}(\tau_h; \Psi = 0) < 0$ *for all* $\tau_h > 0$.

(*ii*) *If* $\Psi = 1$, *then* $\tau_h^*(\Psi = 1) = \tau_h^U < \tau_{Laffer}$.

(*iii*) $\tau_h^*(\Psi)$ *is strictly increasing in the importance of social insurance* Ψ.

PROOF:

See online Appendix A.

Item (i) says that if $\Psi = 0$ and there are no benefits from social insurance, then $\tau_h^* = 0$. The benefits of transfers to poor households l are exactly offset by the costs of these transfers away from the rich. Since taxes distort the labor supply

[7] This assumption ensures that even at the revenue-maximizing tax rate τ_h^{Laffer} the utility of high-productivity individuals is larger than that of low-productivity individuals, and that low-productivity individuals have earnings below the threshold $\bar{z}$, at which the top marginal tax rate τ_h sets in.

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

of top earners if $\chi > 0$ and labor supply is elastic, they are strictly suboptimal. Hence, $\mathcal{V}(\tau_h; \Psi = 0) < 0$ for all $\tau_h > 0$. Thus, if the population is composed of ex ante different types and the tax system merely redistributes between these types, there are no welfare gains from a positive top marginal rate and strict welfare losses if the tax system distorts labor supply.

The second and third parts show that as the social insurance benefits measured by Ψ increase, so does the optimal top rate τ_h^*, reaching its maximum at the tax rate that maximizes utilitarian social welfare τ_h^U when the population is composed only of ex ante identical households facing income risk. That rate is strictly increasing in the extent of income risk and risk aversion of households γ, as shown in Proposition 5.

Therefore, under a welfare metric based on equivalent variation, the benefits of a pure transfer between ex ante different types are zero, and they are negative if labor supply decisions are endogenous and distorted. If, by contrast, the scope for social insurance among ex ante similar individuals is high (because there are many such households, and because they face significant income risk or because they strongly value insurance due to high risk aversion), then the optimal rate is positive, close to the utilitarian rate τ_h^U, but below the revenue-maximizing Laffer rate τ_{Laffer}.

Back-of-the-Envelope Quantification.—To get a first sense of the magnitude of the optimal tax rate, how it relates to the revenue-maximizing rate, and how it depends on the scope of social insurance, we perform a back-of-the-envelope calculation of the simple model. We calibrate the model, motivated by Diamond and Saez (2011), such that the peak of the Laffer curve lies at exactly the 73 percent rate these authors advocate for. To make this section consistent with our quantitative work, we choose a policy elasticity of $\epsilon(z_h) = 0.21$ and productivity dispersion resulting in a Pareto coefficient for earnings of $a = 1.79$.[9] Section IIIC offers more details on how we arrive at these choices. Therefore, in this section, we choose $\chi = 0.21$, since in the absence of income effects the policy elasticity equals χ, and target incomes of $z_l = 0.63$, $\bar{z} = 7.62$, $z_h = 17.18$, resulting in a Pareto coefficient of $a = 1.79$ and a ratio of labor incomes between the top 1 percent and bottom 99 percent of 27, as in the quantitative model.[10] We set risk aversion $\gamma = 1.5$, as in the quantitative model.

In Figure 1 we plot, against the share of people facing income risk Ψ, the revenue-maximizing rate and the welfare-maximizing rate for the benchmark calibration (solid black line), as well as for economies with alternative parameters, motivated by our quantitative sensitivity analysis in Section VII of the paper. By construction, the peak of the Laffer curve is at 73 percent, independent of Ψ.[11] The figure shows that the optimal rate τ_h^* is 0 at $\Psi = 0$, and is strictly increasing in Ψ, as Proposition 6 has demonstrated. Importantly, for Ψ close to a value of one,

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

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

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

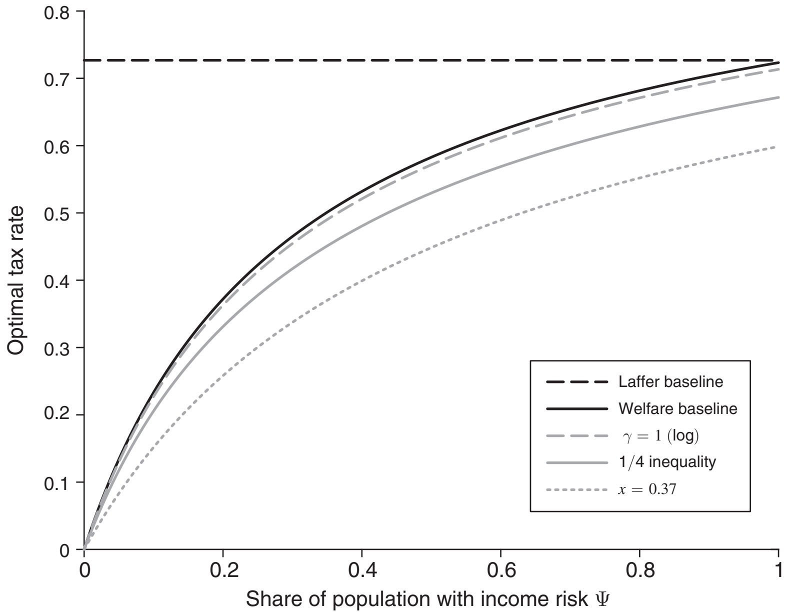

FIGURE 1. OPTIMAL TAX RATE

when most people face significant income risk and, thus, the scope for social insurance is large, the optimal rate τ_h^* is close to the revenue-maximizing rate of 73 percent. At $\Psi = 1$, the optimal rate (equal to the utilitarian rate τ_h^U) equals 72 percent, and remains above 60 percent as long as a majority of individuals face income risk ($\Psi \geq 50$ percent).

These findings are robust to plausible changes in parameters. Lowering risk aversion (and thus the benefit of social insurance) to 1 reduces the optimal top rate (the dashed gray line in the figure), but the quantitative impact is small. Reducing the extent of income risk (solid gray line) has a more noticeable impact, but the top optimal rate still exceeds 60 percent if at least 2/3 of the population is subject to income risk. Finally, making labor supply significantly more elastic (dotted gray line)[12] reduces the optimal rate further, but even then the optimal rate is 60 percent as long as all individuals in society face the risk of never making it into the top 1 percent of earners and, thus, the benefits of social insurance are large.

The results in this section convey two basic insights. First, the optimal top income tax rate depends crucially on the shape of the top income distribution, the earnings policy elasticity, the desire for social insurance, and the use of the extra tax revenue, represented in the simple model by the statistics $(a, \epsilon(z_h), \gamma, \tau_a(\bar{z}))$. Second, even though the welfare-maximizing rate is in general lower than the revenue-maximizing rate, these rates are quantitatively close as long as most people face income risk and the gap between top income earners and the rest is large and, thus, the scope for social insurance is sizeable.

[12] In a sensitivity analysis of our quantitative model in Section VIB, we consider a case with a high value for the Frisch elasticity of labor supply that results in a policy elasticity of $\epsilon(z_h) = 0.37$.

Finally, even in the static model these sufficient statistics are not invariant to the top marginal tax rate. In the dynamic model they are not constant over time, either. We will show that the dynamics of the wealth distribution matter considerably for the size of the policy elasticity $\epsilon(z_h)$ and, hence, for the location of the peak of the Laffer curve. Since wealth is a slow-moving object, it is crucial to consider the transitional dynamics induced by a tax policy reform. We will show that in the quantitative model the revenue-maximizing and welfare-maximizing rates differ greatly in the short run and in the long run.

II. The Quantitative Model

We now study a standard large-scale overlapping generations model in the spirit of Auerbach and Kotlikoff (1987), augmented by exogenous ex ante heterogeneity across households by education levels as well as by 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 ends of the distribution. We first set out the model using recursive language and define a stationary recursive competitive equilibrium. We then turn to a description of the potential policy reforms and the transition dynamics induced by them.

A. *Technology*

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

$$Y = \Omega K^{\epsilon} L^{1-\epsilon}, \tag{22}$$

where $\Omega \geq 0$ parameterizes the level of technology and the parameter $\epsilon \in [0,1]$ measures the elasticity of output with respect to capital. Capital depreciates at rate δ_k in every period. Given our assumptions of perfect competition in all markets and constant returns to scale in production, the number of operative firms and their size is indeterminate and, without loss of generality, we can assume the existence of a representative, competitively behaving firm producing according to the aggregate production function (22).

B. *Preferences and Endowments*

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

Households have preferences defined over stochastic streams of consumption and labor $\{c_j, n_j\}$ determined by the period utility function $U(c_j, n_j)$ in (1) and the time discount factor β. They maximize expected (with respect to idiosyncratic longevity risk and wage risk) lifetime utility and are ex ante heterogeneous with respect to the education they have acquired, a process we do not model endogenously. Let $s \in \{n, c\}$ denote the education level of a household, with $s = c$ denoting some college education and $s = n$ representing high school education (or less). The fraction of college-educated households is exogenously given by ϕ_s. In addition, prior to labor market entry, households draw a fixed effect[13] α from an education-specific distribution $\phi_s(\alpha)$. The wage a household faces in the labor market is given by

$$w \cdot e(j, s, \alpha, \eta), \tag{23}$$

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

Idiosyncratic wage risk determined by the process for η and mortality risk parameterized by the survival probabilities ψ_j cannot be explicitly insured because of market incompleteness, as in Bewley (1986), Huggett (1993), or Aiyagari (1994). However, households can self-insure against these risks by saving at a risk-free after-tax interest rate $r_n = r(1 - \tau_k)$. In addition to saving $a' - a$ the household spends her income—composed of earnings $z = we(j, s, \alpha, \eta)n$, capital income $r_n a$, and transfers $b_j(s, \alpha, \eta)$[14]—on consumption $(1 + \tau_c)c$, including consumption taxes, and on paying labor income taxes $T(z)$ as well as payroll taxes $T_{ss}(z)$. Implicit in these formulations is that the consumption tax and capital income tax are assumed to be linear, whereas the labor earnings tax is given by the potentially nonlinear function $T(\cdot)$.

