Taxing Top Earners: A Human Capital Perspective*

Alejandro Badel
Georgetown University
ale.badel@gmail.com

Mark Huggett
Georgetown University
mh5@georgetown.edu

Wenlan Luo
Tsinghua University
luowenlan@gmail.com

this draft: August 15, 2019

Abstract
An established view is that the revenue maximizing top tax rate for the US is approximately 73 percent. In contrast, the revenue maximizing top tax rate is approximately 49 percent in our quantitative human capital model. The key reason for the lower top tax rate is the presence of two new forces not captured by the model underlying the established view. These new forces are strengthened by the endogenous response of top earners' human capital to a change in the top tax rate.

Keywords: Human Capital, Marginal Tax Rates, Top Earners, Laffer Curve

JEL Classification: D52, E21, H2, J24

*Corresponding Author: Mark Huggett, Economics Department, Georgetown University, Washington, DC 20057, (email: mh5@georgetown.edu). Acknowledgements: This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Grant SES 130025, supported by the National Science Foundation, to access the Stampede cluster at the Texas Advanced Computing Center. This work also used the cluster at the Federal Reserve Bank of Kansas City and Georgetown University. We thank the editor and three referees for valuable comments.

1 Introduction

How should the tax rate on top earners be set? An established view is described by Diamond and Saez (2011) and Piketty and Saez (2013), among others, and is incorporated into the Mirrlees Review - an important document providing tax policy advice.[1]

The established view is based on first determining the revenue maximizing top tax rate via a widely-used formula. The coefficient a in the formula is $a = \bar{y}/(\bar{y} - \underline{y})$, where $\underline{y}$ is an earnings threshold at which the top tax rate applies and $\bar{y}$ is average earnings above this threshold. The parameter ϵ is an earnings elasticity, corresponding to all earners beyond the threshold, with respect to a change in one minus the top tax rate. Diamond and Saez (2011) suggest that $(a, \epsilon) = (1.5, 0.25)$ approximate US values and therefore that the revenue maximizing rate τ^* in the US is approximately 73 percent. They then argue that a revenue maximizing top tax rate will approximate a welfare maximizing top tax rate under certain conditions.

$$\tau^* = \frac{1}{1 + a\ \epsilon} = \frac{1}{1 + 1.5 \times 0.25} \doteq 0.73$$

We analyze the revenue maximizing top tax rate using a dynamic human capital model in place of the static Mirrlees model that underlies the established view. In the human capital model, agents with high learning ability are disproportionately top earners and become top earners largely later in life. These high learning ability agents put more time into skill accumulation than other agents over the lifetime and have strikingly large earnings growth rates over the working lifetime.

We analyze a tax reform that increases the tax rate on the top 1 percent, holding all other tax rates and government spending unchanged. Any extra revenue collected in the new steady state is used to fund a lump-sum transfer to all agents. The human capital model is calibrated to properties of the US age-earnings distribution and the structure of US marginal tax rates. The steady state model Laffer curve peaks at a top tax rate equal to 49 percent.

Why does the human capital model have a revenue maximizing top tax rate of 49 percent when the established view suggests that 73 percent is revenue maximizing? Badel and Huggett (2017, Theorem 1) derive a generalization of the widely-used tax rate formula. The Badel-Huggett formula has three elasticities and applies very widely to static models and to steady states of dynamic models whereas the widely-used formula applies only to some static models.

$$\tau^* = \frac{1 - a_2\epsilon_2 - a_3\epsilon_3}{1 + a_1\epsilon_1}$$

[1]See chapter 2 of the Mirrlees Review by Brewer, Saez and Shephard (2010).

In general there are three forces ($a_1\epsilon_1, a_2\epsilon_2$ and $a_3\epsilon_3$) that determine the top of the Laffer curve. One old force is governed by the elasticity ϵ_1 of earnings (income) for all earners that are above the threshold for the top tax rate to a change in $(1-\tau)$ (i.e. one minus the top tax rate). One new force is governed by ϵ_2 which is the percentage change in all the taxes paid by agents who are below the threshold to a percentage change in $(1-\tau)$. The other new force is governed by ϵ_3 which is the percentage change in all the other taxes paid by agents who are above the threshold to a percentage change in $(1-\tau)$. In the US, consumption and sources of capital income with preferential rates are taxed separately from ordinary income and are some of these other taxes paid by top earners.[2] When the two new elasticities are positive, then an increase in τ results in a decrease in tax revenue from agents below the threshold and a decrease in other taxes paid by top earners. Thus, total revenue would be maximized when the taxes collected from top earners, based on the tax base that τ applies to, are still increasing.

$$(*)\quad \tau^* = \frac{1 - a_2\epsilon_2 - a_3\epsilon_3}{1 + a_1\epsilon_1} = \frac{1 - 4.508 \times .040 - .120 \times .739}{1 + 1.7 \times .317} \doteq 0.49$$

We calculate all the terms entering the formula in the human capital model and apply the tax rate formula to determine the importance of the three forces. Equation $(*)$ indicates that the elasticities underlying the new forces are positive in the human capital model and depress the revenue maximizing top tax rate. The two new forces are quantitatively quite important as the counterfactual case of $\epsilon_2 = \epsilon_3 = 0$ would imply a 65 percent (i.e. $\tau^* = \frac{1-0-0}{1+1.7\times.317} = .65$) revenue maximizing top rate.[3]

Endogenous skill accumulation is quantitatively important for the model's revenue maximizing top tax rate. To determine this, we consider an observationally-equivalent, exogenous-skill model where skills evolve in the same way as in the benchmark human capital model independent of the value of the top tax rate. The revenue maximizing top tax rate is 59 percent in the exogenous-skill model and is larger because the skills of top earners fall when the top tax rate increases in the human capital model but are unaffected in the exogenous-skill model. This then implies that all three elasticities are smaller in the exogenous-skill model.

What is the mechanism by which the skills of top earners fall? An increase in the top tax rate decreases the marginal benefits of skill investment that are received later in life without changing the marginal cost of investment earlier in life. This leads to a fall in skill investment and a fall in skills later in life. For this logic to hold, top earners must have upward sloping earnings profiles. The model produces strikingly large earnings growth rates for top lifetime

[2]The federal tax rate on long-term capital gains and qualified dividends was 15 percent in 2010 for those with high incomes whereas the top federal tax rate on ordinary income was 35 percent.

[3]Section 3.3 explains why we focus on a value of $a_1 = 1.7$, rather than $a_1 = 1.5$, and what US data determines this value.

earners that are roughly similar to those documented for top US lifetime earners.

In theory, all three elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ can be measured in dynamic models by the size of the shift in the log of the balanced-growth path of aggregate income and tax revenue measures as a ratio to a permanent change in $\log(1-\tau)$ applying to top earners. We construct these three aggregate US time series and posit a reduced-form regression equation that relates these measures to an empirical proxy for a permanent change in $\log(1-\tau)$. We present regression evidence showing that the point estimates for $(\epsilon_1, \epsilon_2, \epsilon_3)$ are all positive, consistent with the implications of the human capital model, although the standard errors on these point estimates are large.

The paper is organized as follows. Section 2 presents the model framework. Section 3 documents properties of the US age-earnings distribution and US marginal tax rates. Section 4 calibrates the model and analyzes model validity. Section 5 presents the model Laffer curve, uses the formula to decompose the determinants of the revenue-maximizing rate and presents regression evidence for the values of the two new elasticities. Section 6 concludes.

Related Literature

The widely-used revenue maximizing top tax rate formula is described in Saez (2001). It was developed within a specific static model - the Mirrlees model. Brewer et al. (2010), Diamond and Saez (2011), Piketty and Saez (2013) and others discuss and apply this formula and related optimal tax formulae. Badel and Huggett (2017) derive a generalization of this formula that applies very widely to static and dynamic models and to any component of income.

Badel and Huggett (2017), henceforth BH, "bench test" a deterministic version of the model analyzed in this paper to determine whether their formula accurately predicts the top of the Laffer curve.[4] A model of top earners, to be used to provide quantitative insight into taxing top earners, needs to (i) address the role of earnings risk in becoming a top earner, (ii) be consistent with evidence on the low empirical value of the Frisch labor elasticity and (iii) be consistent with evidence on the income elasticity of top earners with respect to the net-of-tax rate. The model in BH does none of these, whereas the model in this paper does all of these.

The human capital model we employ adds valued leisure to the model developed by Huggett, Ventura and Yaron (2011). Heckman, Lochner and Taber (1998a), Erosa and Koreshkova (2007), Guvenen, Kuruscu and Ozkan (2014), Krueger and Ludwig (2016) and Heathcote, Storesletten and Violante (2017) analyze income tax progression using human capital models

[4] The formula holds exactly when the inputs are calculated at the top but not when they are calculated away from the top of the Laffer curve. The previous literature (e.g. Diamond and Saez (2011)) uses such formulae to predict the revenue maximizing top tax rate when inputs are calculated away from the top. Thus, to credibly use such formulae in this way, bench testing specific models would seem to be important.

but do not focus on tax reforms directed at the extreme upper tail. Blandin (2016) analyzes the elimination of the cap on social security earnings taxation in the US within a human capital model and finds that this reduces skill accumulation. The economic mechanism underlying his results is the same as the mechanism that we highlight.

Guner, Lopez-Daneri and Ventura (2016) and Kindermann and Krueger (2017) analyze tax reforms that are directed at the upper tail in models where an agent's productivity or skills are unaffected by a tax reform. The revenue maximizing top tax rate is aproximately 35 percent on the top 1 percent in Guner et al. (2016) but 77 percent in Kindermann and Krueger (2017).[5] Guner et al. (2016) model skills as having a permanent fixed effect component and a persistent autoregressive component - a specification that is often estimated in the earnings and wage rate literature. In contrast, Kindermann and Krueger add a transient superstar shock to such a process. The superstar shock increases productivity by more than a factor of 100 and makes the earnings of superstars somewhat unresponsive to the top tax rate. The superstar shock is not estimated from panel data on earnings or wage rates. The model employed in our paper differs chiefly in that skills are endogenous and top earners become top earners late in life partly by their own costly investments. The endogeneity of skills reduces the revenue maximizing top rate compared to what would hold in an observationally-equivalent, exogenous-skill model similar to those used by all of these authors.

Bruggemann (2017) also analyzes tax reforms that are directed at the upper tail but within a model where top earners have entrepreneurial income and also experience superstar labor income shocks. Allowing for entrepreneurs reduces the top 1 percent tax rate that is welfare maximizing compared to an otherwise similar model without self-financing enterpreneurs. Self financing makes entreprenuerial income responsive to the top tax rate even though all skills are exogenously determined in the model. The welfare maximizing top tax rate in the model is 75 percent.

Many papers characterize the optimal wedges between marginal rates of substitution and transformation in a solution to a planning problem. Optimal wedges are not analyzed in our paper. Kapicka (2015) and Stantcheva (2017) present results on optimal wedges based on different assumptions about the observability of human capital or human capital investments and the type of human capital investment.

[5] This comparison uses the same risk aversion coefficient (log utility) in both models.

2 Framework

An agent maximizes expected utility which is determined by consumption $c = (c_1, ..., c_J)$, work time $l = (l_1, .., l_J)$ and learning time decisions $s = (s_1, ..., s_J)$.

Problem P1: $\max E[\sum_{j=1}^{J} \beta^{j-1} u(c_j, l_j + s_j)]$ subject to

$$
\begin{aligned}
c_j + k_{j+1} &\leq e_j + k_j(1+r) - T_j(e_j, c_j, rk_j) \text{ and } k_{j+1} \geq 0, \forall j \geq 1 \\
e_j &= wh_j l_j \text{ for } j < \textit{Retire} \text{ and } e_j = 0 \text{ otherwise} \\
h_{j+1} &= H(h_j, s_j, z_{j+1}, a),\ 0 \leq l_j + s_j \leq 1 \text{ and } k_1 = 0.
\end{aligned}
$$

Consumption c_j, work time l_j and learning time s_j decisions at age j are functions of initial conditions $\hat{x} = (h_1, a) \in \hat{X}$, age j and shock histories $z^j = (z_1, ..., z_j)$. An agent enters the model with initial skill level h_1 and an immutable learning ability level a. Shocks z_{j+1} impact an agent's skill level. Shocks are idiosyncratic in that the probabilities of shock histories coincide with the fraction of agents that receive that history.

An agent faces a budget constraint where period resources equal labor earnings e_j, the value of financial assets $k_j(1+r)$ that pay a risk-free return of r less net taxes T_j. These resources are divided between consumption c_j and savings k_{j+1}. Each period an agent divides up one unit of available time into distinct uses: work time l_j and learning time s_j. Leisure time is $1 - l_j - s_j$. Earnings e_j equal the product of a rental rate w, skill h_j and work time l_j before an exogenous retirement age, denoted *Retire*, and is zero afterwards. Learning time s_j and learning ability a affect future skill through the law of motion $h_{j+1} = H(h_j, s_j, z_{j+1}, a)$.

The economy has an overlapping generations structure. The fraction μ_j of age j agents in the economy satisfies $\mu_{j+1} = \mu_j/(1+n)$, where n is the population growth rate. There is an aggregate production function $F(K, L)$ with constant returns by which output is produced from capital K and labor L. Physical capital depreciates at rate δ.

The variables (K, L, C, T) are aggregate quantities of capital, labor, consumption and net taxes per agent. Aggregates are straightforward functions of the decisions of agents, population fractions $(\mu_1, \mu_2, ..., \mu_J)$ and the distribution ψ of initial conditions. For example, aggregate capital and labor are the weighted sum of the mean capital and labor within each age group.

$$K = \sum_{j=1}^{J} \mu_j \int_{\hat{X}} E[k_j(\hat{x}, z^j)|\hat{x}]d\psi \text{ and } L = \sum_{j=1}^{J} \mu_j \int_{\hat{X}} E[h_j(\hat{x}, z^j)l_j(\hat{x}, z^j)|\hat{x}]d\psi$$

$$T = \sum_{j=1}^{J} \mu_j \int_{\hat{X}} E[T_j(wh_j(\hat{x}, z^j)l_j(\hat{x}, z^j), c_j(\hat{x}, z^j), rk_j(\hat{x}, z^j)|\hat{x}]d\psi$$

Definition: *A steady-state equilibrium consists of decisions (c, l, s, k, h), factor prices (w, r) and government spending G such that (1) Decisions: (c, l, s, k, h) solve Problem P1, (2) Prices: $w = F_2(K, L)$ and $r = F_1(K, L) - \delta$, (3) Government Budget: $G = T$ and (4) Feasibility: $C + K(n + \delta) + G = F(K, L)$.*

This paper analyzes the steady state implications of changing the top tax rate in this model, while keeping government spending per person G unchanged. Transitional paths between steady states, that are due to changing the top tax rate, are also computed.

3 Empirics

Three issues are addressed. How does the US age-earnings distribution move with age? How do US marginal tax rates vary with a measure of income? What are US values for the coefficients (a_1, a_2, a_3) that enter the tax rate formula? The answers are used to calibrate the model.

3.1 Age Profiles

Tabulated Social Security Administration (SSA) male earnings data from Guvenen, Ozkan and Song (2014) and Panel Study of Income Dynamics (PSID) male hours data are used to describe how earnings and hours move with age. The statistics that we analyze at each age and year of the data sets are (i) real median earnings, (ii) the 10-50, 90-50 and 99-50 earnings percentile ratios, (iii) the Pareto statistic at the 99th percentile threshold of earnings and (iv) the mean fraction of time spent working. The Pareto statistic is $\bar{y}/(\bar{y} - \underline{y})$, where $\underline{y}$ is a threshold and $\bar{y}$ is mean earnings for all observations above this threshold. Time spent working is measured as work hours divided by total discretionary time (i.e. 14 hours per day times 365 days per year). These data sets are described in Appendix A.

Figure 1 highlights age profiles. These profiles are determined by regressing each statistic, measured for each age and year in the data set, on a third-order polynomial in age and a time dummy variable. The estimated age polynomials are plotted in Figure 1 after adding a constant term so that the adjusted polynomial passes through US data values at age 45 in 2010.[6] Earnings profiles are plotted up to age 55 as SSA earnings data goes up to age 55.

Figure 1 shows that median earnings more than double over the working lifetime and that the 90-50 and the 99-50 earnings percentile ratio both increase over most of the working lifetime. The 99-50 earnings percentile ratio roughly doubles from more than 4 at age 25 to 9 at age 55, whereas the Pareto statistic decreases with age. These facts imply that earnings dispersion increases with age above the median and increases quite strongly at the very top.[7]

3.2 Tax Function

TAXSIM is used to characterize marginal tax rates in 2010.[8] Based on earnings for one earner in thousand dollar increments, TAXSIM calculates total taxes, which include federal and state income taxes and the employee and employer parts of all social security and medicare taxes, for a couple filing jointly living in a specific state. A marginal tax rate is computed as the change in total taxes divided by the change in total earnings, where total earnings also include the employer component of social security and medicare taxes.

Figure 2 displays the relationship between this income measure and marginal tax rates when averaged across states.[9] The marginal tax rate schedule tends to increase with income. It jumps at thresholds where federal income tax brackets increase and it falls where some tax rates no longer apply (e.g. the cap on social security taxes). The schedule is somewhat flat for a range of income beyond 300 thousand dollars.

The model marginal tax rate function is constructed by passing a piecewise-linear function through the empirical schedule. The last point in this approximation is set to $\tau = 0.422$ which is the marginal rate evaluated at the 99th percentile of income in the US in 2010. The 99th percentile of income is calculated in Table 1 in section 3.3. The tax function, labeled $T^{prog}(e; \tau)$, is constructed by integrating the model marginal tax rate function.

3.3 Tax Formula Coefficients

The coefficients (a_1, a_2, a_3) and elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ that enter the tax rate formula are defined in theoretical terms in section 5.3. In terms of data, the coefficients (a_1, a_2, a_3) are simple

[6] This holds for all the statistics except median earnings, which is scaled to equal 1 at age 25. The profile for the average fraction of time spent working is based on estimating age and time dummy variables rather than time dummies and an age polynomial.

[7] Appendix B examines the robustness of the profiles in Figure 1 to controlling for cohort effects.

[8] TAXSIM is a computer program that encodes the relationship between sources of income and statutory federal and state income taxes, given household characteristics. See Feenberg and Coutts (1993).