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

$$v(j, s, \alpha, \eta, a) = \max_{c,n,a'} U(c, n) + \beta \psi_{j+1} \sum_{\eta'} \pi_s(\eta'|\eta) v(j+1, s, \alpha, \eta', a') \tag{24}$$

subject to

$$(1 + \tau_c)c + a' + T(z) + T_{ss}(z) = (1 + r_n)a + b_j(s, \eta) + z \tag{25}$$

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

[14] Transfers include Social Security for the retired and accidental bequests for all working households.

with

$$z = we(j,s,\alpha,\eta)n$$

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

C. *Government Policy*

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

We let Φ denote the cross-sectional distribution[15] of households, constant in a stationary equilibrium, and indicate aggregate quantities derived from individual decisions and Φ by capital letters. The budget constraint of the government in a stationary recursive competitive equilibrium with population growth then reads as

$$(26) \qquad r\tau_k(K+B) + \tau_c C + \int T\big(z(j,s,\alpha,\eta,a)\big)\, d\Phi = G + (r-g_n)B.$$

In addition, the PAYGO Social Security system is characterized by a payroll tax rate τ_{ss}, an earnings threshold $\bar{z}_{ss}$ below which households pay Social Security taxes, and benefits $p(s,\alpha,\eta)$ that depend on the last realization of the persistent wage shock η of working age,[16] education s and the fixed effect α (which determine expected wages over the life cycle). Thus $(\tau_{ss}, \bar{z}_{ss})$ completely pin down the payroll tax function T_{ss}. The specific form of the function $p(s,\alpha,\eta)$ is discussed in Section III.

The budget constraint of the Social Security system then reads as

$$(27) \qquad \int p(s,\alpha,\eta)\cdot \mathbf{1}_{\{j>j_r\}}\, d\Phi = \tau_{ss}\int \min\big\{\bar{z}_{ss}, z(j,s,\alpha,\eta,a)\big\}\, d\Phi.$$

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

$$(28) \qquad Tr = \frac{\int (1+r^n)(1-\psi_{j+1})a'(j,s,\alpha,\eta,a)\, d\Phi}{\int \mathbf{1}_{\{j\leq j_r\}}\, d\Phi},$$

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

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

so that transfers received by households are given as

$$b(j,s,\alpha,\eta) = \begin{cases} Tr & \text{if } j \leq j_r \\ p(s,\alpha,\eta) & \text{if } j > j_r. \end{cases} \tag{29}$$

A definition of a stationary recursive competitive equilibrium is given in online Appendix C.

D. *Transition Paths*

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

$$\begin{aligned} G + (1+r_t)B_t = {} & (1+g_n)B_{t+1} + r_t\tau_k(K_t + B_t) + \tau_c C_t \\ & + \int T_t\big(z_t(j,s,\alpha,\eta,a)\big)\, d\Phi_t \end{aligned} \tag{30}$$

and

$$(1+g_n)(K_{t+1} + B_{t+1}) = \int a_t'(j,s,\alpha,\eta,a)\, d\Phi_t. \tag{31}$$

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.[17]

III. Mapping the Model into Data

Conceptually, we proceed to map the initial stationary equilibrium of our model into US data in two steps. 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 US economy.

Most of the calibration is fairly standard for quantitative OLG models with idiosyncratic risk. However, given the purpose of the paper, there are two essential issues that require special attention. First, it is important that the model-generated cross-sectional earnings and wealth distribution is characterized by the same concentration at the top as in the data. We follow Castañeda, Díaz-Giménez, and Ríos-Rull (2003) and augment fairly standard stochastic wage processes derived

[17] 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).

from the PSID by Krueger and Ludwig (2013) with labor productivity states that occur with low probability, but which induce persistently large earnings when they occur. This allows the model to match the high earnings concentration and the even higher wealth concentration at the top of the distribution. In addition, the explicit life-cycle structure, including a fully articulated Social Security system, permits us to generate a distribution of earnings and wealth at the bottom and the middle of the distribution that matches the data quite well.

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

A. *Demographics*

We set the population growth rate to $g_n = 1.1$ percent; the long-run average value for the US data on survival probabilities from the Human Mortality Database (2013) for the United States in 2010 are used to determine the age-dependent survival probabilities $\{\psi_j\}$.

B. *Technology*

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

C. *Endowments and Preferences*

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

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

$$\ln e(j,s,\alpha,\eta) = \begin{cases} \alpha + \varepsilon_{j,s} + \eta & \text{if } \eta \in \{\eta_{s,1}, \dots, \eta_{s,6}\} \\ \eta & \text{if } \eta = \eta_{s,7}. \end{cases} \tag{32}$$

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, Kitao, and Krueger 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.[18]

We now specify the seven states of the Markov chain $\{\eta_{s,1}, \ldots, \eta_{s,7}\}$ and the transition matrices π_s. In addition, we need to determine the education-specific distribution of the fixed effect $\phi_s(\alpha)$ and the deterministic, education-specific age-wage profile $\{\varepsilon_{j,s}\}$. For the latter, we use the estimates by Krueger and Ludwig (2013) derived from PSID data. 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 and the variances and persistence parameters $\{\sigma^2_{\alpha,s}, \sigma^2_{\eta,s}, \rho_s\}$ of the AR(1) processes, together with the share of college-educated households ϕ_s. Table 2 summarizes our choices.

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

(33)
$$\pi_s = \begin{bmatrix} \pi_{11,s}(1-\pi_{16,s}) & \ldots & \pi_{13,s}(1-\pi_{16,s}) & \ldots & \pi_{15,s}(1-\pi_{16,s}) & \pi_{16,s} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & 0 \\ \pi_{51,s}(1-\pi_{56,s}) & \ldots & \pi_{53,s}(1-\pi_{16,s}) & \ldots & \pi_{55,s}(1-\pi_{56,s}) & \pi_{56,s} & 0 \\ 0 & \ldots & 1-\pi_{66,s}-\pi_{67,s} & \ldots & 0 & \pi_{66,s} & \pi_{67,s} \\ 0 & \ldots & 0 & \ldots & 0 & 1-\pi_{77,s} & \pi_{77,s} \end{bmatrix}.$$

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

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

Table 2—Labor Productivity Process

	ρ_s	$\sigma^2_{\eta,s}$	$\sigma^2_{\alpha,s}$	ϕ_s
$s = n$	0.9850	0.0298	0.1546	0.59
$s = c$	0.9850	0.0155	0.1138	0.41

Table 3—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

summarize, along with the empirical targets, in Table 3. Online Appendix E gives the values of the transition probabilities and states of the Markov chains.[19]

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

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

$$\tau_{Laffer} = \frac{1}{1 + a \cdot \epsilon(z_h)}, \tag{34}$$

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

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

The calibration for labor productivity that targets both the earnings and wealth distribution implies a value of $a = 1.79$ for the Pareto coefficient (in the initial steady state).[20] With this value for a, a policy elasticity of $\epsilon(z_h) = 0.21$ is needed to obtain a recommended rate of 73 percent to reach the peak of the top 1 percent curve.[21] In our full dynamic general equilibrium model, we cannot provide a closed-form solution for the policy elasticity anymore.[22] Hence, we calculate the policy elasticity numerically within our model.[23] By choice of $\gamma = 1.509$, our model delivers a policy elasticity of $\epsilon(z_h) = 0.21$. A similar value for γ has been estimated by Heathcote et al. (2014) in a life-cycle model using cross-sectional data on earnings and consumption from PSID and CEX data.

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

D. *Government Policies*

The two government policies we model explicitly are the tax system and the Social Security system.[24] We discuss both in turn now.

The Tax System.—We assume that the labor earnings tax function is characterized by the marginal tax rate function $T'(z)$ depicted in Figure 2. It is thus characterized by two tax rates τ_l, τ_h and two earnings thresholds $\bar{z}_l, \bar{z}_h$. As in the simple model of Section II, earnings below $\bar{z}_l$ are not taxed and earnings above $\bar{z}_h$ are taxed at the highest marginal rate τ_h. For earnings in the interval $[\bar{z}_l, \bar{z}_h]$, marginal taxes increase linearly from τ_l to τ_h. This tax code strikes a balance between approximating the current income tax code in the United States, being parameterized by few parameters and being continuously differentiable above the initial earnings threshold $\bar{z}_l$, which is crucial for our computational algorithm. Varying τ_h permits us to control the extent to which labor earnings at the top of the earnings distribution are taxed, and changing $\bar{z}_h$ controls at which 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

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

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

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

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

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

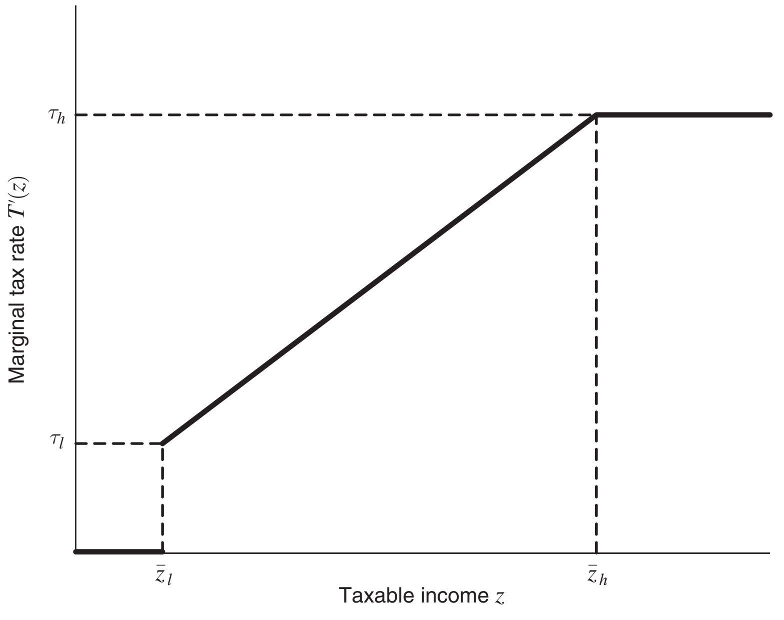

Figure 2. Marginal Labor Income Tax Function

tax system is more progressive than the original one, as is the case in the simple model of Section II.