[9] Averages are calculated using state employment as weights. Source: http://www.bls.gov/lau/rdscnp16.htm

ratios of income received or taxes paid to a measure of income above a threshold. To calculate these coefficients using US data, start with an empirical measure Tax of aggregate taxes and decompose Tax into the three components stated below: (i) $N_1Tax_1 + N_1Tax_3$ are total taxes paid by tax units with income above a threshold $\underline{y}$, (ii) N_2Tax_2 are total taxes paid by tax units below the threshold. The total tax paid by units above $\underline{y}$ is decomposed into a component N_1Tax_1 paid on one specific income type and a component N_1Tax_3 which captures all other taxes paid by these top earners. Let N denote the total number of tax units and let N_1 and N_2 denote the number of tax units above and below the threshold $\underline{y}$.

$$Tax = N_1Tax_1 + N_1Tax_3 + N_2Tax_2 \text{ and } N = N_1 + N_2 = .01N + .99N$$

$$a_1 = \frac{N_1\bar{y}}{N_1(\bar{y} - \underline{y})} = \frac{\bar{y}}{\bar{y} - \underline{y}}, \quad a_2 = \frac{N_2Tax_2}{N_1(\bar{y} - \underline{y})} = 99\frac{Tax_2}{(\bar{y} - \underline{y})} \text{ and } a_3 = \frac{N_1Tax_3}{N_1(\bar{y} - \underline{y})} = \frac{Tax_3}{(\bar{y} - \underline{y})}$$

Table 1 summarizes the results of applying this accounting framework to US data when the threshold $\underline{y}$ is the 99th percentile of income, the income measure is based on the income categories for tax units in Statistics of Income (SOI) data and $\bar{y}$ is mean income for tax units beyond the income threshold.[10]

To understand the logic behind these calculations, start with Panel (a) of Table 1. Tax is the sum of personal taxes, social insurance taxes and taxes on production from Bureau of Economic Analysis (BEA) tables.[11] Personal taxes equal federal, state and local income taxes. Social insurance taxes equal social security and medicare taxes. Taxes on production include sales, excise and property taxes.

Panel (b) calculates the part of each aggregate tax measure paid by the top 1 percent based on SOI data. This involves first calculating $\underline{y}$, the 99th percentile of our income measure. The value N_2Tax_2 is calculated as a residual based on Tax and the part of these taxes paid by the top 1 percent.

Panel (c) divides the taxes paid by the top 1 percent into two parts. N_1Tax_3 equals the sum of taxes on production paid by the top 1 percent plus a measure of capital income taxes paid by the top 1 percent, whereas N_1Tax_1 equals personal taxes and social insurance taxes paid by the top 1 percent less these capital income taxes. The quantity *capital income tax* equals the

[10] Our measure is the sum of (i) wages and salaries, (ii) interest, (iii) non-qualified dividends, (iv) business income, (v) IRA distributions, (vi) pensions and annuities, (vii) total rent and royalty, (viii) partnership and S-corporation income and (ix) estate and trust income. It excludes qualified dividends and capital gains.

[11] The main tax category in BEA tables that our empirical analysis does not account for is corporate income taxes.

Table 1 - Tax Formula Coefficients: 2010 US Data

Panel (a)	Tax = Personal Taxes + Social Insurance Taxes + Taxes on Production $Tax = 1239.3 + 844.0 + 1057.1 = 3140.4$ billion dollars $N = 156.167$ million tax units $N_1 = 0.01 \times N$ and $N_2 = 0.99 \times N$
Panel (b)	$N_1 Tax_1 + N_1 Tax_3 = (1) + (2) + (3) = 548.44$ billion dollars (1) personal taxes = 453.73 (2) social insurance taxes = 47.24 (3) taxes on production = 47.47 $N_2 Tax_2 = Tax - (N_1 Tax_1 + N_1 Tax_3) = 2591.96$
Panel (c)	$N_1 Tax_1 = (1) + (2)$ - capital income tax = 434.91 billion dollars $N_1 Tax_3 = (3) +$ capital income tax = 113.52 capital income tax = 66.05
Panel (d)	$a_1 = \frac{\bar{y}}{\bar{y}-\underline{y}} = 1.70, \quad a_2 = 99\frac{Tax_2}{(\bar{y}-\underline{y})} = 3.64$ and $\quad a_3 = \frac{Tax_3}{(\bar{y}-\underline{y})} = 0.16$ $(\underline{y}, \bar{y}) = (319.5, 775.8)$ and $(Tax_1, Tax_2, Tax_3) = (278.5, 16.8, 72.7)$ [income and taxes are in thousand dollar units]

Notes: Appendix A describes the data sources and the details for all the calculations in Table 1.

sum of all the qualified dividends and capital gains received by tax units with income beyond $\underline{y}$ multiplied by the tax rate $\tau_k = 0.20$ calculated in section 4 based on federal and state tax rates in 2010. Capital income taxes are subtracted from personal taxes because, for tax units with high income, qualified dividends and long-term capital gains are taxed at a 15 percent federal tax rate as compared to the top federal rate of 35 percent in 2010 on ordinary income. We do so because the goal is to determine the revenue consequences of increasing the top tax rate without changing other aspects of the tax system, including the preferential rate on sources of capital income.

The Pareto statistic $a_1 = \bar{y}/(\bar{y} - \underline{y})$ is 1.70 in Table 1 at the 99th percentile of our income measure. Diamond and Saez (2010) present evidence that $a_1 = 1.5$ for the US in 2005 for a range of top income thresholds, including the 99th percentile, based on using adjusted gross income (AGI) as an income measure. AGI includes capital gains and qualified dividends that are taxed at preferential rates. Our income measure excludes capital gains and qualified dividends. If one wants to determine the top tax rate on ordinary income that is revenue maximizing, holding other tax rates fixed, then a_1 needs to be calculated excluding income sources taxed at preferential rates.

4 Model Calibration and Validation

We calibrate the model and then analyze its validity.

4.1 Calibration

The functional forms for the utility function u, production function F and human capital law of motion H are widely used in the literature and are stated below. The bivariate distribution ψ of initial conditions is characterized by 6 parameters. Marginal distributions for learning ability and initial human capital are both Pareto-Log-Normal (PLN) distributions.[12]

Benchmark Model Functional Forms:

Utility: $u(c, l+s) = \log(c) - \phi \frac{(l+s)^{(1+\frac{1}{\nu})}}{1+\frac{1}{\nu}}$

Production: $Y = F(K, L) = AK^{\gamma}L^{1-\gamma}$

Human Capital: $H(h, s, z, a) = \exp(z)[h + a(hs)^{\alpha}]$ and $z \sim N(\mu_z, \sigma_z^2)$

Initial Conditions: $a \sim PLN(\mu_a, \sigma_a^2, \lambda_a), \log h_1 = \beta_0 + \beta_1 \log a + \log \epsilon$ and $\epsilon \sim LN(0, \sigma_\epsilon^2)$

Some parameters are set to fixed values without computing equilibria to the model economy. Parameters governing demographics, technology and the tax system are set in this way. The parameter governing the standard deviation of human capital shocks is set to an estimate from Huggett, Ventura and Yaron (HVY) (2011). The remaining parameters are set so that equilibrium properties of the model best match empirical targets. Appendix B describes the computation of an equilibrium and the objective that is minimized.

Demographics An agent enters the model at a real-life age of 23, retires at age 65 and lives up to age 85. These three ages correspond to $j = 1, 43$ and 63. The population growth rate $n = 0.01$ is set to the geometric average growth rate of the U.S. population over the period 1960-2015. Population fractions μ_j sum to 1 and decline with age by the factor $(1 + n)$.

Technology US national accounts data imply that capital's share, the investment-output ratio and capital-output ratio averaged $(0.352, 0.174, 3.22)$ over the period 1960-2015. Set γ to match capital's share. Set δ to be consistent with the investment-output ratio and the capital-output ratio, given n. Normalize A so that the wage is 1 when the model produces the capital-output ratio measured in the data.

[12]Initial human capital, thus, has a Pareto tail. Appendix B presents basic properties of PLN distributions. The random variable ϵ used in the construction is independent of learning ability.

Table 2 - Benchmark Model Parameter Values

Category	Functional Forms	Parameter Values
Demographics	$\mu_{j+1} = \mu_j/(1+n)$	$Retire = 43, n = 0.01$ $j = 1, ..., 63$ (ages 23-85)
Technology	$Y = F(K, L) = AK^{\gamma}L^{1-\gamma}$ and δ	$(A, \gamma, \delta) = (0.878, 0.352, 0.044)$
Tax System	$T_j = T^{prog}(e;\tau) + \tau_c c + \tau_k kr$ for $j < Retire$ $T_j = \tau_c c + \tau_k kr - transfer$ for $j \geq Retire$	T^{prog} see Figure 2 $\tau_c = 0.10$ and $\tau_k = 0.20$ $transfer = 18115$
Preferences	$u(c, l+s) = \log c - \phi \frac{(l+s)^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}}$	$\phi = 12.4, \beta = 0.962, \nu = 0.614$
Human Capital	$H(h, s, z, a) = \exp(z)\,[h + a(hs)^{\alpha}]$ and $z \sim N(\mu_z, \sigma_z^2)$	$\alpha = 0.543$ $(\mu_z, \sigma_z) = (-0.047, 0.111)$ σ_z follows HVY (2011)
Initial Conditions	$a \sim PLN(\mu_a, \sigma_a^2, \lambda_a)$ and $\epsilon \sim LN(0, \sigma_\epsilon^2)$ $\log h_1 = \beta_0 + \beta_1 \log a + \log \epsilon$	$(\mu_a, \sigma_a^2, \lambda_a) = (0.383, 3.0E-8, 3.58)$ $(\beta_0, \beta_1, \sigma_\epsilon^2) = (3.60, 1.01, 0.211)$

Note: Demographic, Technology and Tax System parameters and σ_z are set without solving for equilibrium. All remaining model parameters are set so that equilibrium values best match targeted moments. Parameters are typically rounded to display 3 significant digits.

Tax System Taxes in the model are $T_j(e_j, c_j, k_j r) = T^{prog}(e_j) + \tau_c c_j + \tau_k k_j r$ for $j < Retire$ and $T_j(e_j, c_j, k_j r) = \tau_c c_j + \tau_k k_j r - transfer$ for $j \geq Retire$. The function T^{prog} was calculated in section 3. The consumption tax rate $\tau_c = 0.10$ is the ratio of taxes on production in 2010 to total consumption expenditures from BEA Table 3.1 and Table 1.1.5. The common social security transfer equals 18115 dollars. This is the yearly old-age benefit for a worker retiring in 2010 based on an earnings history equal to average earnings.[13]

We calculate a marginal tax rate τ_k on capital as follows. For each state, input long-term capital gains in thousand dollar increments into TAXSIM for a couple filing jointly with earnings equal to 160 thousand dollars in 2010. The marginal tax rate equals the change in total taxes divided by the change in income. The US schedule is the employment-weighted average of the state marginal rate schedules and is flat beyond 280 thousand dollars of capital gains. Set $\tau_k = 0.20$ which is the US marginal rate calculated at this level. The federal marginal rate on long-term capital gains was 15 percent in 2010.

Preferences The period utility function over consumption is log utility. This choice controls the strength of the income effect of a tax reform. Chetty (2006, p.1830) states *"A large literature on labor supply has found that the uncompensated wage elasticity of labor supply is not very negative. This observation places a bound on the rate at which the marginal utility of consumption diminishes, and thus bounds risk aversion in an expected utility model. The central estimate of the coefficient of relative risk aversion implied by labor supply studies is 1*

[13]See Table 2.A26 in the 2010 Annual Statistical Supplement of the Social Security Bulletin available at https://www.ssa.gov/policy/docs/statcomps/supplement/2010/2a20-2a28.pdf

(log utility) and an upper bound is 2 ... ."

Remaining Parameters All remaining model parameters are set to minimize the sum of squared differences between model moments and data moments.[14] We use the following data moments as targets: (i) the age profiles documented in Figure 1, (ii) the cross-sectional Pareto statistic $a_1 = 1.7$ and income threshold calculated in Panel (d) of Table 1, (iii) the US capital-output ratio $K/Y = 3.22$ and (iv) the regression coefficient estimated by MaCurdy (1981). Remaining parameters are those governing (i) initial conditions, (ii) the elasticity of the human capital production function α and the mean of the human capital shock μ_z and (iii) the utility function parameters (β, ϕ, ν).

The last target mentioned above is based on evidence from the literature on the Frisch elasticity of labor supply. The regression equation used by MaCurdy (1981) is stated below. The target value for θ_1 is 0.125.[15]

$$\Delta \log hours_j = \theta_0 + \theta_1 \Delta \log wage_j + \epsilon$$

To connect to this evidence, calculate model wages as earnings divided by model hours and estimate the coefficients in the linear regression based on agents age 25-55. Appendix B discusses the results of the estimation of this regression, the instrumental variable methods employed and reasons why the model regression coefficient is below the model value of ν. Disciplining the model to match the empirical regression coefficient θ_1 would seem to be important. This evidence is behind the view that labor hours are not very elastically supplied by prime-age males and has been used to support the view (see Keane (2011) and Saez, Slemrod and Giertz (2012 p. 3-4)) that a very high top tax rate may be revenue maximizing.

Figure 3 graphs the age profiles in US data and in the model using the model parameters which best match these targets. The model exactly reproduces (up to three digits) the Pareto statistic $a_1 = 1.70$, the 99th income threshold, the capital-output ratio $K/Y = 3.22$. The regression coefficient is $\theta_1 = 0.125$ from US data while it is $\theta_1 = 0.124$ in the model. The model parameters that produce these results are stated in Table 2.

Figure 3 shows that the 99-50 earnings ratio increases with age and that the Pareto statistic $(\bar{y}_j/(\bar{y}_j - \underline{y}_j)$ at the 99th percentile of the earnings distribution decreases with age. The model economy has two forces that can produce these properties: shocks and learning ability differences. Shocks lead ex-ante identical agents to have different earnings ex-post, whereas learning

[14] Appendix B5 specifies the objective that is minimized and the minimization procedure.

[15] This is the average from MaCurdy (1981, Table 1 row 5-6) who uses data for white males age 25-55. Altonji (1986) finds similar results. Domeij and Floden (2006, Table 5) estimate $\theta_1 = 0.16$ for male household heads age 25-60 using 1984-94 PSID data. Keane (2011) and Keane and Rogerson (2012) review this literature.

ability differences rotate age-earnings profiles. Good learners have steep mean age-earnings profiles as a result of human capital investment. The standard deviation of shocks $\sigma_z = 0.111$ is set to an estimate from Huggett, Ventura and Yaron (2011), who estimate this parameter using specific moments of log wage rate changes for older workers in panel data. Given this estimate, the parameters of the distribution of initial conditions are set to match the earnings facts. Table 2 shows that learning ability dispersion is needed to match the facts. Thus, the growing earnings dispersion at the top with age in the data dictates a role for learning ability differences beyond that due to shocks impacting earnings.

4.2 Validation

We conduct three validation exercises by examining the degree to which the model matches data properties that were not used to calibrate the model.

4.2.1 Short-term Income Elasticities

Evidence on the response of income measures to a change in the net-of-tax rate comes from the elasticity literature. Saez, Slemrod and Giertz (SSG) (2012) review this literature. They highlight the panel regression framework as the framework that most convincingly identifies a short-term income response to a change in the net-of-tax rate arising from a tax reform. The panel approach regresses the growth in income of tax unit i on the growth of the marginal net-of-tax rate for unit i, an income control $f(z)$ and time dummies α_t.

$$\log\left(\frac{z_{it+1}}{z_{it}}\right) = \epsilon \log\left(\frac{1-\tau_{t+1}(z_{it+1})}{1-\tau_t(z_{it})}\right) + \beta f(z_{it}) + \alpha_t + \nu_{it+1}$$

We apply this regression framework to a tax reform in the model and compare model regression results to those from US data. We compute a transition path for the model economy due to a tax reform in model period 3 that permanently increases the tax rate at the 99th percentile in Figure 2 from $\tau = 0.422$ to $\tau = 0.49$. Figure 4 plots the transition path arising from the reform. Aggregate capital and labor input per agent fall by a similar percentage.[16] The aggregate labor input falls immediately in the reform year by an adjustment in work time and adjusts slowly afterwards due to skill change.

We construct 100 randomly-drawn, balanced panels each with 30 thousand agents with earnings in the top 10 percent in model period 1 and follow these agents from period 1 to period 7. This mimics the structure of the US 1991-1997 panel used by SSG (2012, Table 2). The key

[16] Along the transition path, any extra tax revenue collected beyond that needed to fund government purchases is returned as a lump-sum transfer each period.

point for US data is that in 1993 a tax reform increased the marginal tax rate on the top 1 percent without a substantial tax rate change at lower income levels.

Table 3 - Panel Regression Based on Model Data

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Elasticity ϵ	0.236	0.214	0.379	0.475	0.379	0.475
S.D.	(0.048)	(0.089)	(0.032)	(0.031)	(0.032)	(0.031)
Income Control $f(z) = \log z$	No	Yes	No	Yes	No	Yes
Time Effects α_t	No	No	Yes	Yes	Yes	Yes
Instrument 1: $1_{\{i \in T_2\}}$	Yes	Yes				
Instrument 2: $1_{\{i \in T_2\}} 1_{\{t=2\}}$			Yes	Yes		
Instrument 3: $\log(\frac{1-\tau_{t+1}(z_{it})}{1-\tau_t(z_{it})})$					Yes	Yes
Use data for time periods	$t = 2, 3$	$t = 2, 3$	All	All	All	All

Note: Means and standard deviations of the point estimates of ϵ are based on 100 randomly drawn balanced panels. We employ the same instrument definitions and estimation technique (i.e. two-stage-least squares) as were employed in Saez et al. (2012, Table 2 Panel B2). Instrument 1 is $1_{\{i \in T_2\}}$, where T_2 is the set of agents in the top 1 percent at $t = 2$. Instrument 2 is the interaction of being in the top 1 percent in year 2 and the year being $t = 2$. Instrument 3 is the predicted log marginal net-of-tax rate change.

Table 3 reports means and standard deviations of the point estimate of the model regression coefficient ϵ across the 100 panels. The instruments and income controls used in Table 3 were employed in SSG (2012, Table 2, Panel B1) who find that point estimates based on US data vary widely using only two years of data and depend sensitively on controls. For example, using 1992 and 1993, they find that point estimates range from -0.721 to 1.395. SSG (2012, Table 2, Panel B2) find that, based on using all years of the 1991-1997 data and using income controls, point estimates range from 0.143 to 0.564 depending on the precise instruments employed. This is the type of evidence that underlies the view that $\epsilon = 0.25$ is a relevant US (short term) estimate to input into the widely-used formula $\tau^* = 1/(1 + a\epsilon)$. Table 3 shows that, using model data for all seven time periods, the mean of the model point estimates range from a low of 0.379 to a high of 0.475.[17] Broadly, the model regression coefficients are consistent with the range of estimation results documented by SSG.[18]

4.2.2 Earnings Growth By Lifetime Earnings Group

We examine the relationship between lifetime earnings for individual males in Social Security Administration panel data and the growth rate of earnings over the lifetime. Guvenen, Karahan, Ozkan and Song (2015) divide males into 100 equal-sized bins based on percentiles of lifetime earnings. Figure 5 plots the ratio of mean real bin earnings at age 55 to mean real bin earnings at age 25 and at age 30. Mean earnings grow dramatically over the working lifetime for the top

[17] As the model tax reform occurs at exactly the 99th percentile, Instrument 2 and 3 are perfectly (negatively) correlated.

[18] Appendix B examines US long-run response estimates and runs the same regressions on model data.

several bins of lifetime earnings. For example, real earnings increase in the top bin by a factor of 15 between age 25 and 55 and by a factor of 7 between age 30 and 55.

Figure 5 shows that the human capital model displays the qualitative patterns observed in US data for top lifetime earners when model agents are put into 100 bins following the procedure used on US data. The key model feature that produces the data patterns is that agents differ in learning ability and better learners optimally choose steeper mean earnings profiles via time investments in skill formation. Top lifetime earners are disproportionally those entering the model with high learning ability.[19] Some existing quantitative-theoretical models of top earners that rely on shocks with a transitory impact on earnings (e.g. Kaymak and Poschke (2016) and Heathcote, Storesletten and Violante (2017)) do not produce the strikingly large earnings growth rates documented for US top lifetime earners.

Badel, Daly, Huggett and Nybom (2018) use administrative earnings data for males from Canada, Denmark and Sweden and apply the same data construction procedure used on US data. Figure 5 shows that earnings ratios over the working lifetime are strikingly large for the top several bins of lifetime earnings in all countries.[20] Thus, theories of top earners will broadly need to account for this empirical regularity.

4.2.3 Distribution of Taxes

Table 4 describes the distribution of US federal income taxes and model taxes by income group. US federal income taxes are largely paid by the top of the income distribution. In 2010, the top 1 and top 10 percent of taxpayers by AGI paid roughly 37 and 71 percent of total federal income taxes. A comparison of this tax with model taxes cannot be made directly as model taxes combine federal and state income taxes and payroll taxes. Nevertheless, the second column labeled "Model Federal Income Tax" is an attempt to arrive at such an apples-to-apples comparison. Model federal income taxes are very concentrated at the top but are less concentrated than in US data. Total model income taxes plus payroll taxes are less concentrated at the top compared to model federal income taxes because payroll taxes are capped and because state income taxes are roughly a flat tax.

[19] Table 2 shows that (log) learning ability and (log) initial human capital are positively correlated in the model.

[20] Results for Canada are not graphed to allow a clear visualization, but are very similar to the US profile.

Table 4 - Distribution of Taxes: Model vs. Data

Income Group	US Data: Federal Income Tax 2010	Model Federal Income Tax	Model Income Tax plus Payroll Tax
Top 50 percent	97.6	97.0	91.7
Top 20 percent	83.0	81.0	71.2
Top 10 percent	70.6	67.0	55.4
Top 1 percent	37.4	31.0	22.0

Note: US Data is http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Rates-and-Tax-Shares. Model income is earnings plus interest income. Model Federal Income Tax is based on adding earnings taxes and interest income taxes, where earnings taxes are based on applying TAXSIM to model earnings and where interest income taxes are based on mutiplying model interest income by 15 percent - the federal capital gains tax rate in 2010. Model Income Tax plus Payroll Tax are total model taxes less model consumption taxes.

5 Analyzing the Tax Reform

5.1 Laffer Curve and Welfare

We analyze a reform that permanently alters the top tax rate on earnings but leaves all other tax rates and government spending unchanged. The top tax rate is increased at the earnings threshold for the top tax rate. If more revenue is collected under the new tax system, then the extra revenue is returned in equal, lump-sum transfers to all agents.

Figure 6 displays the Laffer curve in the benchmark model. The horizontal axis measures the top tax rate and the vertical axis measures the aggregate lump-sum transfer as a percentage of pre-reform output. The Laffer curve peaks at a 49 percent tax rate. The transfer is below a tenth of one percent of the pre-reform, steady-state output level. Thus, the Laffer curve in the benchmark model is flat in that little additional revenue is raised.

Figure 6 also presents Laffer curves when the target value for θ_1, governing the Frisch hours elasticity with respect to wage rate, is altered from the empirical benchmark value of $\theta_1 = 0.125$. The target value is varied by multiples of MaCurdy's estimate of the standard error $SE(\theta_1) = 0.139$. The peak of the model Laffer curves occur at larger top tax rates when the target value for θ_1 decreases and at smaller top tax rates when the target value for θ_1 increases. We display the resulting Laffer curves only for the cases when the target values are decreased. The Laffer curve peaks at $\tau^* = 0.55$ when the target value is $\theta_1 = 0.056$ and at $\tau^* = 0.63$ when the target value is $\theta_1 = -0.014$. When the target value for θ_1 decreases then the utility function parameter ν decreases.[21]

The steady state welfare consequences of increasing the top tax rate are measured as the

[21] When the one target is changed then we reestimate all model parameters to minimize the same distance measure between model and data moments used in Table 2. The value of the utility function parameter ν is $0.614, 0.489, 0.427$ when the target value of θ_1 is $0.125, 0.056, -0.014$.

percentage change in consumption at all ages in the benchmark model that is equivalent in ex-ante expected utility terms for age $j = 1$ agents (age 23 in real life) to the ex-ante expected utility achieved at alternative values of the top tax rate. Figure 6 shows that there is a small welfare loss of 0.037 percent of consumption in the benchmark model from increasing the top tax rate to the revenue-maximizing level. While there is a redistributional and insurance benefit arising from the lump-sum transfer, there is also a small fall in the wage w. Factor prices change very little across steady states as aggregate capital and labor input fall by nearly the same percentage when the top tax rate increases.

The small equivalent consumption change in Figure 6 for all age 1 agents masks larger changes, conditional on initial conditions. Moving to the revenue maximizing top tax rate of $\tau = 0.49$ is equivalent to a decrease in consumption of 3 percent at each age over the lifetime for agents in the top 1 percent of learning ability and to a small increase in consumption for agents in the bottom 99 percent of learning ability. Good learners will be directly impacted by the tax later in life if they cross the income threshold associated with the top tax rate.

While the steady-state comparison in Figure 6 does not produce a large welfare gain, the reform is quite popular. Specifically, we compute the transition path associated with the reform that permanently increases the top tax rate to $\tau = 0.49$. It turns out that, among the agents that are alive at the first period of the reform, 95 percent of the agents increase their discounted expected utility. Thus, 95 percent of these agents prefer the reform to the status quo. While not a single top earner (i.e. those in the top 1 percent of the earnings distribution before the reform) prefers the reform, almost all of the retired agents prefer the reform. The lump-sum transfer that these retired agents receive more than compensates for the small initial fall in the interest rate. This transfer declines over time as the economy converges to the new steady state. Figure 4 previously indicated that this reform produces a monotonic decline in aggregate capital, labor and output that is eventually about 1 percent below the initial steady state levels. The steady state fall in output and consumption per capita that is associated with a permanent increase in the top tax rate to $\tau = 0.73$ is 3.5 and 4.5 percent respectively.

5.2 Understanding the Role of Human Capital Accumulation

What role does skill change play in accounting for the shape of the model Laffer curve? To answer this question, we construct an exogenous human capital model with the same preferences, technology, initial conditions and tax system as the benchmark model. Both models are observationally equivalent in that they produce the same joint distribution of consumption, wealth, earnings and income by age and in that they have the same dynamics of these variables over time for individual agents under the benchmark tax system. The key difference is that when the tax system changes then human capital investments change in the human capital model but remain unchanged in the exogenous human capital model.[22]

Figure 7 plots the Laffer curve in both models. The top of the Laffer curve for the exogenous human capital model raises additional tax revenue of roughly two-tenths of one percent of output in the benchmark model steady state. Thus, endogenous skill change flattens out the Laffer curve compared to an otherwise similar model that ignores the possibility of skill change in response to changes in the tax system. The top of the Laffer curve in the exogenous human capital model occurs at a top tax rate of 59 percent.

Intuitively, the Laffer curves differ because labor input is more elastic with respect to a change in the top rate in the human capital model. We decompose the change in the aggregate labor input across steady states and find that slightly more than half of the fall in the aggregate labor input from the original steady state to the steady state with top rate set to 49 percent is due to skill change at fixed work time levels as opposed to changes in work time at fixed skill levels.[23] The largest percentage fall in skills occurs at the end of the working lifetime from agents endowed with high learning ability.[24]

$$wh_j(1-\tau_j') = \sum_{k=j+1}^{Retire-1} (\frac{1}{1+\hat{r}})^{k-j} \frac{dh_k}{ds_j} wl_k(1-\tau_k')$$

The mechanism behind the fall in skill is easily grasped from the Euler equation governing skill investment above, where we abstract from idiosyncratic risk for simplicity. At a best choice an agent equates the marginal cost $wh_j(1-\tau_j')$ of an extra unit of time spent in skill production at age j to the discounted marginal benefit of the extra skill production $\frac{dh_k}{ds_j}$ in future periods, where $\hat{r}$ is the after-tax real interest rate. Now consider an increase in the top tax rate. Absent any adjustment, the left-hand side of the Euler equation does not change for an agent with earnings below the top tax rate but some of the marginal net-of-tax-rate terms $(1-\tau_k')$ decrease for an agent that will be above the threshold in the future. Thus, some adjustment must occur. A decrease in time investment in skill production increases the future marginal product terms $\frac{dh_k}{ds_j}$. If future labor hours l_k decrease in response to the increase in the top tax rate, consistent with model behavior for many agents with high learning ability, then an even larger fall in skill investment occurs at age j. In summary, an increase in the top tax rate decreases the marginal benefit of skill investment without changing the marginal cost for agents who become top earners later in life.

[22]In the exogenous human capital model all decisions, other than the time investment decision, are allowed to be adjusted to maximize expected utility when the tax system changes. Appendix B describes computation.

[23]The decomposition is $\Delta L = \sum_j \mu_j \int E[\hat{h}_j l_j - h_j l_j|\hat{x}]d\psi + \sum_j \mu_j \int E[\hat{h}_j \hat{l}_j - \hat{h}_j l_j|\hat{x}]d\psi$, where hat variables are calculated at the new tax rate.

[24]Skills early in life (i.e. at age $j=1$ in the model) are assumed to be invariant to the top tax rate.

5.3 Applying the Tax Rate Formula

The goal of this section is to apply the formula in Badel and Huggett (2017, Theorem 1) to the model to understand the quantitative role of the three proximate forces that determine the tax rate at the top of the model Laffer curve. The formula is based on three elements: (i) a probability space of agent types $(X, \mathcal{X}, \mathrm{P})$, (ii) functions $(y_1, ..., y_n)$ that map agent type $x \in X$ and a top tax rate τ into income and expenditure decisions and (iii) a class of tax functions T indexed by τ. T is separable (i.e. $T(y_1, ..., y_n; \tau) = T_1(y_1; \tau) + T_2(y_2, ..., y_n)$) and has a constant top tax rate τ beyond a threshold (i.e. $T_1(y_1; \tau) - T_1(\underline{y}; \tau) = \tau[y_1 - \underline{y}], \forall y_1 > \underline{y}$) .

The formula is stated below. The variables entering the formula are integrals over the sets $X_1 \equiv \{x \in X : y_1(x, \tau^*) > \underline{y}\}$ and $X_2 \equiv \{x \in X : y_1(x, \tau^*) \leq \underline{y}\}$. These are the sets of agent types that have "income" y_1 above and below the threshold $\underline{y}$.

$$\tau^* = \frac{1 - a_2\epsilon_2 - a_3\epsilon_3}{1 + a_1\epsilon_1}$$

$$(a_1, a_2, a_3) = \left(\frac{\int_{X_1} y_1 dP}{\int_{X_1} \left[y_1 - \underline{y}\right] dP}, \frac{\int_{X_2} T(y_1, ..., y_n; \tau^*) dP}{\int_{X_1} \left[y_1 - \underline{y}\right] dP}, \frac{\int_{X_1} T_2(y_2, ..., y_n) dP}{\int_{X_1} \left[y_1 - \underline{y}\right] dP} \right)$$

$$(\epsilon_1, \epsilon_2, \epsilon_3) = \left(\frac{d \log(\int_{X_1} y_1 dP)}{d \log(1 - \tau)}, \frac{d \log(\int_{X_2} T(y_1, ..., y_n; \tau^*) dP)}{d \log(1 - \tau)}, \frac{d \log(\int_{X_1} T_2(y_2, ..., y_n) dP)}{d \log(1 - \tau)} \right)$$

Why are there three elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ in the formula above but only a single elasticity in the widely-used formula? To see why, write aggregate tax revenue below, restate it in three useful parts and note that τ^* maximizing revenue is equivalent to τ^* maximizing an objective with three terms.[25] When the top tax rate τ changes there are exactly three broad reasons why aggregate taxes change: (1) taxes from "top earners" based on income source y_1 change, (2) other taxes collected from "top earners" change and (3) taxes on agent types who are not "top earners" change. The widely-used formula accounts for only the first source of tax revenue variation whereas the Badel-Huggett formula accounts for all three.

$$\int_X T(y_1, ..., y_n; \tau) dP = \int_{X_1} T_1(y_1; \tau) dP + \int_{X_1} T_2(y_2, ..., y_n) dP + \int_{X_2} T(y_1, ..., y_n; \tau) dP$$

[25]We suppress the arguments of the functions being integrated to allow for a compact presentation.

$$\tau^* \in argmax \int_X TdP \Leftrightarrow \tau^* \in argmax \int_{X_1} T_1 dP + \int_{X_1} T_2 dP + \int_{X_2} TdP$$

To apply the formula to the model, define an agent type $x = (h_1, a, j, z^j)$ to be a quadruple of initial skill h_1, learning ability a, age j and (partial) shock history z^j and set (y_1, y_2, y_3) to be labor income, consumption and capital income. For example, $y_1(x, \tau) \equiv w(\tau)h_j(h_1, a, z^j; \tau)l_j(h_1, a, z^j; \tau)$ for $j < Retire$ and 0 otherwise, where τ is the top tax rate, $w(\tau)$ is the equilibrium wage and the functions (h_j, l_j) describe human capital and labor input. Lastly, set $T(y_1, y_2, y_3; \tau)$ to equal the sum of the earnings, consumption and capital income taxes in the model.

Table 5 calculates the model coefficients and elasticities in the human capital model and the exogenous human capital model from the last section. The coefficients (a_1, a_2, a_3) are the same in both models because the distribution of incomes, expenditures and taxes at the initial steady state is exactly the same in both models. The model economy value $a_1 = 1.70$ exactly equals the Pareto statistic $a_1 = 1.70$ calculated in US data in Table 1 as this was a calibration target. The model coefficient $a_2 = 4.508$ is above the US data value of $a_2 = 3.64$ from Table 1, whereas the model coefficient $a_3 = 0.12$ is below the US value of $a_3 = 0.16$ from Table 1.[26]

Table 5 - Revenue Maximizing Top Tax Rate Formula

Terms	endogenous human capital model	exogenous human capital model
$a_1 \times \epsilon_1$	$1.700 \times .317 = .539$	$1.700 \times .222 = .377$
$a_2 \times \epsilon_2$	$4.508 \times .040 = .180$	$4.508 \times .020 = .090$
$a_3 \times \epsilon_3$	$0.120 \times .739 = .089$	$0.120 \times .633 = .076$
$\tau^* = \frac{1 - a_2\epsilon_2 - a_3\epsilon_3}{1 + a_1\epsilon_1}$	.476	.604
τ at peak of Laffer curve	.49	.59

Note: Coefficients (a_1, a_2, a_3) are calculated at the original steady state with $\tau = 0.422$. Calculate the set X_1 of top earner types using the earnings threshold for the top tax rate. Calculate elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ as a difference quotient using $\tau = 0.422$ and the tax rate at the top of the Laffer curve.

Table 5 shows that the two models have different revenue maximizing top tax rates because the exogenous human capital lowers all three elasticities compared to the human capital model. The economic mechanism behind the lower elasticity for ϵ_1 was articulated in section 5.2. This mechanism also reduces (ϵ_2, ϵ_3). For example, ϵ_3 is lower because the consumption and wealth accumulation of top earners are reduced to a lesser degree in the exogenous human capital model by an increase in the top tax rate and, thus, taxes paid on consumption and capital

[26]One reason why the model value a_2 is above the data value is that the data value for the denominator includes earnings and some capital income (e.g. interest income and non-qualified dividends) that are taxed at the ordinary income tax rate, whereas for tractability reasons the model value of the denominator covers only labor market earnings of top earners.

income are reduced to a lesser degree.[27] Factor price changes play little role in determining the three elasticities in the human capital model because aggregate capital and labor decrease by about the same percent when the top tax rate increases.

Why does the Laffer curve for the human capital model peak at a lower tax rate than the 73 percent rate suggested by the established view? Table 4 gives a concise answer. There are two new forces, captured by the terms $a_2\epsilon_2$ and $a_3\epsilon_3$, that in general need to be considered that are not accounted for in the formula $\tau^* = 1/(1 + a\epsilon)$. These new forces are positive in the human capital model and depress the revenue maximizing top tax rate.

How important are the two new forces? This question is answered by plugging into the formula the model values for $\epsilon_1 = .317$ but the counterfactual values $\epsilon_2 = \epsilon_3 = 0$. The revenue maximizing tax rate implied by the formula is then $\tau^* = 1/(1 + 1.70 \times .317) = .65$. Thus, the new forces reduce the top rate from 65 percent to 49 percent in this decomposition.