For the initial equilibrium, we choose the highest marginal tax rate $\tau_h = 39.6$ percent, equal to the highest marginal income tax rate of the federal income tax code prior to the 2018 federal income tax reform.[25] That tax rate applies to labor earnings in excess of four times the average household *income*, or $\bar{z}_h = 4\bar{y}$. Households below 35 percent of median income do not pay any taxes;[26] $\bar{z}_l = 0.35\, y^{med}$ and we determine τ_l from budget balance in the initial stationary equilibrium, given the other government policies discussed below.[27] This requires $\tau_l = 11.2$ percent, which lies in between the 2 lowest marginal tax rates of the current US federal income tax code (10 percent and 12 percent).

The initial proportional capital income tax rate is set to $\tau_k = 28.3$ percent and the consumption tax rate to $\tau_c = 5$ percent. We choose exogenous government spending G such that it constitutes 17 percent of GDP; outstanding government debt B is set such that the debt-to-GDP ratio is 60 percent in the initial stationary equilibrium.

[25] This value for the highest marginal tax rate is also close to the value assumed by Diamond and Saez (2011). They use a top tax rate of 42.5 percent that includes a 2.9 percent Medicare tax; we instead treat Medicare as part of the Social Security system.

[26] 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 percent of median household income fairly independently of household composition.

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

These choices coincide with those in Krueger and Ludwig (2013), who argue that they reflect US policy prior to the Great Recession well.

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

$$p(s, \alpha, \eta) = f\big(\tilde{z}(s, \alpha, \eta = \eta_{65})\big). \tag{35}$$

We discuss the details of the benefit formula $f(\cdot)$ in online Appendix E.

E. *Calibration Summary*

Tables 4 and 5 summarize the choice of the remaining exogenously set and endogenously calibrated parameters. The exogenously chosen ones include policy parameters describing current US fiscal policy, as well as the capital share ϵ and the preference parameter χ. The choices for these parameters are standard relative to the literature, with the possible exception of the Frisch labor supply elasticity $\chi = 0.6$, which is larger than the microeconomic estimates for White prime-age males. However, it should be kept in mind that we are modeling household labor supply, including the labor supply of the secondary earner. Note that this choice implies, ceteris paribus, strong disincentive effects on labor supply from higher marginal tax rates at the top of the earnings distribution.

The set of parameters calibrated within the model include the technology parameters (δ_k, Ω), the preference parameters (β, γ, λ) and 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 equilibrium is consistent with a capital-output ratio of 2.9 and a share of time spent on market work equal to 33 percent of the total time endowment available to households. The technology parameters are then determined to reproduce a real pretax return on capital of 4 percent and a wage rate of 1, the latter being a normalization of Ω. Table 5 summarizes the values of these parameters.[28]

IV. Characteristics of the Benchmark Economy

Prior to turning to our tax experiments, we first discuss the aggregate and distributional properties of the initial stationary equilibrium. This is more important than

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

Table 4—Exogenously Chosen Parameters

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

Table 5—Endogenously Calibrated Parameters

Parameter	Value	Target/Data
Technology level Ω	0.921	$w = 1$
Depreciation rate δ_k	7.5%	$r = 4\%$
Initial marginal tax rate τ_l	11.2%	Budget balance
Time discount factor β	0.981	$K/Y = 2.9$
Disutility from labor λ	24	$\bar{n} = 33\%$
Coeff. of relative risk aversion γ	1.509	$\epsilon(z_m) = 0.21$

it is for most applications since a realistic earnings and wealth distribution, especially at the top of the distribution, is required to evaluate a policy reform that will entail potentially massive redistribution of the burden of taxation across different members of the population.

A. *Macroeconomic Aggregates*

In Table 6 we summarize the key macroeconomic aggregates in the initial stationary equilibrium. It shows that the main source of government tax revenues are labor earnings taxes.

B. *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 7 displays the model-implied earnings distribution and Table 8 contains the wealth distribution. When comparing the model-implied earnings and wealth

Table 6—Macroeconomic Variables

Parameter	Value	Parameter	Value
Capital	288%	Tax revenues	
Government debt	60%	– Consumption	2.9%
Consumption	58%	– Labor	11.9%
Investment	25%	– Capital income	3.9%
Government consumption	17%	Pension system	
Av. hours worked (in %)	33%	Contribution rate (in %)	12.5%

Note: All variables in percent of GDP if not indicated otherwise.

Table 7—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	5.6	10.9	17.3	66.2	10.9	18.9	22.6	0.648
US Data	–0.1	4.2	11.7	20.8	63.5	11.7	16.6	18.7	0.636

Table 8—Wealth Distribution in Benchmark Economy

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

quintiles to the corresponding data statistics,[29] we observe that the model fits the data very well, even at the top of the distribution. The same is true for the earnings and wealth Gini coefficients.

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, the 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.

V. 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

[29] As reported by Díaz-Giménez, Glover and Ríos-Rull (2011), based on the 2007 Survey of Consumer Finances.

steps. First, we display top-income Laffer curves, showing at what top marginal tax rate revenue from the top 1 percent of earners is maximized, and we 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 percent 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 percent stem primarily from enhanced insurance against not ascending to the very top of the earnings ladder and only secondarily from redistribution across ex ante heterogeneous households, and that these gains outweigh the macroeconomic costs (as measured by the decline in aggregate consumption) of the reform. In a last subsection, we show that these conclusions are robust to alternative preference specifications of households, but that they do, crucially, depend on a productivity and thus an earnings process that delivers the empirically observed earnings *and* wealth inequality in the data.

A. *The Thought Experiments*

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

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

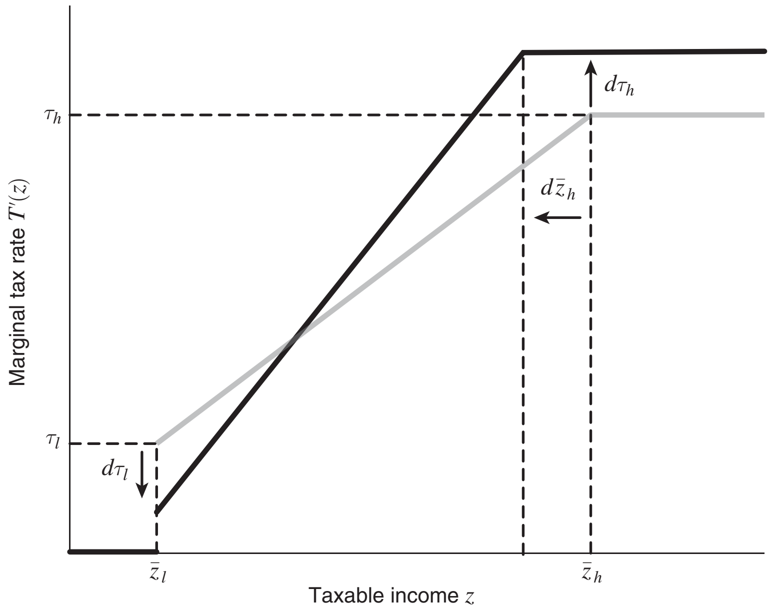

Figure 3. Laffer Curve of Labor Income Tax Receipts from Top 1 Percent

B. *The Top 1 Percent Laffer Curve*

In Figure 4 we plot (in percent deviation from the initial stationary equilibrium) labor income tax receipts from the top 1 percent earners against the top marginal labor income tax rate.[30] The three lines correspond to tax revenues in the first period of the transition (the short run), new steady-state tax revenues (the long run), and the present discounted value of tax receipts along the entire transition path (and the final steady state), where the discount rates used are the time-varying interest rates along the transition path.

From this figure we observe that the revenue-maximizing top marginal tax rate, independent of the time horizon used, is very high, in excess of 80 percent (see the solid black line). However, we also note that the time horizon does matter significantly: when maximizing tax revenue from top 1 percent earners in the short run, the revenue-maximizing rate is 80 percent and the extra revenue that can be generated is roughly 35 percent higher than in the benchmark economy. (See the dashed gray line in Figure 4). As we will show, households reduce their wealth holdings along the transition and become more *inelastic* when faced with higher top marginal tax rates. Consequently, the longer the time horizon, the higher the revenue-maximizing top rate, and the larger the extra revenues that can be generated by this rate. If one restricts attention solely to a steady-state analysis, then the peak of the top 1 percent

[30] Since, in the benchmark tax system, the top marginal tax rate does not apply exactly to the top 1 percent income earners (whereas in our tax experiments we ensure that it does), the Laffer curve does not intersect the 0 line at exactly 39.6 percent but, rather, at a slightly higher level. This is, of course, irrelevant to the question of where the peak of the Laffer curve (and the optimal rate) is located.

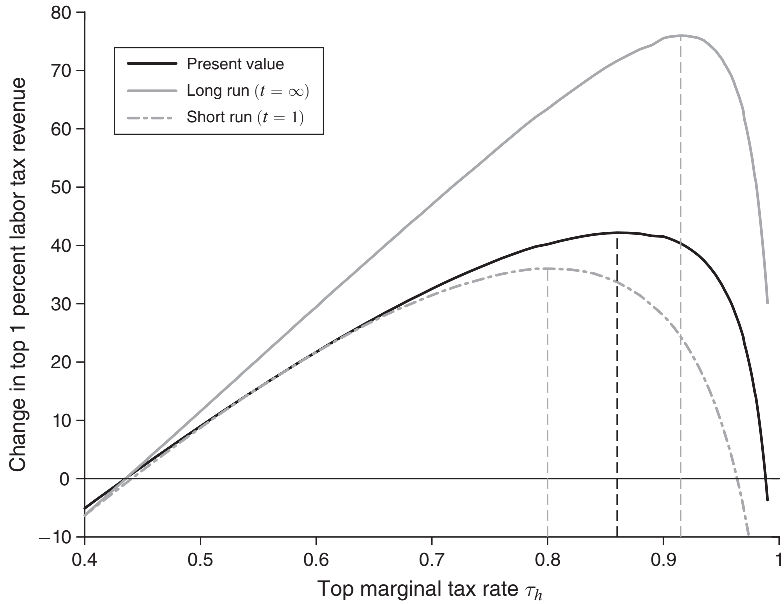

Figure 4. Laffer Curve of Labor Income Tax Receipts from Top 1 Percent

Laffer curve is attained at a tax rate of 92 percent, with 70 percent higher tax revenues from the highest income earners than in the initial stationary equilibrium. (See the solid gray line in Figure 4). The peak of the Laffer curve when maximizing the present discounted value of tax revenues, which is most informative for our welfare calculations, lies in the middle, between the short-run and long-run results (at a rate of 87 percent). Thus, we deduce two main points from Figure 4. First, revenue-maximizing rates are very high relative to the status quo. Second, and more importantly, the time horizon plays a crucial role for the quantitative results due to endogenous wealth accumulation, a finding that can only be uncovered through an explicit analysis of the transition path of a dynamic model with endogenous capital accumulation.