Although the two new forces account for the bulk of the reduction in the revenue maximizing top rate relative to an established view, it is still useful to explain what accounts for why model parameters are set so that the model Pareto statistic is $a_1 = 1.70$ rather than $a_1 = 1.5$. Diamond and Saez (2011) calculate that a_1 is approximately 1.5 in 2005 for a wide range of income thresholds in the upper tail when income is measured by adjusted gross income (AGI). However, the use of AGI in computing the coefficient a_1 is not consistent with our goal of determining the revenue maximizing top tax rate when changing only the tax rate on ordinary income. This is because AGI includes income sources, qualified dividends and long-term capital gains, that are taxed at a lower rate. Based on 2010 Statistics of Income (SOI) data, we calculate that $a_1 = 1.70$ when income excludes qualified dividends and capital gains and that $a_1 = 1.50$ using AGI as an income measure. Thus, a key reason why we calculate that $a_1 = 1.70$ in 2010, whereas Diamond and Saez (2011) calculate that $a_1 = 1.5$ in 2005, is not that the upper tail differs significantly across years. Instead, it is that the upper tail is thinner after excluding capital income sources that are taxed at a preferential rate.

5.4 Cross Elasticities: US Evidence

The previous section concluded that the key reason that the human capital model has a revenue maximizing top tax rate of 49 percent when the established view suggests that 73 is revenue maximizing is due to two forces not analyzed in the established view. This section presents reduced-form regression evidence for the magnitude of the two new elaticities based on US data

[27] The elasticity ϵ_2 is lower in the exogenous human capital model: (i) earnings, income and wealth fall less and thus taxes on these sources fall less for agents before they become top earners, (ii) retirees' consumption taxes and capital income taxes fall less for agents who were top earners before retirement but who fall below the threshold during retirement.

that focuses on the long-term response. Saez, Slemrod and Giertz (2012, Table 1) present some US long-term response evidence for ϵ_1 based on time series regressions.

The elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ in the tax rate formula involve how aggregate income and tax revenue measures for different groups change when $(1-\tau)$ changes. These elasticities can be calculated as the ratio of the vertical shift in the balanced-growth path of log aggregate income and tax revenues to a permanent change in $\log(1-\tau)$ within any model with a balanced-growth path. This section estimates the two new (cross) elasticities (ϵ_2, ϵ_3) using a statistical model that allows for a permanent shift in the net-of-tax rate and that captures a shift in a balanced-growth path.

We construct an aggregate measure of US taxes, labeled Tax_t, that consists of federal, state and local personal income taxes, social insurance taxes and taxes on production. This is the aggregate tax measure used previously in Table 1. Aggregate taxes are decomposed following the logic of the formula: $Tax_t = Rev_{1t} + Rev_{2t} + Rev_{3t}$.[28] Taxes paid by tax units with income above the top 1 percent equal $Rev_{1t} + Rev_{3t}$, whereas Rev_{2t} is the part paid by tax units below the top 1 percent. Rev_{3t} equals taxes on production plus capital income taxes paid by the top 1 percent on sources of capital income that have preferential tax rates (e.g. capital gains and qualified dividends). Top 1 percent income is measured without capital gains as described in Appendix B.

Figure 8 plots the tax revenue series we calculate and the average marginal tax rate for the top 1 percent of tax units calculated by Mertens and Montiel-Olea (2017).[29] Much of the tax rate variation appears to be transitory; however, the tax rate declined in a dramatic and roughly permanent fashion in the mid 1980s.[30]

$$\begin{pmatrix} \log(1-\tau_t) \\ \log Rev_{2t} \\ \log Rev_{3t} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & 0 \\ \gamma_1 & \gamma_2 & \gamma_3 \\ \eta_1 & \eta_2 & \eta_3 \end{pmatrix} \begin{pmatrix} 1 \\ 1_{\{t \geq T\}} \\ t \end{pmatrix} + \begin{pmatrix} \delta_{1t} \\ \delta_{2t} \\ \delta_{3t} \end{pmatrix} \tag{1}$$

We posit a reduced-form relationship whereby the log net-of-tax rate $\log(1-\tau_t)$ on the top 1 percent is the sum of an intercept α_1, a regime-switch component $\alpha_2 1_{\{t \geq T\}}$ and a residual δ_{1t}. The regime-switch component permanently alters the top rate after period T. We set $T = 1987$

[28]Using the notation from Table 1, $Rev_{1t} = N_{1t}Tax_{1t}, Rev_{2t} = N_{2t}Tax_{2t}$ and $Rev_{3t} = N_{1t}Tax_{3t}$, where (N_{1t}, N_{2t}) are the number of tax units with income above and below the 99th percentile at time t and Tax_{1t}, Tax_{2t} and Tax_{3t} are tax revenues per tax unit in the relevant income group.

[29]Mertens and Montiel-Olea calculate the income-weighted average, marginal tax rate based on federal income taxes and social security and medicare taxes. We use this measure because standard tax calculators (e.g. TAXSIM) encode state tax rules starting from the late 1970's.

[30]The central policy question that the literature tries to answer is how much to permanently change the top tax rate. To address this question using reduced-form methods, it is important to have tax rate variation in the data that roughly corresponds to such a permanent change.

so that the regime switch is associated with the Tax Reform Act of 1986.

Log revenue in the statistical model can shift in response to such a permanent tax rate change. $\log Rev_{2t}$ depends on an intercept γ_1, a regime-switch component $\gamma_2 1_{\{t \geq T\}}$, a time trend $\gamma_3 t$ and on unmeasured sources of variation. The statistical model for $\log Rev_{3t}$ follows the same formulation. Ratios of the estimated parameters $(\epsilon_2, \epsilon_3) = (\gamma_2/\alpha_2, \eta_2/\alpha_2)$ are our estimates of the elasticities that enter the tax rate formula.

Table 6 - Revenue Elasticities: US Data 1964-2012

EQUATION	PARAMETER	(1)	(2)	(3)
$\log(1-\tau)$	α_2	0.236	0.243	0.225
		(0.028)	(0.028)	(0.026)
$\log Rev_2$	γ_2	0.068	0.075	0.048
		(0.046)	(0.053)	(0.067)
$\log Rev_3$	η_2	-0.090	0.020	0.050
		(0.157)	(0.148)	(0.195)
Elasticity	$\epsilon_2 = \gamma_2/\alpha_2$	0.288	0.307	0.212
		(0.191)	(0.215)	(0.297)
Elasticity	$\epsilon_3 = \eta_2/\alpha_2$	-0.380	0.081	0.223
		(0.673)	(0.609)	(0.865)
	Time Dummies	None	1986-87	1986-90
	N. Obs.	41	41	41

Note: Parameters are estimated using the exactly-identified GMM estimator. Standard errors (in parentheses) are computed using the Newey-West method with 1 lag and, for the elasticities, the delta method. Appendix B describes data sources and data construction.

While this statistical model captures in a simple and transparent way a shift in balanced-growth paths and a permanent shift in the top tax rate, there are several ways in which it is too simple. First, many tax reforms are announced before they come into force and there are very large anticipatory responses. For example, the Tax Reform Act of 1986 increased the capital gains tax rate in 1987 while decreasing the top ordinary income tax rate in 1987. This led to a dramatic surge in capital gains realizations in 1986 and a corresponding surge in capital gains taxes in 1986 and a collapse in 1987.[31] Second, the statistical model does not allow for transitional dynamics that occur in the human capital model. Both of these issues can be addressed in a simple way by adding time dummines to the system of equations.[32]

Table 6 presents estimation results. Column (1) presents estimates based on the system in equation (1), whereas column (2) and (3) add time dummies to the system. Observations for

[31]Burman, Clausing and O'Hare (1994) document the behavior of capital gains realizations.

[32]Appendix B shows that with short time series of 49 observation, all three model elasticities are underestimated in the statistical model based on the transitional path calculated for the human capital model. Adding time dummies for 1986-1990 (1 year before and 4 years after the reform) and estimating the resulting statistical model reduced the underestimation of elasticities (ϵ_2, ϵ_3) compared to the case of no time dummies or just time dummies for 1986-7.

the years around the Tax Reform Act of 1986 are "dummied out" in column (2) and the years from 1986 to 1990 are dummied out in column (3). The estimate for α_2, governing the shift in the net-of-tax rate, is always positive in Table 6. This accounts for the dramatic drop in the top tax rate in Figure 8. The estimate for the elasticity ϵ_2 is positive in all columns. The estimate for ϵ_3 is negative in column (1) but is positive in columns (2) and (3).

Our view is that the elasticity estimates in columns (2)-(3) of Table 6 are the most relevant. Some means of controlling for the dramatic variation in capital gains that occured in 1986 and 1987 is needed as the leading explanation is that this is due to (preannounced) changes in the capital gains tax rate. Point estimates for ϵ_2 and ϵ_3 are positive based on these columns. Of course, the standard errors are large so that there is no strong evidence for positive elasticities.[33]

The elasticity ϵ_3 can in theory be restated as the revenue-weighted average of the elasticities of its two components: taxes on production paid by the top 1 percent and capital income taxes paid by the top 1 percent. Figure 9 shows that in 1986 capital income taxes paid by the top 1 percent are more than double the 1985 value and that the 1987 value collapses below the 1985 value. Taxes on production paid by the top 1 percent do not display similar fluctuations, even qualitatively, over this time period.

We estimate the elasticity of each of these two components to a movement in the net-of-tax rate. The estimate for the elasticity of taxes on production paid by the top 1 percent to the net-of-tax rate range from a low of 0.205 to a high of 0.447 as is documented in Table B2 in Appendix B. Thus, the negative point estimate for the elasticity ϵ_3 in Table 6, for the case without time dummies, is entirely driven by a negative point estimate for the elasticity of capital income taxes paid by the top 1 percent.

Piketty and Saez (2003) and Saez and Zucman (2016) provide a narrative argument for a positive value for ϵ_3.[34] They document a strong U-shaped pattern for the top 1 percent income and wealth shares in the US over the last hundred years. The narrative is that the decrease in income tax rates for the top 1 percent since the 1980s has fostered growing capital income and wealth concentration. If valid, then the lower top tax rate would seem to be a force behind an increase in consumption taxes and capital income taxes from top earners leading to a positive value for ϵ_3.[35]

[33] A more comprehensive empirical analysis would investigate the importance of adding a list of plausible controls or modeling a shift of a balanced growth path differently. The results in this paper are a first step in such an analysis.

[34] Piketty and Saez (2003, p. 37) state "In the United States, due to the very large rise of top wages since the 1970s, the coupon-clipping rentiers have been overtaken by the working rich. Such a pattern might not last for very long because our proposed interpretation also suggests that the decline of progressive taxation observed since the early 1980s in the United States could very well spur a revival of high wealth concentration and top capital incomes during the next few decades."

[35] Kaymak and Poschke (2016) and Hubmer, Krusell and Smith (2016) employ quantitative-theoretical models

6 Conclusion

From a human capital perspective, we argue that the revenue maximizing top tax rate in the US is approximately 49 percent. Two new forces are the main quantitative drivers for why the model's top rate is well below the 73 percent top tax rate from the established view. The widely-used tax rate formula, which underlies the established view, does not account for these two new forces. Reduced-form evidence supports positive point estimates for the elasticities underlying these new forces, but standard errors are large so that this evidence is currently weak.

The human capital model is consistent with a number of facts that theory and the literature view as being relevant for determining the revenue maximizing top tax rate. For example, the model exactly matches the US Pareto statistic $a_1 = 1.70$ that enters the tax rate formula. The model approximates the Frisch labor hours elasticity for males. The small empirical value for this elasticity has been viewed (e.g. Keane (2011) and Saez et al. (2012)) as suggesting that the revenue maximizing top tax rate will be large. The model is also consistent with a number of non-targeted, data properties. For example, the model produces a short-term income elasticity to the net-of-tax rate within the range of the estimated values in the literature. This is the main evidence underlying the established view. The model produces strikingly large earnings growth rates for top lifetime earners just as in US data. The human capital model produces large growth rates because top lifetime earners are disproportionally good learners whose investments over the lifetime produce steep age-earnings profiles.

to argue that lower top income tax rates since the 1980s account for an important part of the rise in the top 1 percent wealth share since 1980 in the US.

References

Altonji, J. (1986), Intertemporal Substitution in Labor Supply: Evidence from Micro Data, Journal of Political Economy, vol. 94, pp. S176-S215.

Alvaredo, F., Atkinson, A., Piketty, T. and E. Saez, The World Top Incomes Database, http://topincomes.g-mond.parisschoolofeconomics.eu/, accessed10/10/2013.

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

Badel, A., Daly, M., Huggett, M. and M. Nybom (2018), Top Earners: Cross-Country Facts, Federal Reserve Bank of St. Louis Review, vol. 100, pp. 237-57.

Blandin, A. (2016), Reforming the Social Security Earnings Cap: The Role of Endogenous Human Capital, manuscript.

Brewer, M., Saez, E. and A. Shephard (2010), Means testing and tax rates on earnings, in Dimensions of Tax Design: The Mirrlees Review, Editors J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba, Oxford University Press.

Bruggemann, B. (2017), Higher Taxes at the Top: The Role of Entrepreneurs, manuscript.

Burman, L., Clausing, K. and J. O'Hare (1994), Tax Reform and Realizations of Capital Gains in 1986, National Tax Journal, vol. 47, pp. 1-18.

Chetty, R. (2006), A New Method of Estimating Risk Aversion, American Economic Review, vol. 96, pp. 1821- 1834.

Diamond, P. and E. Saez (2011), The Case for a Progressive Tax: From Basic Research to Policy Recommendations, Journal of Economic Perspectives, vol. 25, pp. 165-90.

Domeij, D and M. Floden (2006), The Labor-supply Elasticity and Borrowing Constraints: Why Estimates are Biased, Review of Economic Dynamics vol. 9, pp. 242-262.

Erosa, A. and T. Koreshkova (2007), Progressive Taxation in a Dynastic Model of Human Capital, Journal of Monetary Economics, vol. 54, pp. 667-85.

Feenberg, D. and E. Coutts (1993), An Introduction to the TAXSIM Model, Journal of Policy Analysis and Management, vol. 12, pp. 189-94.

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

Guvenen, F., Kuruscu, B. and S. Ozkan (2014), Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis, Review of Economic Studies, vol. 81, pp. 818-50.

Guvenen, F., Ozkan, S. and J. Song (2014), The Nature of Countercyclical Income Risk, Journal of Political Economy, vol. 122, pp. 621-60.

Guvenen, F., Karahan, F., Ozkan, S. and J. Song (2015), What Do Data on Millions of US Workers Say About Labor Income Risk?, manuscript.

Heathcote, J., Perri F. and G. Violante (2010), Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006, Review of Economic Dynamics, vol. 13, pp. 15-51.

Heathcote, J., Storesletten, K. and G. Violante (2009), Quantitative Macroeconomics with Heterogeneous Households, Annual Review of Economics, vol. 1, pp.319-54.

Heathcote, J., Storesletten, K. and G. Violante (2017), Optimal Tax Progressivity: An Analytical Framework, Quarterly Journal of Economics, vol. 132, pp. 1693-1754.

Heckman, J, Lochner, L. and C. Taber (1998), Tax Policy and Human Capital Formation, American Economic Review, Papers and Proceedings, vol. 88, pp. 293-7.

Hubmer, J., Krusell, P. and A. Smith, Jr.(2016), The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation, NBER Working Paper 23011.

Huggett, M., Ventura, G. and A. Yaron (2011), Sources of Lifetime Inequality, American Economic Review, vol. 101, pp. 2923-54.

Kapicka, M. (2015) Optimal Mirrleesian Taxation in a Ben-Porath Economy, American Economic Journal: Macroeconomics, vol. 7, pp. 219-48.

Kaymak, B. and M. Poschke (2016), The Evolution of Weath Inequality Over Half a Century: The Role of Taxes, Transfers and Technology, Journal of Monetary Economics, vol. 77, pp. 1-25.

Keane, M. (2011), Labor Supply and Taxes: A Survey, Journal of Economic Literature, vol. 49(4), pp. 961-1075.

Keane, M. and R. Rogerson (2012), Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom, Journal of Economic Literature, vol. 50, pp. 464-76.

Kindermann, F. and D. Krueger (2017), High Marginal Tax Rates on the Top 1% ? Lessons from a Life Cycle Model with Idiosyncratic Income Risk, manuscript.

Krueger, D. and A. Ludwig (2016), On the Optimal Provision of Social Insurance: Progressive Taxation Versus Education Subsidies, Journal of Monetary Economics, vol. 78, pp. 72-98.

MaCurdy, T (1981), An Empirical Model of Labor Supply in a Life-Cycle Setting, Journal of Political Economy, vol. 89, pp. 1059- 85.

Mertens, K. and J. Montiel-Olea (2017), Marginal Tax Rates and Income: New Time Series Evidence, manuscript.

Mirrlees, J. (1971), An Exploration into the Theory of Optimum Income Taxation, Review of Economic Studies, vol. 38, pp. 175- 208.

Piketty, T. and E. Saez (2003), Income Inequality in the United States, 1913-1998, Quarterly Journal of Economics, vol. 118, pp. 1-39.

Piketty, T. and E. Saez (2013), Optimal Labor Income Taxation, Handbook of Public Economic, (eds) Atkinson and Bourguignon.

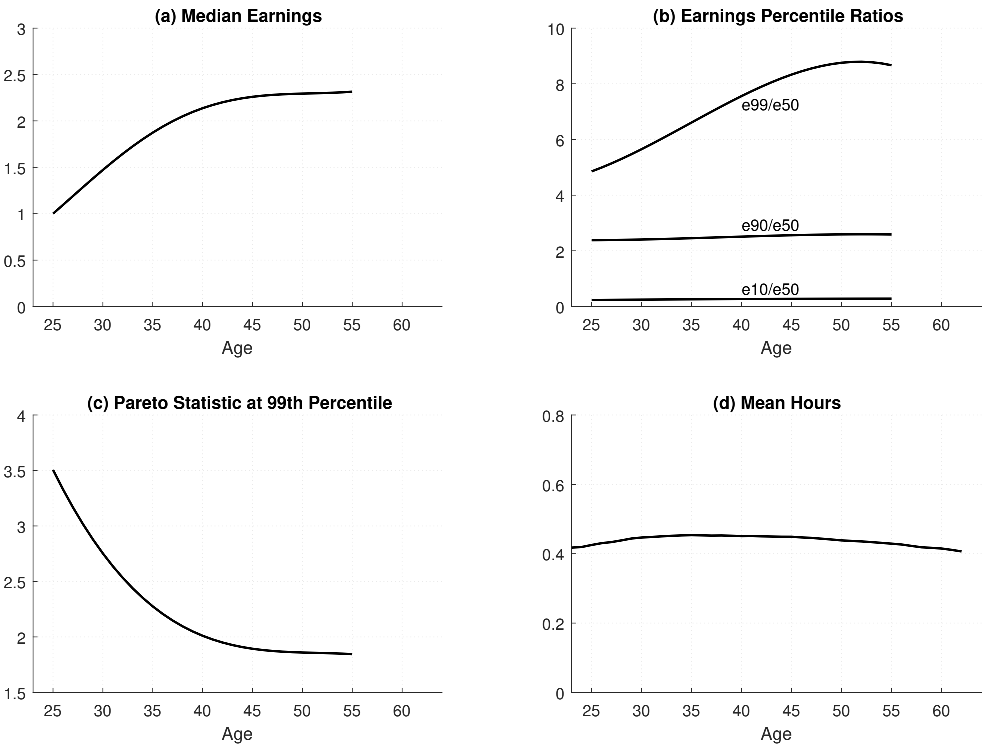
Saez, E. (2001), Using Elasticities to Derive Optimal Income Tax Rates, Review of Economic Studies, vol. 68, pp. 205-29.

Saez, E., Slemrod, J. and S. Giertz (2012), The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review, Journal of Economic Literature, vol. 50(1), pp. 3-50.

Saez, E. and G. Zucman (2016), Wealth Inequality in the United States Since 1913: Evidence from Capitalized Income Tax Data, Quarterly Journal of Economics, vol. 131, pp. 519-78.

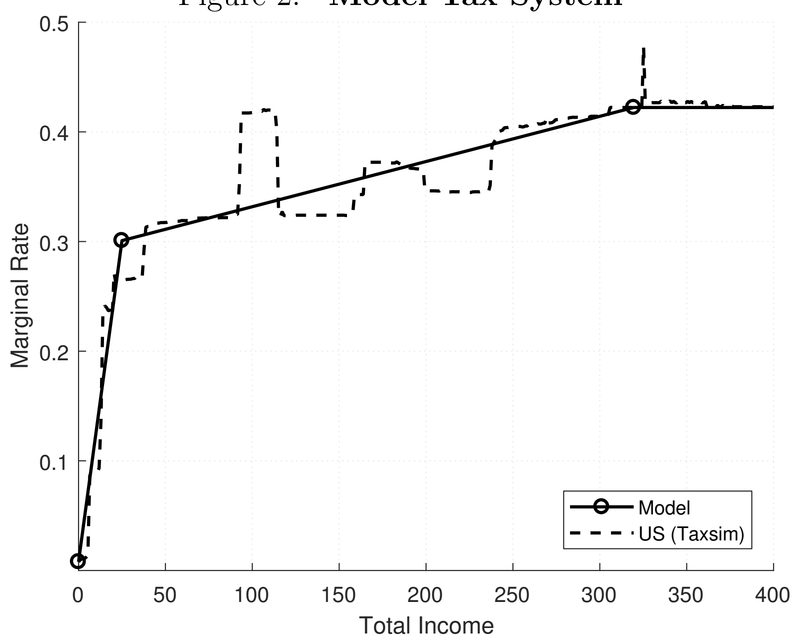
Stantcheva, S. (2017), Optimal Taxation and Human Capital Policies over the Lifetime, Journal of Political Economy, vol. 125, pp. 1931-90.

Figure 1: **Life-Cycle Profiles: Earnings and Hours**



Note: Earnings profiles are based on SSA data. Hours profiles are based on PSID data.

Figure 2: **Model Tax System**



Note: The horizontal axis measures income in thousands of dollars.

Figure 3: **Life-Cycle Profiles: Data and Model**

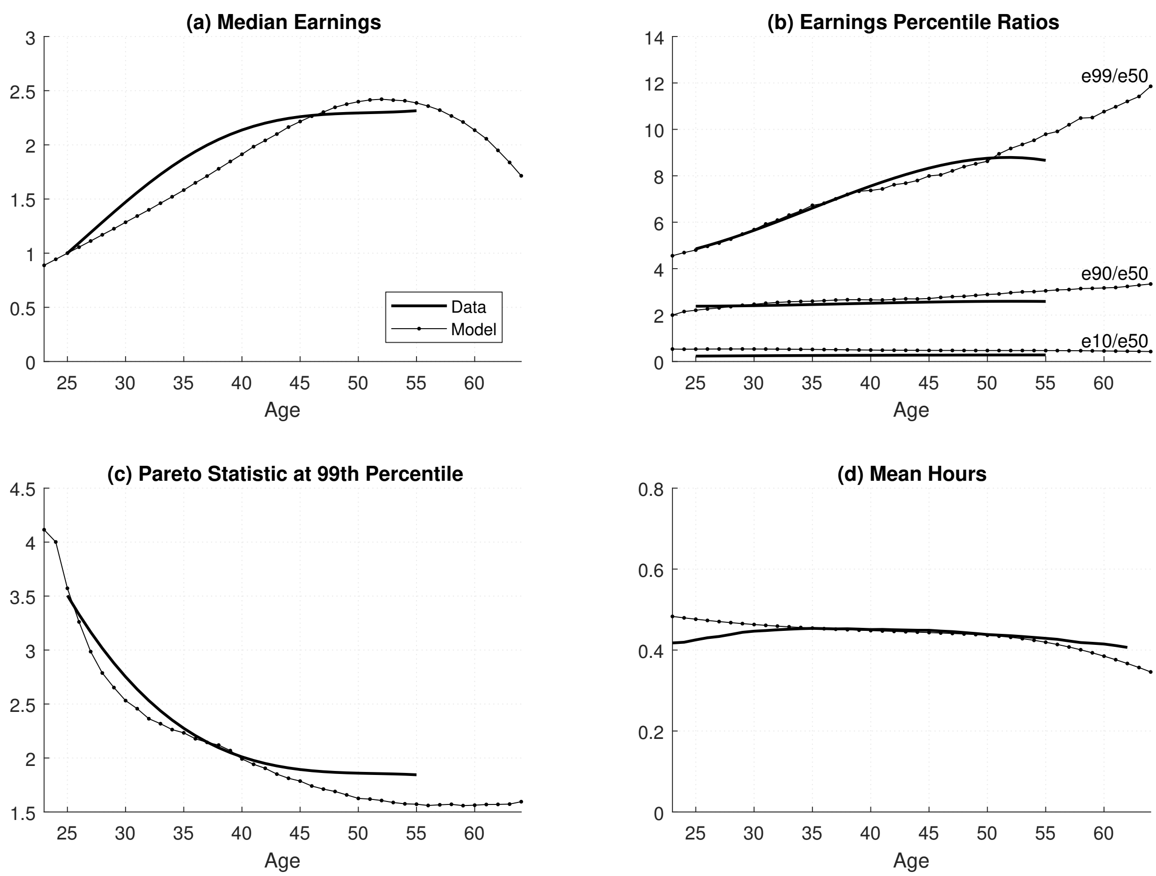
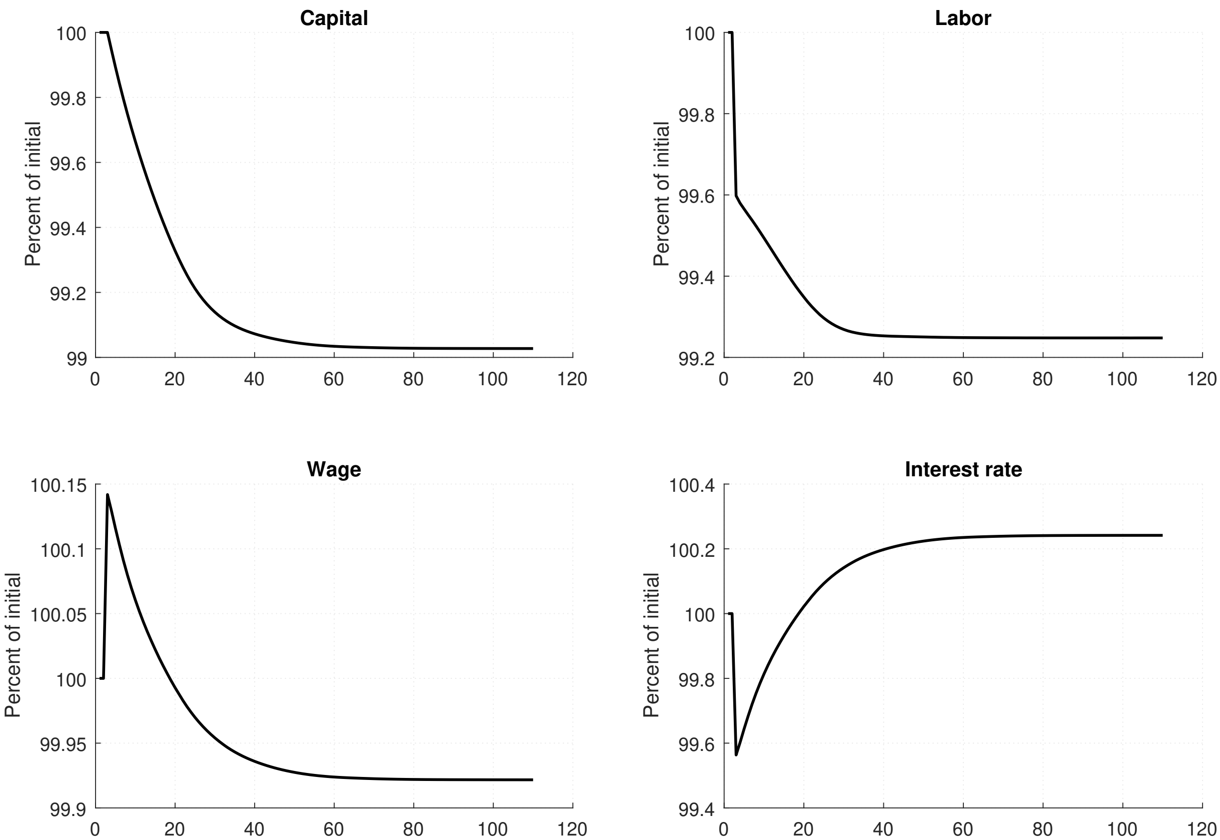
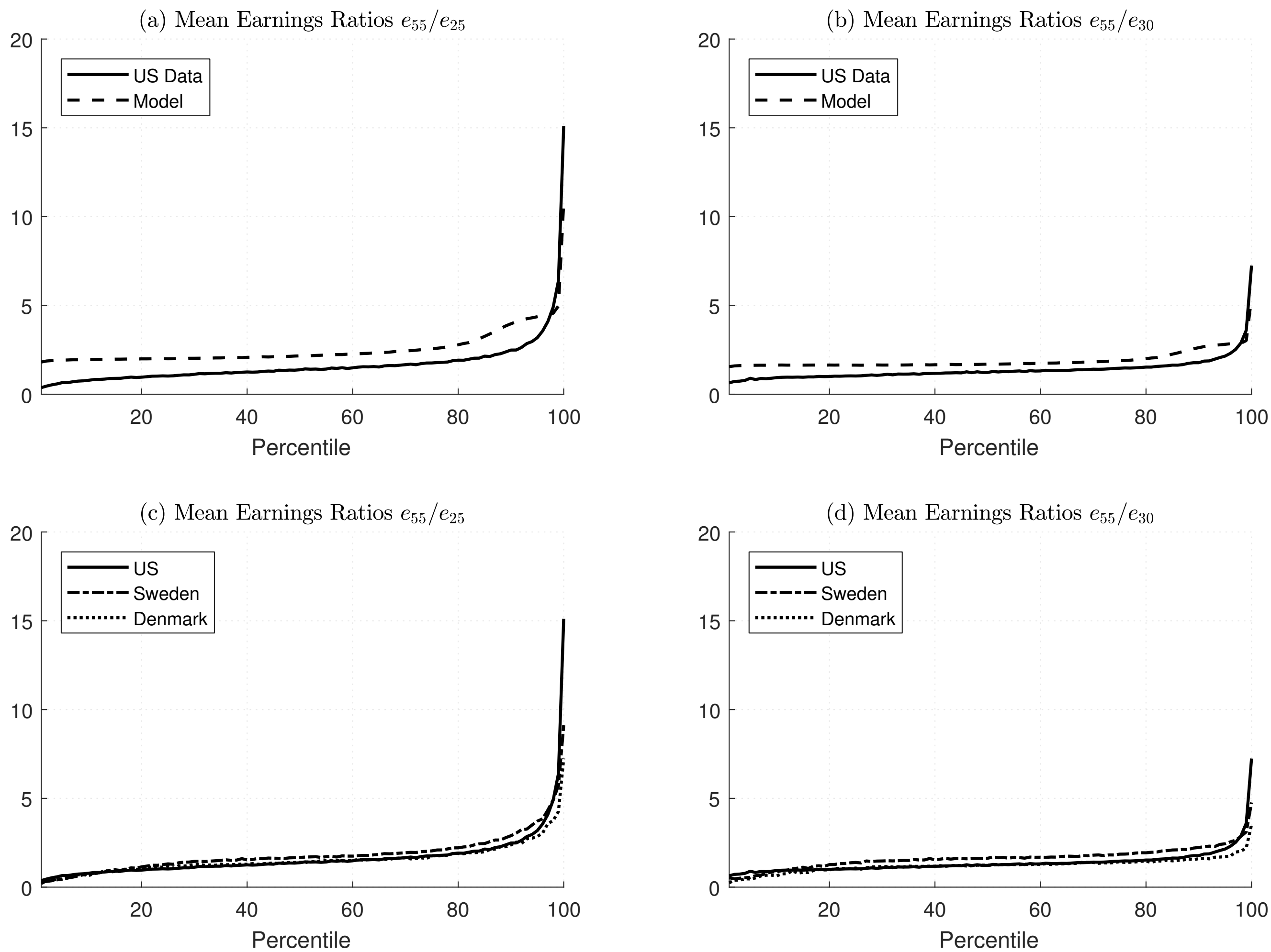


Figure 4: **Model Transition Path**



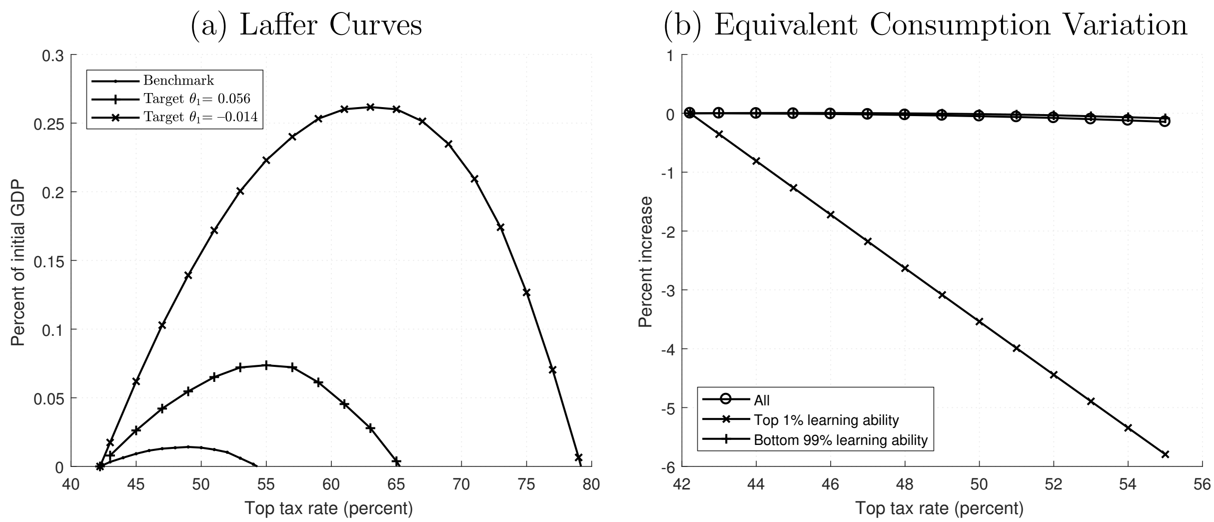
Note: All variables are normalized to 100 in period 1. The tax reform occurs in model period 3 and permanently changes the tax rate from $\tau = 0.422$ to $\tau = 0.49$. Capital and labor are per agent.

Figure 5: **Earnings Ratios By Lifetime Earnings Percentiles**



Note: The figure plots mean earnings ratios for each bin of lifetime earnings. US facts are from Guvenen et al. (2015). Facts for Denmark and Sweden are from Badel, Daly, Huggett and Nybom (2018).

Figure 6: **Laffer Curves and Equivalent Consumption Variation**



Note: The Laffer curve plots the lump-sum transfer per agent stated as a percent of initial output per agent. The Laffer curve is plotted for the benchmark model and for models with different targeted values for the regression coefficient θ_1. Consumption equivalents are calculated for age 1 agents in the benchmark model.