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

C. *Connection to Sufficient Statistics Approach*

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

$$\tau_{Laffer} = \frac{1}{1 + a \cdot \epsilon(z_h)}. \tag{36}$$

Table 9—Sufficient Statistic versus Numerical Simulation

		Augmented Formula			
		$t = 1$		$t = \infty$	
	Original	Initial	Final	Initial	Final
a_t	1.79	1.79	1.58	1.79	1.49
$\epsilon(z_{h,t})$	0.21	0.21	0.26	0.09	0.11
$\epsilon(\tau_a(\bar{z}_t))$		−0.78	−0.60	−0.77	−0.31
$\tau_a(\bar{z}_t)$		0.27	0.36	0.27	0.51
$\tau_{\text{Laffer}}^{\text{predicted}}$	0.73	0.85	0.80	1.00	0.92
$\tau_{\text{Laffer}}^{\text{simulated}}$		0.80		0.92	

The first column in Table 9 summarizes the ingredients into this formula based on simulated statistics from our quantitative model. In our calibration the Pareto parameter a, which summarizes the ratio between average top 1 percent earnings and the top 1 percent earnings threshold, takes a value of 1.79, which is right in the middle of the value of 1.5 assumed by Diamond and Saez (2011) and the value of 2 used by Saez (2001). The policy elasticity, as targeted in the calibration of the preference parameter γ, is 0.21, slightly lower than in Diamond and Saez. As a consequence, the peak of the Laffer curve, as predicted by the original formula, is precisely at 73 percent, as recommended by Diamond and Saez (2011). In other words, if our model is the true data-generating process, and if one were to base policy recommendations on (36), one would arrive at exactly their recommendation.[31]

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

The remaining columns of Table 9 demonstrate these points. They are based on the augmented formula given in Proposition 1,

$$\tau_{Laffer} = \frac{1 - (a-1)\cdot\tau_a(\bar{z})\cdot\epsilon(\tau_a(\bar{z}))}{1 + a\cdot\epsilon(z_h)}, \tag{37}$$

and show the various ingredients of the formula computed from model-generated data in the first period of the transition ($t = 1$) and the final steady state ($t = \infty$). The columns "initial" and "final" calculate these ingredients at the initial status quo

[31] As described in Section III and online Appendix E, this is, of course, how we chose γ in the first place.

and the final (i.e., revenue-maximizing) tax systems, respectively. The additional (relative to the original formula) term $(a-1) \cdot \tau_a(\bar{z}) \cdot \epsilon(\tau_a(\bar{z}))$ summarizes the effects that stem from an adjustment of the tax schedule below the top 1 percent earnings threshold—in our case, of changes in τ_l. Recall that $\tau_a(\bar{z})$ is the average tax rate of a household at the income threshold and $\epsilon(\tau_a(\bar{z}))$ is the elasticity of that rate with respect to $1-\tau_h$.

Focusing first on the short run (columns 2 and 3), we find that at the status quo tax system (with a top marginal rate of 39.6 percent), this elasticity equals -0.78. That is, as the top marginal tax rate τ_h increases, so does the average tax rate at the top 1 percent threshold.[32] This boosts tax revenue collected from this group and, hence, the Laffer curve peaks at a higher rate. Based on the sufficient statistics estimates from our model, we predict a Laffer peak at $\tau_h = 85$ percent; see the second column ("initial") of Table 9.

Comparing this prediction to the actual peak rate of 80 percent shows that the sufficient statistics formula only imprecisely predicts the peak of the Laffer curve. The third column ("final") of Table 9 shows why. There, we summarize the sufficient statistics when the tax system features the Laffer tax rate of $\tau_h = 80$ percent. The results show that these statistics are far from invariant to the very tax system under which they were calculated. With the higher top rate, the Pareto parameter a drops by 0.21 (i.e., the right tail of the earnings distribution becomes fatter), the policy elasticity rises to 0.26, and the average tax rate at the threshold increases by 9 percentage points and its elasticity declines in absolute value. As a consequence of these changes, the augmented formula generates an estimate for the Laffer tax rate of 80 percent when the sufficient statistics are calculated under the Laffer tax system, exactly equal to the true Laffer tax rate.

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

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

depend on the tax system in which they are evaluated, with the peak rate falling to 92 percent.

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

D. *Welfare-Maximizing Tax Rates*

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

Measuring Social Welfare.—The welfare measure we employ is constructed as follows. After solving for the equilibrium path of a specific tax reform, we calculate the amount of initial wealth transfers needed to make an individual indifferent between the status quo and the policy reform—ex post for the currently living and ex ante for future generations.[33] These transfers $\Psi_0(j,s,\alpha,\eta,a)$ satisfy, for currently alive individuals,

$$v_1\big(j,s,\alpha,\eta,a+\Psi_0(j,s,\alpha,\eta,a)\big) = v_0\big(j,s,\alpha,\eta,a\big) \tag{38}$$

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

$$Ev_t\big(j = 1,s,\alpha,\bar{\eta},\Psi_t\big) = Ev_0\big(j = 1,s,\alpha,\bar{\eta},0\big) \tag{39}$$

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

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

$$W = \int \frac{\Psi_0(j, s, \alpha, \eta, a)}{1 + r_0} \, d\Phi_0 + \mu_1 \sum_{t=1}^{\infty} \left(\frac{1 + g_n}{1 + r_0}\right)^t \Psi_t. \tag{40}$$

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

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

$$C \sum_{t=0}^{\infty} \left(\frac{1 + g_n}{1 + r_0}\right)^t = -W. \tag{41}$$

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

$$CEV = 100 * C / C_0. \tag{42}$$

This idea of calculating the welfare consequences of policy reforms follows closely that of Huang, İmrohoroğlu, and Sargent (1997) or Benabou (2002) and, more generally, the hypothetical lump-sum redistribution authority originally envisioned by Auerbach and Kotlikoff (1987).[34, 35]

Finally, we also compute and report a steady-state welfare measure that asks what uniform (over time and across states) percentage increase in consumption a household born into the old steady state, under the veil of ignorance, prior to the realization of the education level s and fixed effect α would need to receive to be indifferent to being born into the steady state associated with a new policy.[36]

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

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

[36] Conesa, Kitao, and Krueger (2009) employ the same long-run welfare measure in their study of optimal capital taxes.

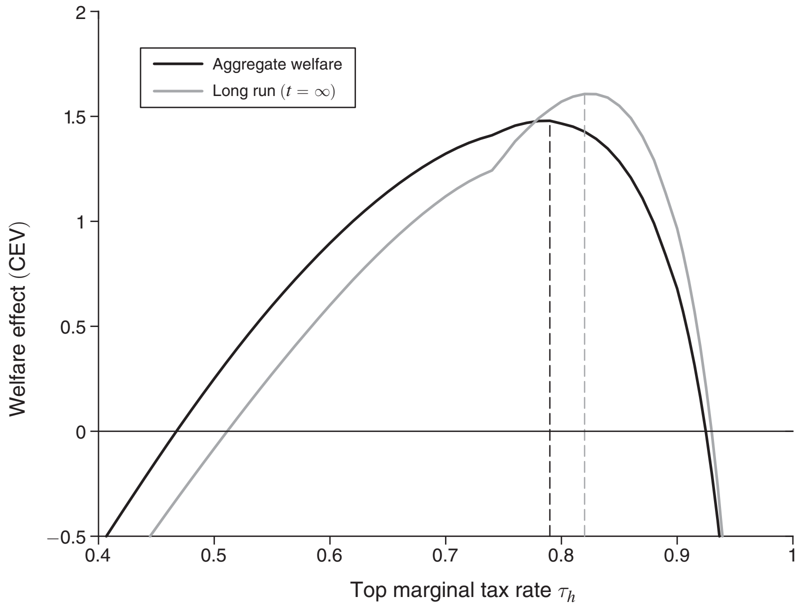

FIGURE 5. THREE AGGREGATE WELFARE MEASURES AS FUNCTIONS OF τ_h

Optimal Size of the Top Marginal Earnings Tax Rate.—In this section we document the optimal top marginal labor earnings tax rate. In Figure 5, we plot two welfare measures against the top marginal tax rate τ_h. The black line plots the aggregate welfare measure *CEV*, whereas the gray line instead displays steady-state welfare as described in the previous subsection.

As Figure 5 shows, the optimal top marginal tax rate is indeed very high, around 80 percent under both welfare measures. Welfare *CEV* including the top 1 percent households and including the transition effects is hump-shaped and maximized at $\tau_h = 79$ percent. Recall from Figure 4 that the top marginal tax rate that maximizes the present discounted value of tax revenues from the top 1 percent earners is 87 percent, higher than this welfare-optimal rate, but quantitatively close.[37] Focusing exclusively on welfare in the long run, the optimal top marginal rate is even larger, at $\tau_h = 82$ percent. Note that the welfare gains induced by these high marginal tax rates are very substantial, on the order of 1.5 percent of permanent consumption. In these thought experiments, as we vary τ_h we adjust the upper threshold $\bar{z}_h$ above which the highest marginal tax rate applies so that in the first period of the transition, the top 1 percent earners face this rate. The government intertemporal budget constraint is balanced by adjusting the entry marginal rate τ_l, holding fixed the lower bend point $\bar{z}_l$.[38]

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

[38] If the required τ_l is non-negative, all households with earnings below $\bar{z}_l$ pay zero taxes; if τ_l is negative, all households with earnings below $\bar{z}_l$ receive a subsidy of τ_l per dollar earned, akin to the Earned Income Tax Credit in the United States. This slight asymmetry about how income below $\bar{z}_l$ is treated induces a small kink in the welfare

E. *Understanding the Welfare Gains*

In order to understand the reported welfare gains from the optimal tax reform, we proceed in two steps. First, we display the transition paths of key macroeconomic variables that the tax reform induces, documenting the significant adverse consequences on output, *aggregate* consumption, and the capital stock in the economy. Second, we quantify the redistributive and insurance benefits of the reform, arguing that the latter are crucial for understanding our welfare results.

The Dynamics of Aggregates along the Transition.—In Figure 6 we plot the evolution of key macroeconomic aggregates along the transition from the old to the new stationary equilibrium. All variables are expressed in percent deviations from their initial steady-state values. Figure 7 displays the transition path of hours worked separately for the bottom 99 percent and the top 1 percent 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 pretax earnings and wealth inequality (as measured by the Gini coefficient) evolve over time.

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

In the medium run, the capital stock falls significantly, partially crowded out by higher public debt used to finance the tax transition but mainly driven by the decline in private saving of the high earners that are now subject to a significantly higher marginal (and thus average) labor earnings tax rate under the new tax system. Whereas in the short run most of the loss in output is absorbed by lower investment, in the long run, aggregate consumption declines strongly as well, by about 6 percent (see the right panel of Figure 6, after 40 periods of the transition).

The left panel of Figure 8 displays the evolution of tax revenue along the transition. Even though overall economic activity falls in response to the tax reform, government tax revenues 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

plot when τ_l turns from positive to negative. This is, of course, irrelevant for the determination of the optimal tax code, as the kink occurs far to the left of the optimal τ_h.

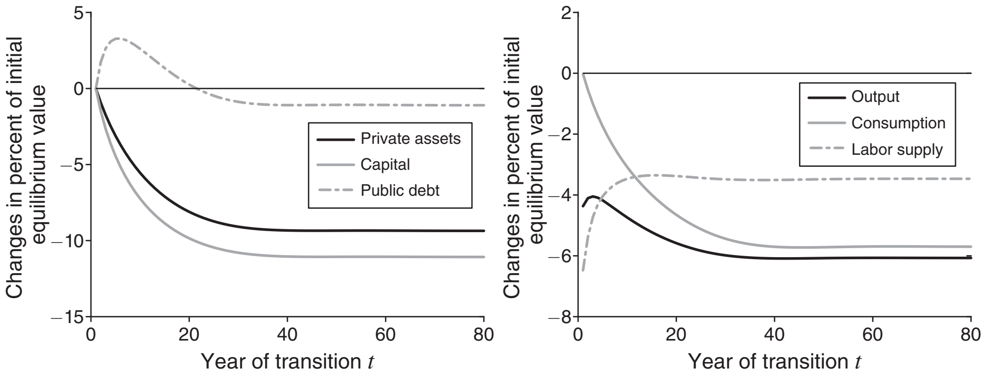

Figure 6. Aggregate Quantities along Transition

Figure 7. Hours and Prices along Transition

from taxing it. On the other hand, once the hours of the top 1 percent earners have partially recovered, labor income tax revenues increase on account of the significantly higher taxes these individuals pay. In the long run this group accounts for close to 80 percent of all revenue from the labor earnings tax. Revenues from capital income taxes also rise due to the higher return a lower capital-labor ratio implies, despite the decline in the capital income tax base.

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 pretax labor earnings falls significantly on impact, reflecting primarily the decline in hours worked and thus the earnings of the top 1 percent earners. As the hours of this group partially recover, so does earnings inequality, without reaching its pre-reform level. Wealth inequality, on the other hand, is monotonically and very substantially declining over time as the lower labor earnings of the households at the top of the distribution translate into lower wealth holdings in that group and, thus, a lesser net worth concentration in the population. In the long run, the wealth Gini is 10 percentage points lower than under the benchmark tax system, indicating that when a wide labor earnings distribution is the main culprit for high wealth inequality, tackling earnings inequality with high

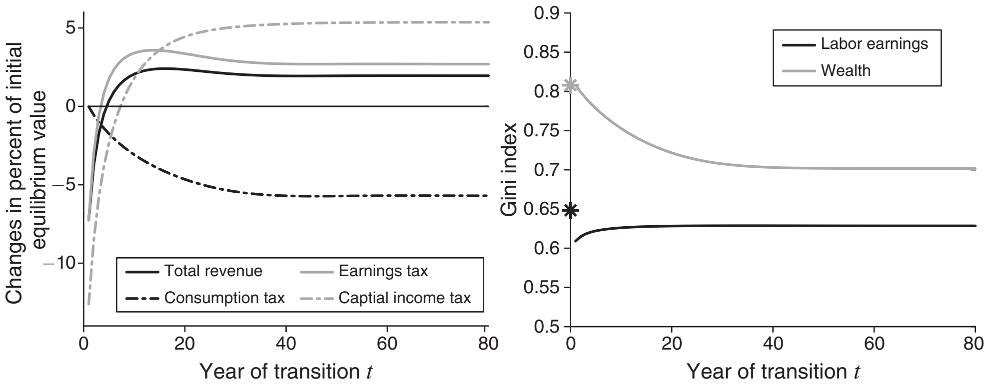

Figure 8. Tax Revenues and Inequality along Transition

marginal earnings taxation at the top is an effective tool for curbing wealth inequality. This is, of course, not an explicit policy goal of the government, but rather a side effect of its desire to provide social insurance and ex ante redistribution, as discussed in Section VE.

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

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

The welfare impact of the reform on the top 1 percent earners currently living 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

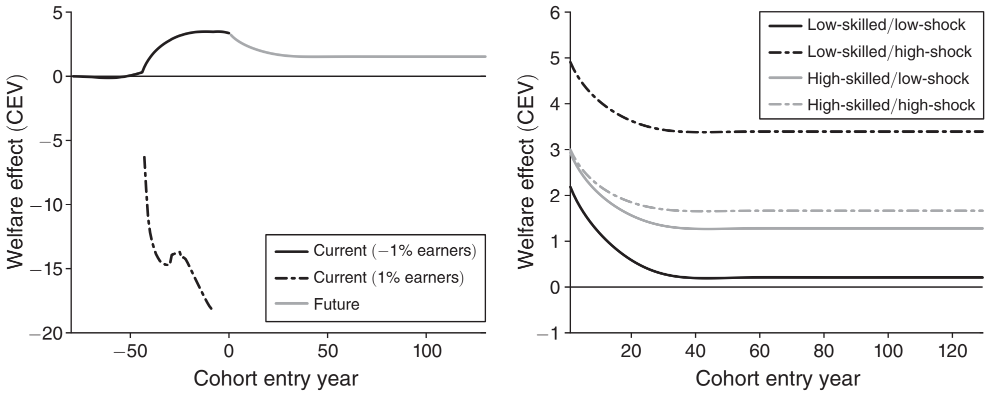

Figure 9. Welfare Effects of Policy Reform by Age Cohort

current non-top-earners the welfare gains are larger the younger they are, since younger workers spend a larger share of their working life under the new tax regime. Finally, the welfare impact of future generations is positive, in the order of 1.5–2.5 percent of lifetime consumption. (See the solid gray line in the left panel of Figure 9.) It falls along the transition as the economy consumes part of its capital stock.[39]

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

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

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

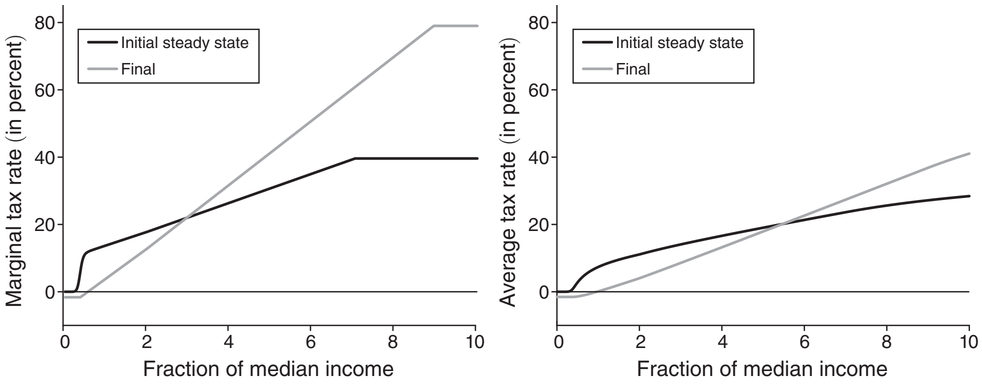

Figure 10. Marginal Tax Schedules, Average Tax Schedule: Benchmark and Optimum

This shows that households with up to about six times median earnings in the initial steady state face lower average taxes, whereas high earners face massively higher marginal and average taxes. In Figures 11 and 12, we display the differences in marginal and average tax rates between the two tax codes in conjunction with box plots to summarize the earnings distribution in the model in the initial (Figure 11) and final (Figure 12) steady state. As our model is populated by four ex ante heterogeneous types that differ in their education and earnings fixed effects, each panel includes four box plots associated with the earnings distributions of each of the four types. The box in the middle contains 50 percent of the probability mass, with the median earnings of the group represented by the vertical line in the middle of the box. The ends of the box plots give the positions of the 2.5 percentile and the 97.5 percentile of the earnings distribution.