Figure 7: **Laffer Curves**

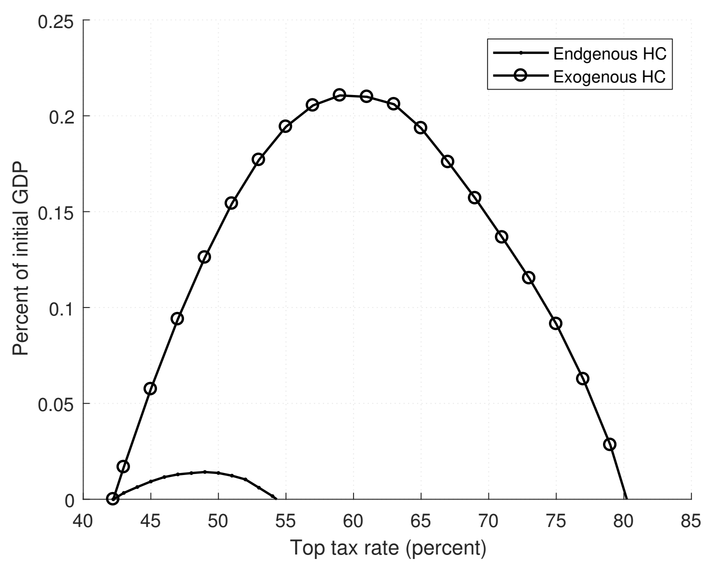
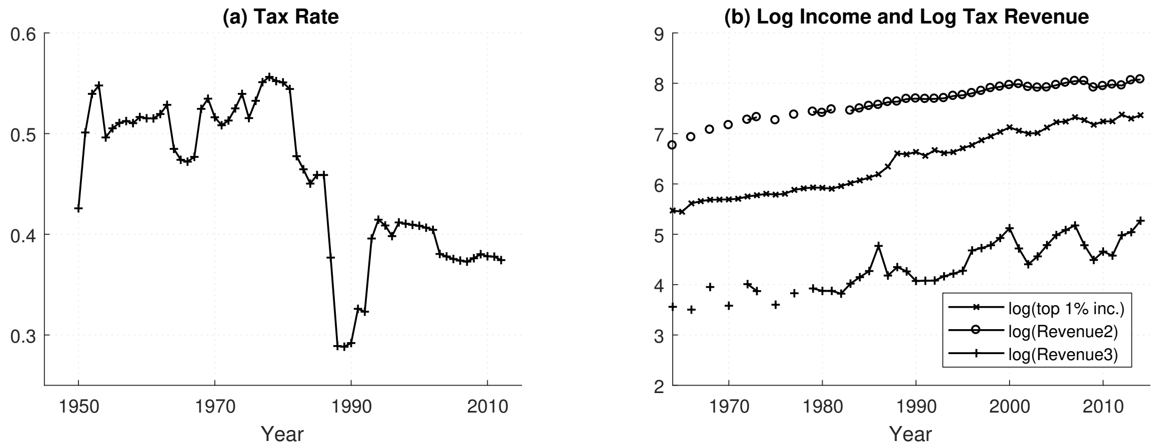
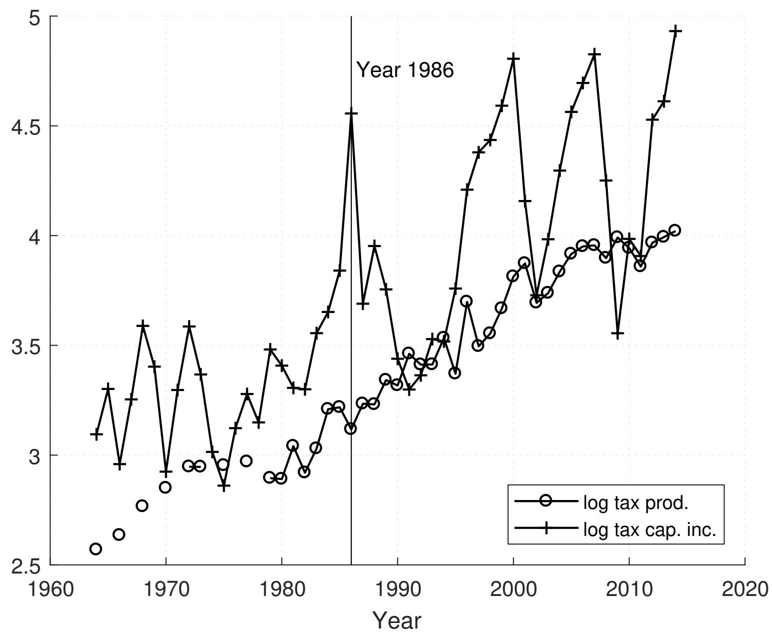


Figure 8: **Top 1 Percent Tax Rate, Income and Tax Revenue**



Note: Tax rate is the average marginal tax rate on the top 1 percent from Mertens and Montiel-Olea (2017). $Revenue_2$ is based on all the enumerated taxes paid by tax units below the 99th percentile. $Revenue_3$ is based on the taxes on production and the capital income taxes paid by the top 1 percent of tax units. Data and methods are described in Appendix B.

Figure 9: **Log Tax Revenues from the Top 1 Percent**



Note: Revenue3 is decomposed into Tax on Production and Tax on Capital Income. Data and methods for constructing these series are described in Appendix B.

A Appendix for "Taxing Top Earners: A Human Capital Perspective", Authors: Badel, Huggett and Luo

A.1 Earnings and Hours Data

SSA Data: We use Social Security Administration (SSA) earnings data from Guvenen, Ozkan and Song (2014). We use age-year tabulations of the 10, 25, 50, 75, 90, 95 and 99th earnings percentile for males age $j \in \{25, 35, 45, 55\}$ in year $t \in \{1978, 1979, ..., 2011\}$. These tabulations are based on a 10 percent random sample of males from the Master Earnings File (MEF). The MEF contains all earnings data collected by SSA based on W-2 forms. Earnings data are not top coded and include wages and salaries, bonuses and exercised stock options as reported on the W-2 form (Box 1). The earnings data is converted into real units. See Guvenen et al. (2014) for details.

We construct the Pareto statistic at the 99th earnings percentile for age j and year t as follows. We assume that the earnings distribution follows a Pareto distribution beyond the 99th percentile for age j and year t. We construct the parameters describing this distribution via the method of moments and the data values for the 95th and 99th earnings percentiles (e_{95}, e_{99}) for a given age and year. The c.d.f. of a Pareto distribution is $F(e; \alpha, \lambda) = 1 - (\frac{e}{\alpha})^{-\lambda}$. We solve the system $.95 = F(e_{95}; \alpha, \lambda)$ and $.99 = F(e_{99}; \alpha, \lambda)$. This implies $\lambda = \frac{\log .05 - \log .01}{\log e_{99} - \log e_{95}}$. To construct the Pareto statistic at the 99th percentile for age j and year t, it remains to calculate the conditional mean $E[e|e \geq e_{99}] = \frac{\lambda e_{99}}{\lambda - 1}$ that is implied by the Pareto distribution.

PSID Data: We use Panel Study of Income Dynamics (PSID) data provided by Heathcote, Perri and Violante (2010), HPV hereafter. The data comes from the PSID 1967 to 1996 annual surveys and from the 1999 to 2003 biennial surveys.

Sample Selection: We keep only data on male heads of household reporting to have worked at least 260 hours during the last year with non-missing records for labor earnings. In order to minimize measurement error, we delete records with positive labor income and zero hours of work or an hourly wage less than half of the federal minimum in the reporting year.

Variable Definitions: The annual earnings variable provided by HPV includes all income from wages, salaries, commissions, bonuses, overtime and the labor part of self-employment income. Annual hours of work is defined as the sum total of hours worked during the previous year on the main job, on extra jobs and overtime hours. This variable is computed using information on usual hours worked per week times the number of actual weeks worked in the last year.

Age-Year Cells: We split the dataset into age-year cells, compute the relevant moment within each cell. We put a PSID observation in the (a, y) cell if the interview was conducted during year y and the head of household's age in year y was in the interval $[a - 2, a + 2]$, where ages range from $a = 23$ to $a = 62$. The life-cycle profiles we calculate correspond to $(\beta_{23} + d, \beta_{24} + d, \beta_{25} + d, ..., \beta_{62} + d)$, where the β_a are the estimated age coefficients and d is a vertical displacement selected in the manner described in section 3.

A.2 Tax Formula Coefficients

Step 1: Calculate $(\underline{y}, \bar{y})$ in Panel (d) of Table 1, using the income measure in section 3.3.

1. SOI Table 1.4 tabulates income, by type of income, for tax units sorted by adjusted gross income (AGI) bins.[36] In 2010, 250 and 500 thousand dollars are the $p_1 = .9823$ and $p_2 = .9947$ percentiles of AGI based on the number of potential tax units in 2010 reported in Piketty and Saez (2003 update, Table A0).

2. Assume that tax units in the $[250, 500)$ and $[500, \infty)$ AGI bins (in thousands of dollars) are also the tax units that would fall in the two corresponding bins based on the same percentiles for our definition of income. Assume that income is distributed Pareto beyond the p_1 percentile of income. The Pareto cdf is $F(y) = 1 - (\alpha/y)^\lambda$. Conditional means satisfy $E[y|y > y_i] = y_i(\lambda/(\lambda - 1))$.

3. Denote (y_1, y_2) the (p_1, p_2) percentiles of our income measure and $(\bar{y}_1, \bar{y}_2)$ the respective means, conditional on income exceeding (y_1, y_2). Calculate $(\bar{y}_1, \bar{y}_2)$ based on tabulated income types in the $[250, \infty)$ and $[500, \infty)$ AGI bins. Solve the equations below to get $(\alpha, \lambda, y_1, y_2)$:

$$1 - p_1 = (\alpha/y_1)^\lambda,\ 1 - p_2 = (\alpha/y_2)^\lambda,\ \bar{y}_1 = y_1(\lambda/(\lambda - 1)),\ \bar{y}_2 = y_2(\lambda/(\lambda - 1))$$

4. Solve $(\underline{y}, \bar{y})$ using: $.99 = 1 - (\alpha/\underline{y})^\lambda$ *and* $\bar{y} = \underline{y}(\lambda/(\lambda - 1))$.

Step 2: Calculate capital income tax in Panel (c) of Table 1.

1. $capital\ income\ tax = \tau_k \times Capital\ Income$. We calculate $\tau_k = 0.20$ in section 4.

2. $Capital\ Income$ equals qualified dividends plus capital gains for tax units in the top percentile. Capital gains equal the measure for the top percentile in Piketty and Saez (2003, update) based on Table A0, A4 and A6. Qualified dividends equal those in the $[500, \infty)$ AGI bin from SOI Table 1.4 and a fraction ϕ of those in the $[250, 500)$ AGI bin. ϕ is the fraction of income in the $[250, 500)$ bin due to units with income above the 99th percentile.

$$\phi = \frac{(1 - .99)\bar{y} - (1 - p_2)\bar{y}_2}{(1 - p_1)\bar{y}_1 - (1 - p_2)\bar{y}_2}$$

Step 3: Calculate entries (1)-(3) in Panel (b) of Table 1.

1. Personal taxes are the sum of federal, state and local income taxes. Federal taxes equal the federal income tax in the $[500, \infty)$ bin in SOI Table 3.3 plus a fraction ϕ of this tax in the $[250, 500)$ bin. State and local taxes equal the state and local taxes in the $[500, \infty)$ bin in SOI Table 2.1 plus a fraction ϕ of the tax in the $[250, 500)$ bin. State and local income taxes are based on tax units that use itemized deductions. A high fraction of the tax units in the $[250, \infty)$ bin are itemizers.

2. Social insurance taxes are the sum of self-employment taxes and the taxes on the employed. Self-employment taxes are the sum of self-employment taxes in the $[500, \infty)$ bin in SOI Table 3.3 plus a fraction ϕ of this tax in the $[250, 500)$ bin. Taxes on the employed are in two parts. First, for units in the $[500, \infty)$ bin taxes equal $number \times 13243 + 0.029 \times wages$, where $number$ is the number of units in the bin with wage and salary income from SOI Table 1.4, 13243 is the maximum OASDI tax, 0.029 is the combined medicare tax rate and *wages* is the bin total of reported wage and salary income from SOI Table 1.4. Second, we make the same calculation for the $[250, 500)$ bin and multiply by ϕ.

[36] SOI Tables: https://www.irs.gov/uac/soi-tax-stats-individual-income-tax-returns-publication-1304-complete-report

3. Taxes on production paid by the top 1 percent equal real estate taxes paid by the top 1 percent plus an imputed measure of residual taxes paid by the top 1 percent. Real estate taxes equal the real estate tax in the $[500, \infty)$ bin in SOI Table 2.1 plus a fraction ϕ of this tax in the $[250, 500)$ bin. Residual taxes paid by the top 1 percent in $t = 2010$ equal $Residual_t \times Consumption\ Share_t$.

   Residual taxes at time $t = 2010$ are: $Residual_t = Taxes\ on\ Production_t - Residential\ Real\ Estate\ Tax_t$. Residential Real Estate Tax equals property taxes, from BEA Table 3.5 line 30, times the average share of residential structures investment in private structures investment from BEA Table 5.4.5. $Consumption\ Share_t$ equals non-housing, consumption expenditures of households with income in the top 1 percent as a ratio to non-housing consumption expenditures of all households. We use CEX data in 2010 to make this calculation.

Step 4: Compute the entries in Panel (a) of Table 1. Personal Taxes are BEA Table 3.1 line 3. Social Insurance Taxes are BEA Table 3.6 line 4 plus line 22. Social insurance taxes are based only on the employer and employee parts of social security and medicare taxes. Taxes on Production are BEA Table 3.1 line 4. The number N is the number of potential tax units in 2010 from Piketty and Saez (2003 Update, Table A0).

B APPENDIX

This appendix provides detailed information on (1) construction of the data series graphed in Figure 8 and 9, (2) Pareto statistics in US data, (3) robustness excercises related to Figure 1, (4) the construction of Pareto-Lognormal distributions, (5) computation of model equilibrium, (6) the labor hours regressions from section 4, (7) estimation of the three elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ based on US data and on model data, (8) training time in model and NLSY data. References related to these calculations are provided at the end of the appendix.

B.1 Decomposing Aggregate Taxes: Data Construction

Construct a time series for aggregate taxes, labeled Tax_t, over 1964-2014. Use the data and methods described in Step 4 of Appendix A.2. Decompose $Tax_t = (Rev_{1t} + Rev_{3t}) + Rev_{2t}$ into three components. $Rev_{1t} + Rev_{3t}$ are all the taxes in Tax_t paid by the top 1 percent of tax units, whereas Rev_{2t} are taxes paid by tax units below the top 1 percent. Using the notation from section 3, $Rev_{1t} = N_{1t}Tax_{1t}, Rev_{2t} = N_{2t}Tax_{2t}, Rev_{3t} = N_{1t}Tax_{3t}$. The decomposition is calculated in three steps.

Step 1: Calculate (1) federal income taxes, (2) state and local income taxes, (3) social insurance taxes and (4) taxes on production paid by the top 1 percent tax units. Calculate $Rev_{1t} + Rev_{3t}$ for a given year over 1964-2014 as the sum of these four measures by applying step 1.1 to 1.8.

1.1 Define income as the sum of the following Statistics of Income (SOI) categories: (i) wages and salaries, (ii) interest, (iii) dividends, (iv) business income, (v) IRA distributions, (vi) pensions and annuities, (vii) total rent and royalty, (viii) partnership and S-corporation income and (ix) estate and trust income.

1.2 Find the AGI bin in SOI Table 1.4 that brackets the top 1 percent of tax units based on AGI. Calculate mean income per tax unit $\bar{y}_1$ among tax units within this AGI bin and higher AGI bins using the income definition in step 1.1. Calculate mean income per tax unit $\bar{y}_2$ based on the income information in next highest AGI bin and higher bins. Calculate (p_1, p_2) as the percentile threshold for the two AGI bins among potential tax units. For example, in 2010 the [250, 500k] AGI bin brackets the top 1 percent of AGI among all potential tax units and $(p_1, p_2) = (.9823, .9947)$. This is the procedure in Step 1 from Appendix A.2.

1.3 Infer $(\underline{y}, \bar{y})$, the 99th percentile of income and the mean income per tax unit beyond this percentile. Use the assumption that income is distributed Pareto and use $(\bar{y}_1, \bar{y}_2, p_1, p_2)$. Follow Step 1 in Appendix A.2.

1.4 Calculate the fraction ϕ of income, defined in step 1.1, in the bracketing AGI bin that is due to tax units with income beyond the 99th percentile $\underline{y}$. Follow Step 2 in Appendix A.2 and use $(\bar{y}_1, \bar{y}_2, \bar{y}, p_1, p_2)$.

1.5 Calculate federal income taxes paid by tax units with income in the top 1 percent as the sum of all the federal income taxes paid by tax units above the bracketing AGI bin plus ϕ times the federal income taxes paid by tax units in the bracketing AGI bin. Use SOI Table 3.3 to calculate federal taxes.

1.6 Calculate state and local income taxes paid by the top 1 percent of tax units by repeating the procedure in step 1.5. Use SOI Table 2.1 to do this.

1.7 Social insurance taxes paid by the top 1 percent of tax units are calculated as self-employment taxes plus imputed social insurance taxes from the employed. Calculate self-employment taxes directly from those reported in SOI Table 3.3 following the procedure in step 1.5. Social insurance taxes from the employed are calculated in two parts. First, for all AGI bins above the bracketing bin, social insurance taxes are the number of tax units in these AGI bins times the maximum taxable earnings for OASDI taxes times the combined employee-employer OASDI tax rate plus the same calculation for HI taxes. Second, make the same calculation for the bracketing bin and multiply by the fraction ϕ. In years without a maximum taxable earnings for HI taxes, follow the calculation in Step 3 of Appendix A.2. Tax rates and maximum taxable earnings in different years are taken from Tax Policy Center.[37]

1.8 Taxes on production paid by the top 1 percent equal the sum of two components. First, calculate residential real estate taxes paid by the top 1 percent directly by repeating the procedure in step 1.5. Use SOI Table 2.1 for real estate taxes paid by itemizers. Second, calculate residual taxes on production paid by the top 1 percent at time t as $Residual_t \times Consumption\ Share_t$. Step 3 from Appendix A.2 explains the construction of $(Residual_t, Consumption\ Share_t)$. We calculate the top 1 percent consumption share directly from CEX data for the period 1980-2012. We set $Consumption\ Share_t$ to the 1980 value before 1980 and to the 2012 value after 2012. As a check, we also calculate $Consumption\ Share_t$ using household income and imputed consumption from the 1967-2010 PSID data constructed by Attanasio and Pistaferri (2014). The consumption share is quite flat in this data set from 1967 to 1980 and is always below the value we calculate in CEX data in 1980.

Step 2: Calculate $Rev_{2t} = Tax_t - (Rev_{1t} + Rev_{3t})$ using Step 1. Calculate Rev_{3t} as taxes on production paid by the top 1 percent tax units from Step 1 plus a measure of the capital income taxes paid by the top 1 percent tax units: $capital\ income\ taxes_t = Capital\ Income_t \times capital\ tax\ rate_t$. Calculate Rev_{1t} as federal, state and local income taxes and social insurance taxes from Step 1 less capital income taxes.

Capital income is measured as capital gains plus qualified dividends. Capital gains is the measure in the top income percentile from Piketty and Saez (2003, update), based on the information in their Tables A0, A4 and A6. Qualified dividends are calculated based on the method in step 1.5 using SOI Table 1.4. The capital income tax rate series used is the maximum federal tax rate on capital gains for 1954-2014 calculated by the US Treasury.[38]

Step 3: Express $(Rev_{1t}, Rev_{2t}, Rev_{3t})$ in 2014 dollars using the CPI-U series. Express aggregate top 1 percent income $Income_t$ in 2014 dollars using the CPI-U series, the top 1 percent income from step 1.3 and the number of top 1 percent tax units.