We make three main observations. First, the overwhelming majority of households are located in parts of the earnings distribution that face lower average and marginal tax rates under the optimal tax system 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 *pretax* 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 percent and 200 percent of median income (see the right panels of Figures 11 and 12). This naturally makes the low-skilled, high fixed effect group and the high-skilled, low fixed effect group the largest beneficiaries of the reform; see the box plots of these two groups. The main difference between these two groups is that high-skilled (college) households have a nontrivial chance of rising to the very top of the earnings distribution (where they are hurt by the high marginal tax rates), whereas the low-skilled households face essentially zero chance of experiencing the same fate. We can see this by comparing the location of the 97.5 percentile of the earnings distribution for each of the two groups. This combination—middle class earnings, on average, and almost no chance of becoming very earnings rich—makes low-skilled households benefit disproportionately from the proposed tax reform.

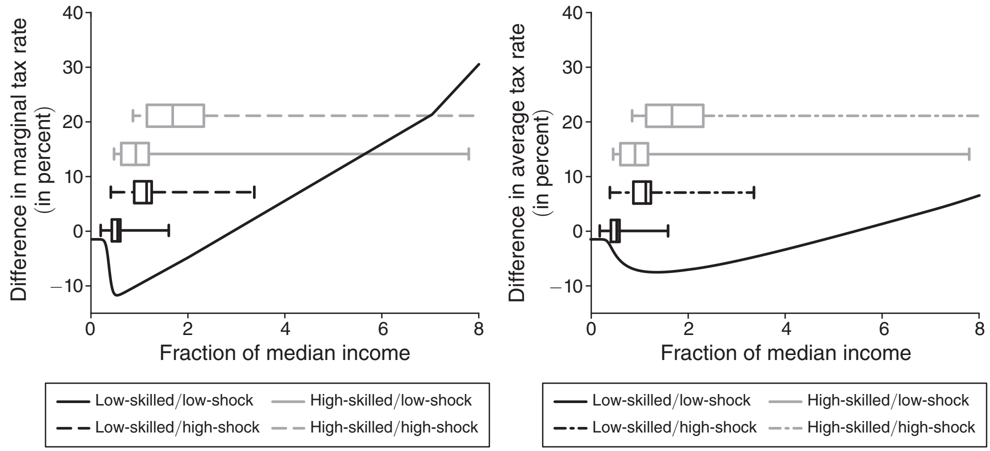

FIGURE 11. DIFFERENCE IN TAX SCHEDULES AND EARNINGS DISTRIBUTION (INITIAL STEADY STATE)

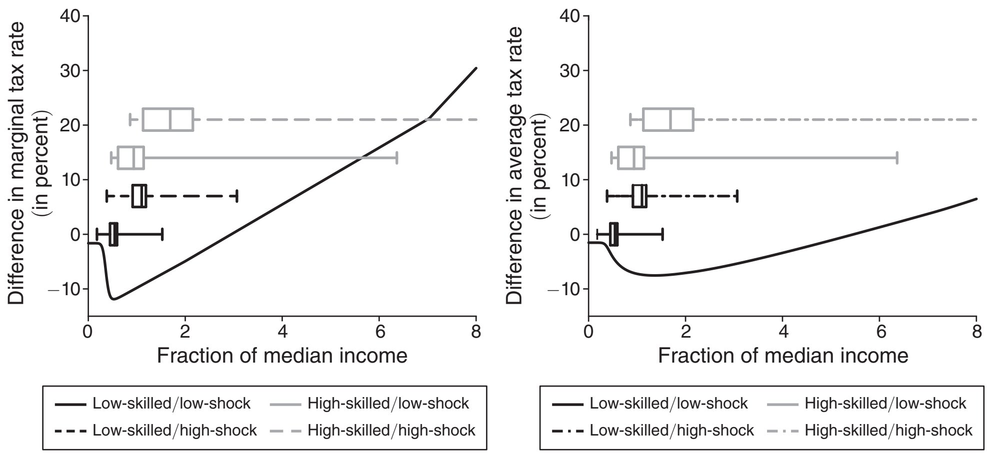

FIGURE 12. DIFFERENCE IN TAX SCHEDULES AND EARNINGS DISTRIBUTION (FINAL STEADY STATE)

The previous discussion does not clarify what the *common* sources of the welfare gains of each of these four groups are. To identify these sources, in Figure 13 we plot mean consumption and hours worked over the life cycle, not counting consumption and hours occurring when households have one of the two high labor productivity shocks (that is, roughly, excluding the hours and consumption of the top 1 percent). Figure 14 does the same for the variance of consumption of hours, and Figure 15 and Figure F1 in online Appendix F repeat the same for the entire population, that is, now *including* the top productivity states in the calculation of means and variances.

The key observation comes from comparing Figures 13 and 15. The average consumption of households outside the top 1 percent is uniformly larger over the life cycle under the new tax system relative to the old tax system (comparing steady states), despite the fact that aggregate consumption is 6 percent lower (as we saw in the right panel of Figure 6). As Figure 15 shows, the reduction of consumption is

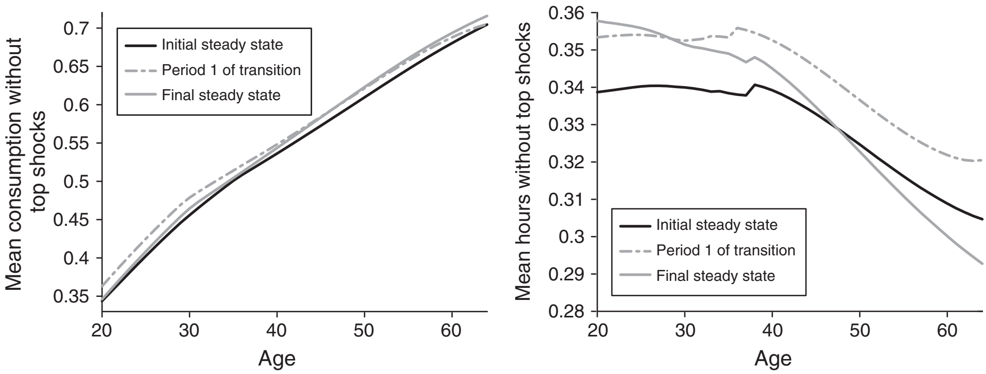

Figure 13. Average Consumption and Hours over the Life Cycles, without Top Two Shocks

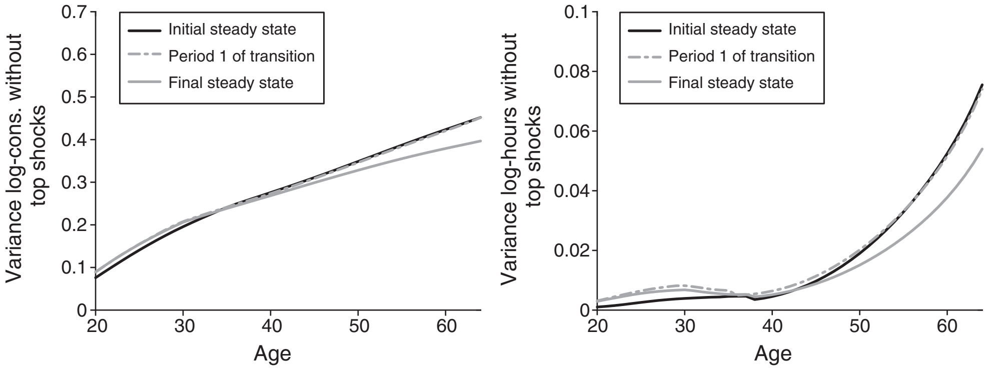

Figure 14. Variance of Consumption and Hours over the Life Cycles, without Top Two Shocks

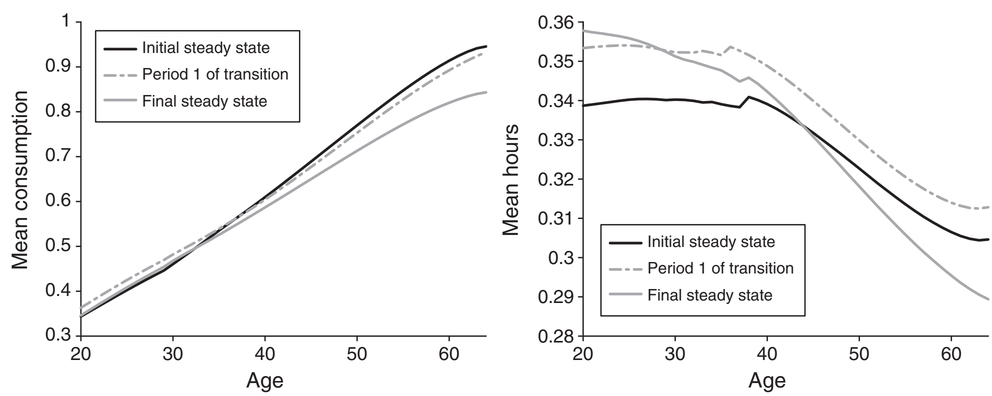

Figure 15. Average Consumption and Hours over the Life Cycles, Entire Population

heavily concentrated among older households at the top of the earnings distribution. In addition, hours worked remain roughly constant in the new steady state relative to the old steady state. Coupled with a sizeable reduction of lifetime consumption risk, approximated by the within-cohort consumption variance (see the left panel of Figure 14 or online Appendix Figure F1), the 1.5 percent steady-state welfare gains documented in Figure 5 emerge.

VI. Sensitivity Analysis

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

A. *The Productivity Process Generating Top Income Earners*

The key quantitative model ingredient to generate the very high earnings at the top of the income distribution and the even more concentrated wealth distribution is the presence of high and persistent labor productivity states. It is well known that this model element is sufficient to generate these distributions. However, we have argued here that the implied desire to provide social insurance against never becoming an earnings superstar or falling back to normal earnings provides a rationale for very high marginal tax rates on these top earners being optimal. We now show that in the absence of this model element, the implications for optimal tax rates at the top change dramatically.

The Model without Superstars.—Suppose first that households face a labor productivity process that does *not* contain the small chance of very high wage and, thus, earnings realizations.[40] In this version of the model the earnings, income, and wealth distributions do not display the degree of concentration observed in US data and thus, the model does not paint an accurate picture of the economic circumstances of the top 1 percent.