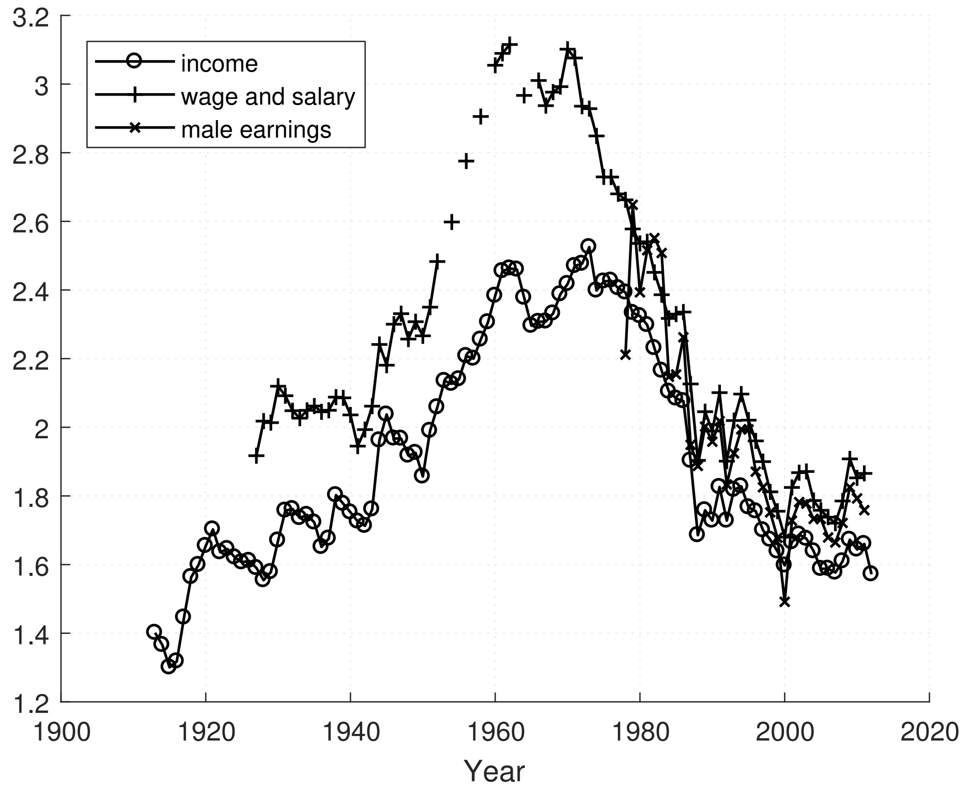
Notes on Step 1-3: First, the definition of income used in step 1 above differs slightly from that in Table 1 as dividends are total dividends rather than non-qualified dividends. Second, SOI table numbering is systematic after roughly 1996. Table numbering above refers to these later years.[39] Third, some of the information needed to carry out steps 1.1 to 1.8 is not available in some years. Real estate tax and state and local income tax information is not present in SOI tables in 1965, 1967, 1969, 1971, 1974, 1976, 1978 and self-employment tax is not present in 1982. The decomposition $(Rev_{1t}, Rev_{2t}, Rev_{3t})$ is not calculated in those years. Fourth, Table 1 Panel (c) calculates capital income taxes paid by the top 1 percent in 2010 using the combined federal and state capital gains tax rate calculated using TAXSIM. The calculation above is based only on the maximum federal capital gains tax rate series as TAXSIM only incorporates state taxes from the late 1970s.

[37] http://www.taxpolicycenter.org/statistics/payroll-tax-rates.

[38] https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/Taxes-Paid-on-Capital-Gains-for-Returns-with-Positive-Net-Capital-Gains.pdf

[39] SOI Tables prior to 1996 are available at https://www.irs.gov/uac/soi-tax-stats-archive-1954-to-1999-individual-income-tax-return-reports

Figure B1: **Pareto Statistic at 99th Percentile: US Measures**



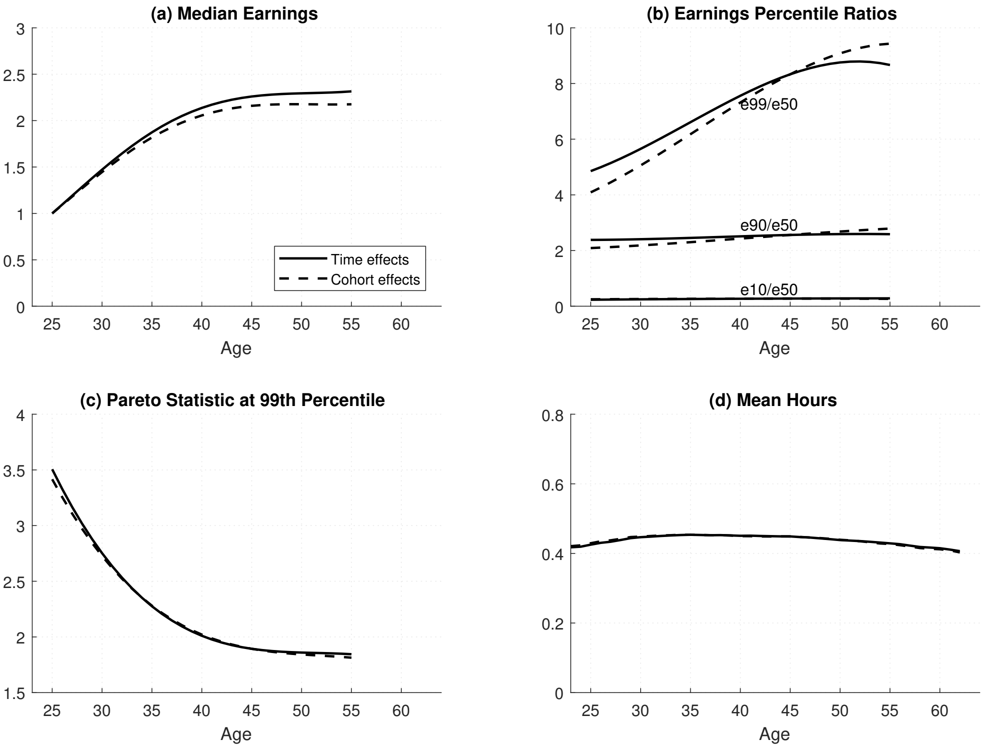
Note: Income measure is without capital gains and comes from the World Top Incomes Data Base. The wage and salary measure is for tax units using data from Piketty and Saez (2003, Update). The earnings measure is for males based on data from Guvenen, Ozkan and Song (2014).

B.2 Pareto Statistic in US Data

Figure B1 plots the Pareto statistic at the 99th percentile in US data. We plot the statistic based on (1) an income definition that excludes capital gains, (2) wage and salary income and (3) earnings for males age 25-60. The measures in (1)-(2) are based on tax units whereas (3) is based on individual males. Measure (1) comes from The World Top Incomes Database, measure (2) comes from Piketty and Saez (2003, update). Measure (3) is based on SSA data tabulated by Guvenen et al. (2014) and is constructed by assuming that earnings beyond the 99-th percentile is distributed Pareto and by using the reported 99-th and 99.999-th percentiles of earnings to infer the Pareto coefficient beyond the 99-th percentile.

The Pareto statistic for income is below that for wage and salary income in all years of the data sets. The Pareto statistic for male earnings is close to the Pareto statistic for wage and salary income in all years for which both statistics can be calculated. In the year 2010, the statistic is 1.64 for income excluding capital gains, 1.79 for male earnings and 1.85 for wage and salary. The Pareto statistic at the 99th percentile is 1.70 in 2010 in Table 1 for an income measure that excludes capital gains and qualified dividends, whereas it is 1.50 in 2010 using AGI as an income measure. The AGI calculation uses SOI Table 1.4, uses the $[250k, \infty)$ and $[500k, \infty)$ AGI bins to calculate average AGI within these bins and uses an interpolation based on the Pareto distribution. Intuitively, the AGI-based Pareto statistic is lower in 2010 than the measures that exclude capital gains or that

Figure B2: **Life-Cycle Profiles: Time and Cohort Effects**



Note: Earnings profiles are based on SSA data. Hours profiles are based on PSID data.

exclude capital gains and qualified dividends because AGI includes capital gains and qualified dividends and these capital income sources are quite concentrated in the upper tail.

B.3 US Profiles: Time Effects vs. Cohort Effects

Figure B2 plots the results of regressing various earnings and work time statistics on a third-order polynomial in age and either time dummy variables or cohort dummy variables. It shows that the results based on time and cohort effects are qualitatively similar. Cohort effects regressions imply a greater growth rate of the 99-50 ratio over the working lifetime. Figure 1 from section 3 of the paper plotted the time effects results. Figure B3 plots model properties for the economies considered in Figure 6.

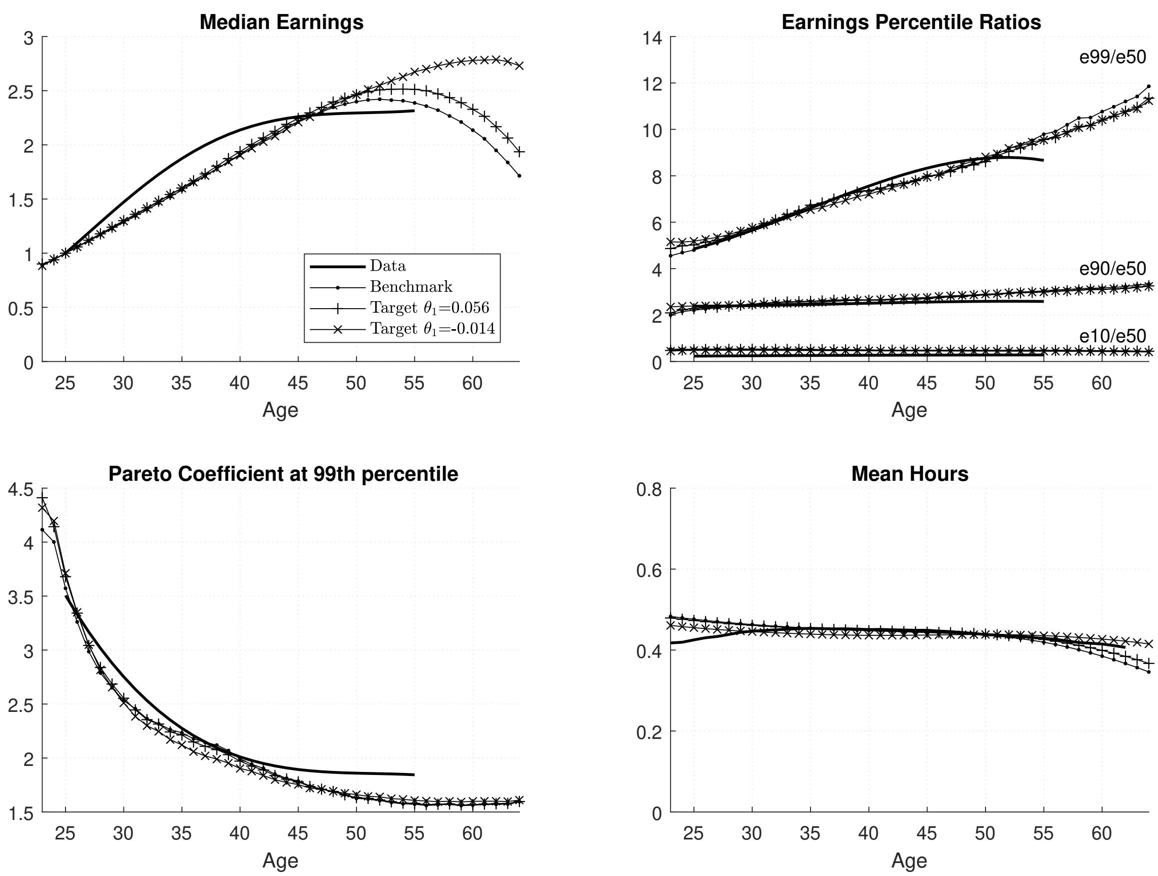
B.4 Initial Conditions

We construct a bivariate distribution based on assumptions A1-2 below. Theorem 1 implies that initial human capital follows a Pareto-Lognormal distribution and, thus, has a Pareto tail.

A1: Let learning ability a be distributed according to a Right-Tail Pareto-Lognormal distribution $PLN(\mu_a, \sigma_a^2, \lambda_a)$. Let ε be independently distributed and lognormal $LN(0, \sigma_\varepsilon^2)$.

A2: $\log h_1 = \beta_0 + \beta_1 \log a + \log \varepsilon$ and $\beta_1 > 0$.

Figure B3: **Model and Data: Different Frisch Elasticity Targets**



Theorem 1: *Assume A1-2. Then* h_1 *is distributed* $PLN(\beta_0 + \beta_1\mu_a, \beta_1^2\sigma_a^2 + \sigma_\varepsilon^2, \lambda_a/\beta_1)$.

Proof: By definition of the PLN distribution, $a \sim PLN(\mu_a, \sigma_a^2, \lambda_a)$ can be expressed as $a = xy$, where $x \sim LN(\mu_a, \sigma_a^2)$ and y is distributed Type-1 Pareto$(1, \lambda_a)$. Substitute this identity into assumption A2 and rearrange.

$$\begin{aligned} \log h_1 &= \beta_0 + \beta_1 \log x + \log \varepsilon + \beta_1 \log y \\ h_1 &= \exp(\beta_0 + \beta_1 \log x + \log \varepsilon) y^{\beta_1} \end{aligned}$$

The first term on the right hand side is distributed $LN(\beta_0 + \beta_1\mu_a, \beta_1^2\sigma_a^2 + \sigma_\varepsilon^2)$. By definition of the Type-1 Pareto distribution, for $y_0 \geq 1$ we have $\text{Prob}(y \leq y_0) = 1 - y_0^{-\lambda_a}$. Let $z \equiv y^{\beta_1}$.

$$\text{Prob}(z \leq z_0) = \text{Prob}(y^{\beta_1} \leq z_0) = \text{Prob}(y \leq z_0^{\frac{1}{\beta_1}}) = 1 - z_0^{-\frac{\lambda_a}{\beta_1}}$$

The second term on the right hand side is distributed Type-1 Pareto with scale parameter 1 and shape parameter λ_a/β_1. $\|$

We discretize this bivariate distribution. Let $\psi(\hat{x}_{ij}) = Prob_i Prob_{ij}$ denote the probability of $\hat{x}_{ij} \in X_1^{grid}$. $Prob_i$ and $Prob_{ij}$ are the probabilities of learning ability a_i and human capital h_{ij}, conditional on ability a_i.

1. $X_1^{grid} = \{(a_i, h_{ij}) : a_i \in \{a_1, ..., a_I\}, \log h_{ij} = \beta_1 + \beta_1 \log a_i + \epsilon_j, \epsilon_j \in \{\epsilon_1, ..., \epsilon_J\}\}$
2. $\{\epsilon_1, ..., \epsilon_J\}$ is an equi-spaced $J = 21$ point Tauchen discretization going from -3 to 3 standard deviations σ_ϵ. $Prob_{ij}$ are probabilities induced by this discretization and by $a_i \in \{a_1, ..., a_I\}$.
3. $\{a_1, ..., a_I\}$ is an $I = 9$ point ability grid, where $a_i = E[a | a_{P_i} \leq a < a_{P_{i+1}}]$, P_i are percentiles and a_{P_i} is the P_i-th percentile of the ability distribution. Expectation is taken using $PLN(\mu_a, \sigma_a^2, \lambda_a)$. Percentiles are chosen to be $(0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.99, 0.995, 0.999, 1.0)$.

B.5 Computation

The algorithm to compute a steady-state equilibrium for the model with top tax rate $\bar{\tau}$, given all model parameters, is outlined below.

Main Algorithm:

1. Given $\bar{\tau}$, guess $(K/L, Tr)$. Calculate $w = F_2(K/L, 1)$ and $r = F_1(K/L, 1) - \delta$.
2. Solve problem DP-1 at grid points $x = (k, h) \in X_j^{grid}(a)$.

   (DP-1) $v_j(x, a) = \max_{(c,l,s,k')} u(c, l + s) + \beta E[v_{j+1}(k', h', a)]$ subject to

   i. $c + k' \leq whl + k(1 + r) - T_j(whl, c, kr; \bar{\tau}) + Tr$ and $k' \geq 0$

   ii. $h' = H(h, s, z', a)$ and $0 \leq l + s \leq 1$
3. Compute (K', L', Tr') implied by the optimal decision rules in step 2.
4. If $K'/L' = K/L$ and $Tr' = Tr$ up to a tolerance, then stop. Otherwise, update the guess and repeat 1-3.

Comments:

Step 1: A guess for the lump-sum transfer Tr, corresponding to the top rate $\bar{\tau}$, is needed as Tr is endogenous.

Step 2: Solve DP-1 at age and ability specific grid points in $X_j^{grid}(a)$. This involves interpolating v_{j+1}. We use bi-cubic interpolation on (k', h'). To compute expectations, discretize the distribution of the shock variable z' with 7 equi-spaced log shocks lying 3 standard deviations on each side of the mean.

Step 3: Compute aggregates (K', L') as follows. First, for each initial conditions $\hat{x} = (h, a) \in X_1^{grid}$ draw $N = 1000$ random histories z^J. Use the decision rules from step 2 to compute lifetime histories. Set $E[k_j(\hat{x}, z^j)|\hat{x}] = \frac{1}{N}\sum_{n=1}^{N} k_j(\hat{x}, z_n^j)$ and $E[h_j(\hat{x}, z^j)l_j(\hat{x}, z^j)|\hat{x}] = \frac{1}{N}\sum_{n=1}^{N} h_j(\hat{x}, z_n^j)l_j(\hat{x}, z_n^j)$, where z_n^j is the n-th draw of the shock history. Compute aggregates as indicated below, where $\psi(\hat{x})$ is the probability of $\hat{x} \in X_1^{grid}$. Shock histories are fixed across all iterations in the Main Algorithm. The lump-sum transfer condition $Tr' = Tr$ holds when aggregate taxes T' implied from the computed decision rules equal $G + Tr$ computed from the benchmark model.

$$
\begin{aligned}
K' &= \sum_{\hat{x} \in X_1^{grid}} \sum_{j=1}^{J} \mu_j E[k_j(\hat{x}, z^j)|\hat{x}]\psi(\hat{x}) \\
L' &= \sum_{\hat{x} \in X_1^{grid}} \sum_{j=1}^{J} \mu_j E[h_j(\hat{x}, z^j)l_j(\hat{x}, z^j)|\hat{x}]\psi(\hat{x}) \\
T' &= \sum_{\hat{x} \in X_1^{grid}} \sum_{j=1}^{J} \mu_j E[T_j(wh_j(\hat{x}, z^j)l_j(\hat{x}, z^j), c_j(\hat{x}, z^j), rk_j(\hat{x}, z^j); \bar{\tau})|\hat{x}]\psi(\hat{x})
\end{aligned}
$$

Setting Model Parameters:

Following the discussion in section 4, some model parameters are fixed and the remaining model parameters are calibrated to minimize the sum of squares of distances between equilibrium model values and data values, subject to the equality constraints for three scalar moments that we want to match perfectly. In particular, the distances include those of the four age profiles presented in Figure 3, and of 10 $\times$ MaCurdy coefficient and 10 $\times$ mean hours. The equality constraints impose the following model values equal to data values: K/Y, Pareto statistic of earnings at the 99th percentile among all earners, and earnings level at the 99th percentile among all earners. The constrained minimization problem is first solved using a global algorithm (genetic algorithm), and then refined using several local methods (the interior-point and the active-set methods by Knitro).