Figure 16 displays the top 1 percent Laffer curve (left panel) and welfare (right panel). As the figure shows, in the absence of the top two productivity shocks—and thus, in the absence of a realistic degree of earnings and wealth dispersion—the optimal top marginal labor earnings tax rate falls and is fairly close to the current US top rate. This happens for two reasons. First, the revenue-maximizing top marginal tax rate falls to 74 percent (see the solid gray line in the left panel), rather than above 87 percent, as in the benchmark economy, on account of a smaller income effect of the now less earnings-rich top 1 percent that makes labor supply more elastic to the top marginal tax rate. Most importantly, now the divergence between the revenue-maximizing (from the top 1 percent) top tax rate (above 70 percent) and welfare-maximizing top rate (below 40 percent) is much more significant, as the

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

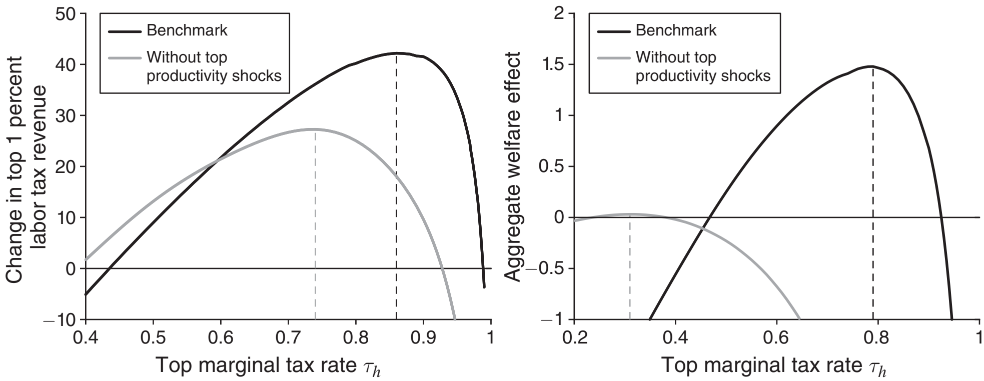

Figure 16. Laffer Curves and Welfare as Function of τ_h, Absent Top Productivity Shocks

solid gray line in Figure 16 shows. Since the largest productivity realizations are now much less severe, the large social insurance benefit of high tax progressivity vanishes. Thus, our main result—of very high marginal tax rates for top earners—depends crucially on a productivity process capable of producing earnings-rich and wealth-rich households, as in the data.

Persistence of High-Productivity States.—To what extent do our results depend on the fact that the large productivity shocks are persistent, but far from permanent (and, thus, a progressive tax system provides both insurance against the risks of never becoming highly productive and of becoming unproductive again after a spell of stardom)? To answer this question we model permanent superstars, but we remain consistent with our benchmark model, in which the probability of becoming very productive is essentially 0 before age 30. Specifically, we proceed as follows.

In the benchmark model, starting at age 30, households may receive the high productivity shock η_6 and, subsequently, the superstar shock η_7 according to the Markov transition matrix specified in the calibration section.[41] By contrast, we now assume that at age 30 a share of households randomly but *permanently* draw shocks η_6 and η_7. These shares are chosen such that the share of households with these productivities in the population are the same as in the benchmark model. In this way, we vary the persistence of the superstar states (by making it permanent) without changing the cross-sectional productivity distribution relative to the benchmark model.[42]

The main change relative to the benchmark model (and consistent with the previous section) is that with permanent top income shocks, the gap between the revenue-maximizing and the welfare-maximizing top rate is significantly larger. (See Figure 17.) Specifically, the welfare-maximizing tax rate is significantly smaller with permanently high productivity states. Effectively, being an earnings

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

[42] Note that whereas the earnings Gini remains close to its empirical counterpart, the wealth Gini falls from 0.81 to 0.74, and the wealth share of the top 1 percent decreases from 30 percent to 21 percent in the model with permanent superstars.

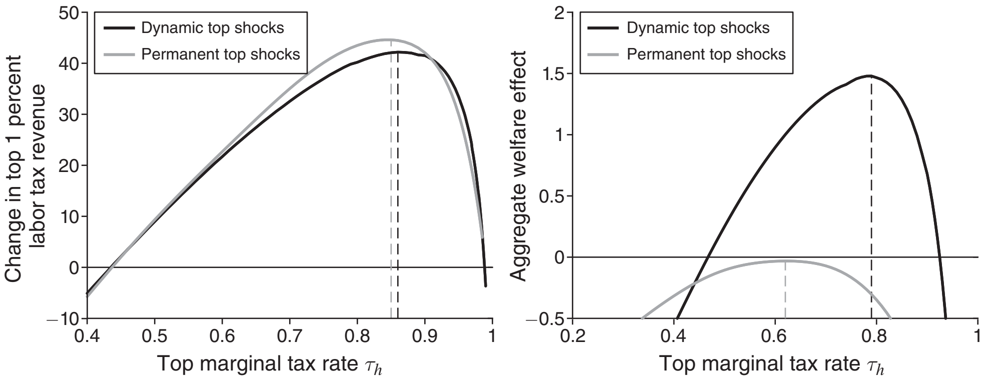

Figure 17. Laffer Curve and Aggregate Welfare, Persistent versus Permanent Highest Shocks

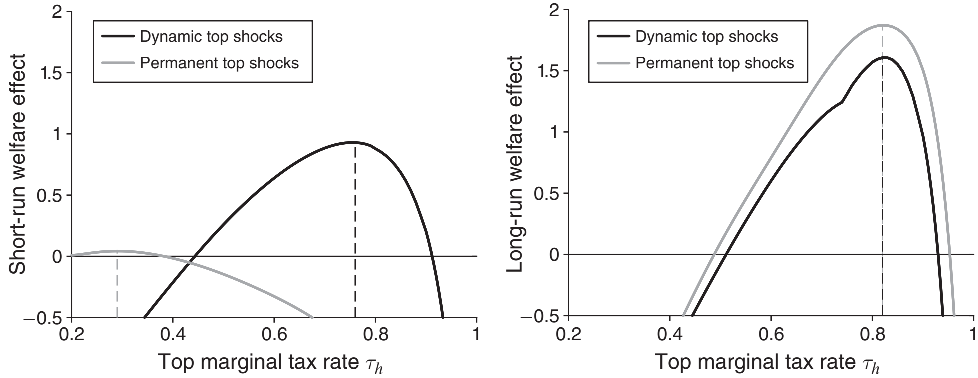

Figure 18. Short-Run and Long-Run Welfare Effects, Persistent versus Permanent Highest Shocks

superstar is now a permanent trait, and a high marginal tax rate on these individuals (with associated lower rates on everyone else) no longer provides useful insurance against reverting to the lower part of the earnings distribution.

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

Evidence on Top Income Earners.—The previous section has argued that mean reversion of earnings at the very top of the distribution is crucial for our optimal tax results. There is significant empirical support for this assumption. For example, Guvenen, Kaplan, and Song (2020) investigate data on the top 1 percent and the top 0.1 percent of wage earners from the Social Security Administration. They estimate the likelihood that a top earner remains in the same earnings bracket in the following year as well as over a five-year horizon. Their data display a significant extent of transitions in and out of the top earnings brackets. They report that in the 2000s, an individual in the top 0.1 percent earnings bracket only had about a 57 percent chance of staying there in the next year (40 percent over the next 5 years). An individual in the next 0.9 percent only stays within this group of earners with a probability of 65 percent (46 percent over a five-year horizon.) In our calibration, the probability of remaining in the very high productivity state for another year (71 percent over one year,18 percent over five years) strikes a balance between these short-run and long-run estimates.[43]

B. *Labor Supply Elasticity*

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

In Table 10 and in Figure F2 in online Appendix F, we document how our optimal tax and welfare results depend on the Frisch labor supply elasticity. The key finding is that, although the positive and normative results change in the expected direction (a larger elasticity reduces the size of the top marginal tax rate and the associated welfare gains from the policy reform), the differences are, quantitatively, fairly small. Even with a household-level Frisch labor supply elasticity of 1.5, arguably at the upper bound of empirical estimates, the optimal top marginal tax rate exceeds 70 percent. (See the second column of Table 10).

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

The left panel of Figure 19 plots the top-earner Laffer curve (the present discounted-value version) for risk aversion of $\gamma = 1$ and $\gamma = 1.509$, whereas the

[43] Auten, Gee, and Turner (2013) use tax return data and administrative records in the IRS Compliance Data Warehouse to document that for individuals who were in the top 1 percent group of taxpayers in 2005, 65 percent were still top 1 percent taxpayers in 2006 and only 27 percent were in 2010, confirming the results of Guvenen, Kaplan, and Song (2020).

Table 10—Sensitivity with Respect to Frisch Labor Supply Elasticity

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

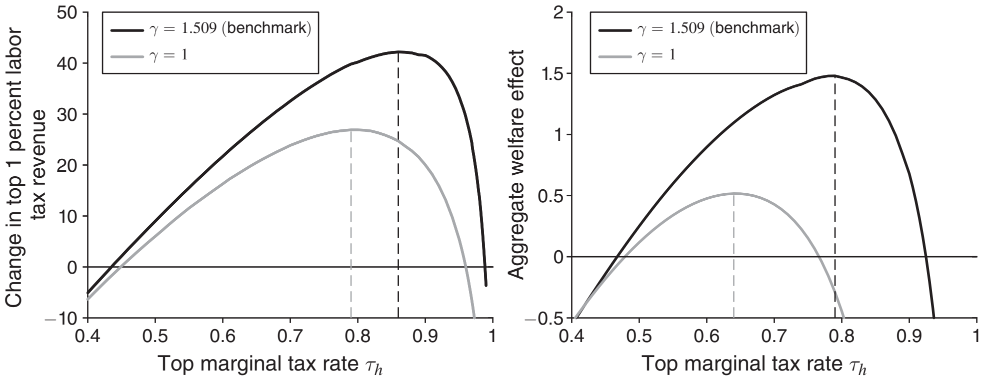

Figure 19. Laffer Curve and Aggregate Welfare, $\gamma = 1.509$ versus $\gamma = 1$

right panel does the same for aggregate welfare. We observe that the magnitude of the income effect is quantitatively important for our findings, but that the key result (top marginal tax rate significantly above current levels) remains unaffected. As Figure 19 (gray lines) shows, with log-utility the *revenue-maximizing* top rate is 79 percent and the welfare-maximizing rate is 64 percent.

Through the lens of the sufficient statistics approach, with a value of $\gamma = 1$ the short-run policy elasticity changes to $\epsilon(z_{m,1}) = 0.30$ and the long-run elasticity changes to $\epsilon(z_{m,\infty}) = 0.23$. Both values are significantly higher than the original ones shown in Table 9. The Pareto parameter rises to $a = 1.89$. Thus, a smaller income effect increases the elasticity of aggregate top earnings with respect to the top marginal net-of-tax rate. As a consequence, the revenue-maximizing tax rate falls from 0.87 to 0.79, as the left panel of Figure 19 shows. Second, the divergence between revenue maximization and welfare maximization again becomes more important as lower risk aversion shrinks the insurance benefits of highly progressive labor income taxes. Thus, the socially optimal top rate is even lower. Yet, it remains at a sizable 64 percent (see the gray line in the right panel of Figure 19), substantially higher than the current values in the United States. We think of the parameter configuration with log-utility as delivering a plausible lower bound for what the top marginal tax rate should be, since logarithmic utility implies a risk aversion at the low end of commonly used values and leads to a high elasticity of earnings with respect to taxes at the upper bound of empirical estimates.[44]

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

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

VII. Conclusion

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

The crucial model ingredient that generates realistic earnings and wealth inequality is a policy-invariant labor productivity process where individuals with small probability receive very high realizations, and these realizations are mean-reverting but persistent. Given the centrality of this assumption for our result, important next steps for 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 described well by our model, whereas high earnings professionals for whom long-term human capital investment decisions are crucial entry tickets into the top 1 percent 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) or Cagetti and De Nardi (2006) or the human capital model analyzed in Badel, Huggett, and Luo (2020).

REFERENCES

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *Quarterly Journal of Economics* 109 (3): 659–84.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. 2013. "The Top 1 Percent in International and Historical Perspective." *Journal of Economic Perspectives* 27 (3): 3–20.

Auerbach, J. Alan, and Laurence J. Kotlikoff. 1987. *Dynamic Fiscal Policy.* Cambridge, UK: Cambridge University Press.

Auten, Gerald, Geofrey Gee, and Nicholas Turner. 2013. "Income Inequality, Mobility, and Turnover at the Top in the U.S., 1987–2010." *American Economic Review* 103 (3): 168–72.

Badel, Alejandro, and Mark Huggett. 2017. "The Sufficient Statistic Approach: Predicting the Top of the Laffer Curve." *Journal of Monetary Economics* 87: 1–12.

Badel, Alejandro, Mark Huggett, and Wenlan Luo. 2020. "Taxing Top Earners: A Human Capital Perspective." *Economic Journal* 130 (629): 1200–25.

Bakış, Ozan, Barış Kaymak, and Markus Poschke. 2015. "Transitional Dynamics and the Optimal Progressivity of Income Redistribution." *Review of Economic Dynamics* 18 (3): 679–93.

Bassetto, Marco. 2014. "Optimal Fiscal Policy with Heterogeneous Agents." *Quantitative Economics* 5 (3): 675–704.

Benabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency." *Econometrica* 70 (2): 481–517.

Bewley, Truman F. 1986. "Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers." In *Contributions to Mathematical Economics in Honor of Gerard Debreu*, edited by Werner Hildenbrand and Andreu Mas-Colell, 79–102. Amsterdam: North-Holland.

Brüggemann, Bettina. 2021. "Higher Taxes at the Top: The Role of Entrepreneurs." *American Economic Journal: Macroeconomics* 13 (3): 1–36.

Brüggemann, Bettina, and Jinhyuk Yoo. 2015. "Aggregate and Distributional Effects of Increasing Taxes on Top Income Earners." Unpublished.

Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy* 114 (5): 835–70.

Castañeda, Ana, Javier Díaz-Giménez, and José-Victor Ríos-Rull. 2003. "Accounting for the U.S. Earnings and Wealth Inequality." *Journal of Political Economy* 111 (4): 818–57.

Chamley, Christophe. 1986. "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives." *Econometrica* 54 (3): 607–22.

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger. 2009. "Taxing Capital? Not a Bad Idea after All!" *American Economic Review* 99 (1): 25–48.

Conesa, Juan Carlos, and Dirk Krueger. 2006. "On the Optimal Progressivity of the Income Tax Code." *Journal of Monetary Economics* 53 (7): 1425–50.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88 (1): 83–95.

Diamond, Peter, and Emmanuel Saez. 2011. "The Case for a Progressive Tax: From Basic Research to Policy Recommendations." *Journal of Economic Perspectives* 25 (4): 165–90.

Díaz-Giménez, Javier, Andrew Glover, and José-Victor Ríos-Rull. 2011. "Facts on the Distributions of Earnings, Income, and Wealth in the United States: 2007 Update." *Federal Reserve Bank of Minneapolis Quarterly Review* 34 (1): 1–31.

Domeij, David, and Jonathan Heathcote. 2004. "On the Distributional Effects of Reducing Capital Taxes." *International Economic Review* 45 (2): 523–54.

Erosa, Andrés, and Martin Gervais. 2002. "Optimal Taxation in Life-Cycle Economies." *Journal of Economic Theory* 105 (2): 338–69.

Fehr, Hans, and Fabian Kindermann. 2015. "Taxing Capital along the Transition—Not a Bad Idea after All?" *Journal of Economic Dynamics and Control* 51: 64–77.

Greenwood, Jeremy, Zvi Hercovitz, and Gregory W. Huffman. 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review* 78 (3): 402–17.

Guner, Nezih, Martin Lopez-Daneri, and Gustavo Ventura. 2016. "Heterogeneity and Government Revenues: Higher Taxes at the Top?" *Journal of Monetary Economics* 80: 69–85.

Guvenen, Fatih, Greg Kaplan, and Jae Song. 2020. "The Glass Ceiling and the Paper Floor: Gender Differences among Top Earners, 1981–2012." *NBER Macroeconomics Annual* 35: 309–73.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2014. "Consumption and Labor Supply with Partial Insurance: An Analytical Framework." *American Economic Review* 104 (7): 2075–2126.

Hendren, Nathaniel. 2016. "The Policy Elasticity." *Tax Policy and the Economy* 30 (1): 51–89.

Holter, Hans A., Dirk Krueger, and Serhiy Stepanchuk. 2019. "How Does Tax Progressivity and Household Heterogeneity Affect Laffer Curves?" *Quantitative Economics* 10 (4): 1317–56.

Hsu, Minchung and C.C. Yang. 2013. "Optimal Linear and Two-Bracket Income Taxes with Idiosyncratic Earnings Risk." *Journal of Public Economics* 105: 58–71.

Huang, He, Selahattin İmrohoroğlu, and Thomas Sargent. 1997. "Two Computations to Fund Social Security." *Macroeconomic Dynamics* 1 (1): 7–44.

Hubmer, Joachim, Per Krusell, and Anthony A. Smith, Jr. 2020. "Sources of U.S. Wealth Inequality: Past, Present, and Future." *NBER Macroeconomics Annual* 35: 391–455.

Huggett, Mark. 1993. "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies." *Journal of Economic Dynamics and Control* 17 (5–6): 953–69.

Jacobs, Bas, and Dirk Schindler. 2012. "On the Desirability of Taxing Capital Income in Optimal Social Insurance." *Journal of Public Economics* 96 (9–10): 853–68.

Jones, Charles I., and Jihee Kim. 2018. "A Schumpeterian Model of Top Income Inequality." *Journal of Political Economy* 126 (5): 1785–1826.

Judd, Kenneth L. 1985. "Redistributive Taxation in a Simple Perfect Foresight Model." *Journal of Public Economics* 28 (1): 59–83.

Keane, Michael P. 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature* 49 (4): 961–1075.

Kindermann, Fabian, and Dirk Krueger. 2022. "Replication Data for: High Marginal Tax Rates on the Top 1 Percent?: Lessons from a Life Cycle Model with Idiosyncratic Income Risk." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E126741V1.

Krueger, Dirk, and Alexander Ludwig. 2013. "Optimal Progressive Labor Income Taxation and Education Subsidies When Education Decisions and Intergenerational Transfers Are Endogenous." *American Economic Review* 103 (3): 496–501.

Krusell, Per, and Anthony A. Smith, Jr. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106 (5): 867–96.

MaCurdy, Thomas E. 1981. "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy* 89 (6): 1059–85.

Mirrlees, J.A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.

Piketty, Thomas, and Arthur Goldhammer. 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118 (1): 1–41.

Quadrini, Vincenzo. 1997. "Entrepreneurship, Saving and Social Mobility." *Review of Economic Dynamics* 3 (1): 1–40.

Reich, Robert B. 2010. *Aftershock: The Next Economy and America's Future*. New York: Random House.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68 (1): 205–29.

Stantcheva, Stefanie. 2020. "Dynamic Taxation." *Annual Review of Economics* 12: 801–31.

University of California, Berkeley, and Max Planck Institute for Demographic Research. *Human Mortality Database*. Accessed August 10, 2013. https://www.mortality.org/.

Vogelgesang, Ulrike. 2000. "Optimal Capital Income Taxation and Redistribution." *FinanzArchiv* 57 (4): 412–34.

This article has been cited by:

1. Jason DeBacker, Vasia Panousi, Shanthi Ramnath. 2023. A Risky Venture: Income Dynamics among Pass-Through Business Owners. *American Economic Journal: Macroeconomics* **15**:1, 444-474. [Abstract] [View PDF article] [PDF with links]
2. Luisa Fuster. 2022. Macroeconomic and distributive effects of increasing taxes in Spain. *SERIEs* **13**:4, 613-648. [Crossref]
3. William B. Peterman, Erick Sager. 2022. Optimal Public Debt with Life Cycle Motives. *American Economic Journal: Macroeconomics* **14**:4, 404-437. [Abstract] [View PDF article] [PDF with links]