The algorithm specified above is used to compute equilibria under tax reforms when all model parameters are determined. A closely-related algorithm is used to compute equilibria for given model parameters when the lump-sum transfer is zero.

The algorithm to compute the Laffer curve for the exogenous human capital model is given below. The skills process is by construction exactly the same as in the original benchmark steady-state equilibrium. An equilibrium in this model is defined in the same way as in the benchmark model with the exception that the decision problem differs.

Algorithm for Computing Equilibria in the Model with Exogenous Human Capital:

1. Given top tax rate $\bar{\tau}$, guess $(K/L, Tr)$. Calculate $w = F_2(K/L, 1)$ and $r = F_1(K/L, 1) - \delta$.
2. Solve problem DP-2 at grid points $x = (k; \bar{k}, \bar{h})$ for fixed values of ability a.

(DP-2) $v_j(k;\bar{k},\bar{h},a) = \max_{(c,l,k')} u(c, l+\bar{s}) + \beta E[v_{j+1}(k';\bar{k}',\bar{h}',a)]$ subject to

i. $c + k' \leq w\bar{h}l + k(1+r) - T_j(w\bar{h}l, c, kr; \bar{\tau}) + Tr$ and $k' \geq 0$

ii. $(\bar{h}', \bar{k}') = (H(\bar{h}, \bar{s}, z', a), k_j^*(\bar{k}, \bar{h}, a))$ and $\bar{s} = s_j^*(\bar{k}, \bar{h}, a)$.

iii. $s_j^*(\bar{k}, \bar{h}, a)$ and $k_j^*(\bar{k}, \bar{h}, a)$ are optimal decision rules solving DP-1 from the benchmark model.

iv. $0 \leq l + \bar{s} \leq 1$ and $\bar{s} = s_j^*(\bar{k}, \bar{h}, a)$.

3. Compute (K', L', Tr') implied by the optimal decision rules in step 2.

4. If $K'/L' = K/L$ and $Tr' = Tr$, then stop. Otherwise, update the guesses and repeat 1-3.

B.6 Labor Hours Regressions

We first describe the construction of the data sets underlying the results in section 4 for the regressions of the change in log hours on the change in log wages.

$$\Delta \log hours_j = \theta_0 + \theta_1 \Delta \log wage_j + \epsilon_j$$

We create a data set of pairs $(\Delta \log hours_j, \Delta \log wage_j)$ in two steps. Step 1: For each initial condition $\hat{x} = (a, h) \in X_1^{grid}$, draw $N = 1000$ lifetime shock histories. Step 2: For each $\hat{x} \in X_1^{grid}$, shock history and age j in the age range, calculate $(\Delta \log hours_j, \Delta \log wage_j)$. We run IV regressions using a two-stage-weighted-least-squares estimator. The instruments in the first stage are cubic polynomials in age and learning ability and their interactions. We use the weighted-least-squares estimator with weight $\frac{1}{N}\mu_j \psi(\hat{x})$ on an observation, where $N = 1000$, μ_j are age shares and $\psi(\hat{x})$ are probabilities of initial conditions. The results are contained in Table B1 for a number of choices of the age range and the measurement of wages.

Table B1 - Model Regression Coefficient θ_1

Panel (a)	Wage Measure	Age 25-55	Age 30-60	Age 50-60
	$wage_j = e_j/(l_j + s_j)$	0.124	0.296	0.378
	$wage_j = e_j/l_j$	0.117	0.274	0.479
	$wage_j = e_j(1-\tau_j')/l_j$	0.130	0.303	0.567
Panel (b)	**Wage Measure**	**Age 25-55**	**Age 30-60**	**Age 50-60**
	$wage_j = e_j/(l_j + s_j)$	0.360	0.458	0.377
	$wage_j = e_j/l_j$	0.482	0.503	0.489
	$wage_j = e_j(1-\tau_j')/l_j$	0.561	0.580	0.570

Note: Panel (a) puts no sample restrictions beyond those stated. Panel (b) puts the additional sample restriction that agents hold more than one quarter of mean earnings in assets. Model hours are $hours_j = l_j + s_j$. The symbol τ_j' denotes the marginal earnings tax rate.

To interpret the results in Table B1, we state a necessary condition for an interior solution to Problem P1 from section 2 and follow an analogous derivation to that in MaCurdy (1981). The intratemporal necessary condition below states that the period marginal disutility of extra time working equals the after-tax marginal compensation to work multiplied by the Lagrange multiplier on the period budget constraint. We restated this necessary condition using the functional form assumption on the period utility function from section 2. The second equation takes first differences of the log of the first equation. The third equation uses the Euler equation for asset holding to replace the change in the Lagrange multiplier with model variables and parameters.

$$
\begin{aligned}
& u_{2,j}(c_j, l_j + s_j) + \lambda_j \left[wh_j(1-\tau_j') \right] = 0 \; implies \; l_j + s_j = \left[\frac{\lambda_j \left(wh_j(1-\tau_j') \right)}{\phi} \right]^{\nu} \\
& \Delta \log(l_j + s_j) = \nu \left[\Delta \log \lambda_j \right] + \nu \Delta \log wh_j(1-\tau_j') \\
& \Delta \log(l_j + s_j) = \nu \left[-\log \beta(1 + r(1-\tau_k)) \right] + \nu \Delta \log wh_j(1-\tau_j')
\end{aligned}
$$

The last step assumes that the agent is off the corner of the borrowing constraint (i.e. $k_{j+1} > 0$) and that there is no risk. We do so for transparency. An extra Lagrange multiplier term enters the last equation when the agent is at a corner (see Domeij and Floden (2006)). When there is risk, the last equation is modified by an additive "forecast error" term (see Keane and Rogerson (2012)) where the additive term is based on a linear approximation.

The last equation above suggests that the human capital model is similar to the exogenous wage model, considered by MaCurdy (1981) and many others, in that the regression coefficient that comes from regressing a particular measure of "hours" growth on a very specific measure of "wage" growth is, at least in principle, a way of estimating the model parameter ν. This holds within the model only when the hours measure is the sum of model work time and model learning time ($hours_j = l_j + s_j$) and only when the wage measure is $wage_j = e_j(1-\tau_j')/l_j = wh_j(1-\tau_j')$. Thus, the hours measure $l_j + s_j$ on the left-hand side of the equation must differ from the hours measure l_j used to calculate the "wage" measure used on the right-hand side. Clearly, this is not consistent with the practice in the empirical literature. Thus, even if borrowing constraints, idiosyncratic risk and progressive taxation were not present, the standard regression approach in the literature does not produce an unbiased estimate of the model parameter ν when the theoretical model is the human capital model.[40]

Table B1 shows a number of regularities. First, the regression coefficient for the 25-55 age group in the first row is positive but below the value of $\nu = 0.614$ from Table 2. Second, even in row 3, where the measures for log hours and log wage changes used are the relevant ones from the perspective of theory and IV techniques are applied, the regression coefficient is still less than the value of ν. Domeij and Floden (2006) argue that in exogenous-wage models standard estimation procedures are biased downward. They demonstrate a downward bias due to borrowing constraints and approximation error of the intertemporal Euler equation.

When we include in the estimation only agents with substantial assets (more than one quarter of mean earnings), then all regression coefficients in Table B1 increase but remain below the value of ν. For example, for the 25-55 age group the estimate moves from $\theta_1 = 0.124$ to $\theta_1 = 0.360$ by restricting the sample in this way. Domeij and Floden (2006, Table 5) estimate in 1984-94 PSID data that $\theta_1 = 0.16$ with a standard error of 0.13. Their point estimate increases when the sample is restricted to include only individuals with larger liquid assets or larger total wealth. They estimate that $\theta_1 = 0.41$ when the sample includes only PSID observations with a total wealth of at least one half of yearly income.

[40] Imai and Keane (2004) argue that an important elasticity is underestimated in a human capital model with learning by doing.

B.7 Elasticity Estimates $(\epsilon_1, \epsilon_2, \epsilon_3)$: US Data and Model Data

We estimate the elasticity of the taxes on production of the top 1 percent and capital income taxes paid by the top 1 percent, $(\epsilon_{3,1}, \epsilon_{3,2})$, using system (2). System (2) modifies system (1) from section 6 by removing the series $\log Rev_{3t}$ while adding the series for taxes on production and capital income taxes paid by the top 1 percent. In theory, the elasticity ϵ_3 in section 6.5 is a revenue-weighted average of these two sub elasticities.

$$\begin{pmatrix} \log(1-\tau_t) \\ \log Rev_{2t} \\ \log TaxProd_t \\ \log CapIncomeTax_t \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & 0 \\ \gamma_1 & \gamma_2 & \gamma_3 \\ \zeta_1 & \zeta_2 & \zeta_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} \begin{pmatrix} 1 \\ 1_{\{t \geq T\}} \\ t \end{pmatrix} + \begin{pmatrix} \delta_{1t} \\ \delta_{2t} \\ \delta_{3t} \\ \delta_{4t} \end{pmatrix} \tag{2}$$

Table B2 - Revenue Elasticities: US Data 1964-2012

EQUATION	PARAMETER	(1)	(2)	(3)	(4)	(5)	(6)
$\log(1-\tau)$	α_2	0.236	0.243	0.225	0.236	0.243	0.225
		(0.028)	(0.028)	(0.026)	(0.028)	(0.028)	(0.026)
$\log Rev_2$	γ_2	0.068	0.075	0.048	0.052	0.055	0.047
		(0.046)	(0.053)	(0.067)	(0.036)	(0.043)	(0.058)
$\log TaxProd$	ζ_2	0.056	0.050	0.101	0.040	0.030	0.100
		(0.047)	(0.053)	(0.062)	(0.057)	(0.064)	(0.072)
$\log CapIncTax$	μ_2	-0.167	-0.016	-0.011	-0.183	-0.035	-0.012
		(0.245)	(0.240)	(0.309)	(0.244)	(0.244)	(0.318)
Elasticity	$\epsilon_2 = \gamma_2/\alpha_2$	0.288	0.307	0.212	0.219	0.227	0.207
		(0.191)	(0.215)	(0.297)	(0.156)	(0.179)	(0.260)
Elasticity	$\epsilon_{3,1} = \zeta_2/\alpha_2$	0.238	0.205	0.447	0.169	0.125	0.442
		(0.199)	(0.218)	(0.261)	(0.246)	(0.266)	(0.313)
Elasticity	$\epsilon_{3,2} = \mu_2/\alpha_2$	-0.706	-0.065	-0.047	-0.775	-0.145	-0.053
		(1.053)	(0.987)	(1.371)	(1.048)	(1.003)	(1.409)
	Time Dummies	None	1986-87	1986-90	None	1986-87	1986-90
	N. Obs.	41	41	41	41	41	41

Note: Columns (1)-(3) measure the dependent variables as log revenue. Columns (4)-(6) measure the dependent variables as log revenue per tax unit. Estimation and standard errors follow the methods in Table 5.

Table B2 estimates the elasticities $(\epsilon_{3,1}, \epsilon_{3,2})$. The point estimate for the elasticity $\epsilon_{3,1}$ is positive in all specifications. The point estimate of the elasticity of capital income taxes $\epsilon_{3,2}$ is negative without time dummies, but increases markedly after controlling for time dummies. Thus, the negative point estimate for ϵ_3 in Table 5 in section 6.5, for the case without time dummies, is driven entirely by a large, negative elasticity of capital income taxes paid by the top 1 percent.

Table B3 estimates all three elasticities, based on system (3) below, using the US data displayed in Figure 7 from section 6.5.

$$\begin{pmatrix} \log(1-\tau_t) \\ \log top1inc_t \\ \log Rev_{2t} \\ \log Rev_{3t} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & 0 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \\ \eta_1 & \eta_2 & \eta_3 \end{pmatrix} \begin{pmatrix} 1 \\ 1_{\{t \geq 1987\}} \\ t \end{pmatrix} + \begin{pmatrix} \delta_{1t} \\ \delta_{2t} \\ \delta_{3t} \\ \delta_{4t} \end{pmatrix} \tag{3}$$

Table B3 - US Data 1964 - 2012

EQUATION	PARAMETER	(1)	(2)	(3)	(4)	(5)	(6)
$\log(1-\tau)$	α_2	0.236	0.243	0.225	0.236	0.243	0.225
		0.028	0.028	0.026	0.028	0.028	0.026
$\log top1inc$	β_2	0.299	0.337	0.333	0.283	0.317	0.331
		0.049	0.051	0.071	0.059	0.064	0.089
$\log Rev_2$	γ_2	0.068	0.075	0.048	0.052	0.055	0.047
		0.046	0.053	0.067	0.036	0.043	0.058
$\log Rev_3$	η_2	-0.090	0.020	0.050	-0.106	0.000	0.049
		0.157	0.148	0.195	0.157	0.154	0.204
	Time Dummies	None	1986-7	1986-90	None	1986-7	1986-90
	Observations	41	41	41	41	41	41
	$\epsilon_1 = \beta_2/\alpha_2$	1.268 (.255)	1.386 (.274)	1.476 (.376)	1.198 (.304)	1.306 (.331)	1.47 (.457)
	$\epsilon_2 = \gamma_2/\alpha_2$	.288 (.191)	.307 (.215)	.212 (.297)	.219 (.156)	.227 (.179)	.207 (.26)
	$\epsilon_3 = \eta_2/\alpha_2$	-.38 (.673)	.081 (.609)	.223 (.865)	-.449 (.673)	.001 (.633)	.218 (.908)

Note: Dependent variables are measured as total income and revenue in columns (1)-(3), but are measured per tax unit in columns (4)-(6). Elasticity point estimates are ratios of regression coefficients as in Table 5 from section 6.5. Appendix B describes data construction. Estimation follows methods in Table 5.

What is new in Table B3 are the results for the income elasticity ϵ_1. The top 1 percent income elasticity estimates are larger than 1 in all specifications. The closest regression specification to that in Table B3 are the level and share regressions in Saez (2004). Saez (2004, Table 3 and 6) estimates that $\epsilon_1 = 0.71$ for the level regression and $\epsilon_1 = 0.85$ for the share regression when the time polynomial is a linear time trend for the time period 1960-2000.[41] The tax rate in Saez is the average marginal tax rate on the top 1 percent based on federal income taxes. The Mertens and Montiel-Olea (2017) tax rate series that we use is the average marginal tax rate based on federal income taxes and social security and medicare taxes. This implies that the (log) net-of-tax rate increase from the early 1980s to the present is smaller than the Saez measure. This is one reason why the estimates for ϵ_1 in Table B3 are larger than those in Saez (2004).[42] Saez, Slemrod and Giertz (2012, Table 1) estimate $\epsilon_1 = 0.82$ based on the top income share regression with a linear time trend and US data 1960-2006.

$$\log(top\ 1\ income_t) = \alpha + \epsilon \log(1-\tau_t) + poly(t) + \nu_t$$

$$\log(top\ 1\ income\ share_t) = \alpha + \epsilon \log(1-\tau_t) + poly(t) + \nu_t$$

We apply the regression methods used in Table B3 to model data in Table B4. We use the model transition path to a one-time tax reform from section 6 to construct time series for all aggregate variables and tax rates. The first model period corresponds to 1964, the last period corresponds to 2012 and the model tax rate increases permanently in 1987 from 0.422 to 0.49. Model aggregates are income and revenue per agent. Table B4 shows that the true model elasticities $(\epsilon_1, \epsilon_2, \epsilon_3) = (.317, .040, .739)$ are all underestimated. One reason for this is that

[41]The level regression in Saez (2004) and the regressions in Table B2 use the CPI-U series to deflate. The Saez (2004) level regression is based on average top 1 percent income and thus is most comparable to the estimate $\epsilon_1 = 1.198$ from column (4).

[42]Effective since 1994, the medicare tax applies to all earned income.

with a short time series the trend term captures part of the fall in income and in tax revenue that is due to the tax rate increase.

Table B4 - System Regression Based on Model Data

EQUATION	PARAMETER	(1)	(2)	(3)
$\log(1-\tau)$	α_2	-0.112	-0.112	-0.112
		0.000	0.000	0.000
log $top1inc$	β_2	-0.032	-0.031	-0.030
		0.001	0.001	0.001
$\log Rev_2$	γ_2	0.002	0.002	0.001
		0.001	0.001	0.001
$\log Rev_3$	η_2	-0.052	-0.054	-0.065
		0.007	0.006	0.004
	Time Dummies	None	1986-7	1986-90
	Observations	49	49	49
	$\epsilon_1 = \beta_2/\alpha_2$	.285 (.008)	.278 (.008)	.27 (.008)
	$\epsilon_2 = \gamma_2/\alpha_2$	-.017 (.007)	-.016 (.007)	-.007 (.007)
	$\epsilon_3 = \eta_2/\alpha_2$	.46 (.06)	.485 (.055)	.583 (.041)

B.8 Training Time: Model and NLSY Data

Figure B4 plots the mean model time investment in learning. Time investments decline with age for agents grouped by learning ability. Higher learning ability groups have larger time investments at any age. Figure B4 also plots the findings for males from National Longitudinal Survey of Youth (NLSY) data. The NLSY records an individual's Armed Forces Qualification Test (AFQT) score, based on tests conducted early in the lifetime, and records training time in training programs that are not associated with formal schooling. We add together the weekly training time in all such training programs. Undoubtedly, the training measure understates time allocated to skill acquistion. Presumably, a part of what any assistant professor, resident heart surgeon, professional musician or aspiring CEO picks up is not via time recorded in such programs. Nevertheless, training time tends to decline with age and to be larger at any age for groups with higher AFQT scores.[43]

We describe sample selection based on the NLSY data on training time and AFQT scores for males.

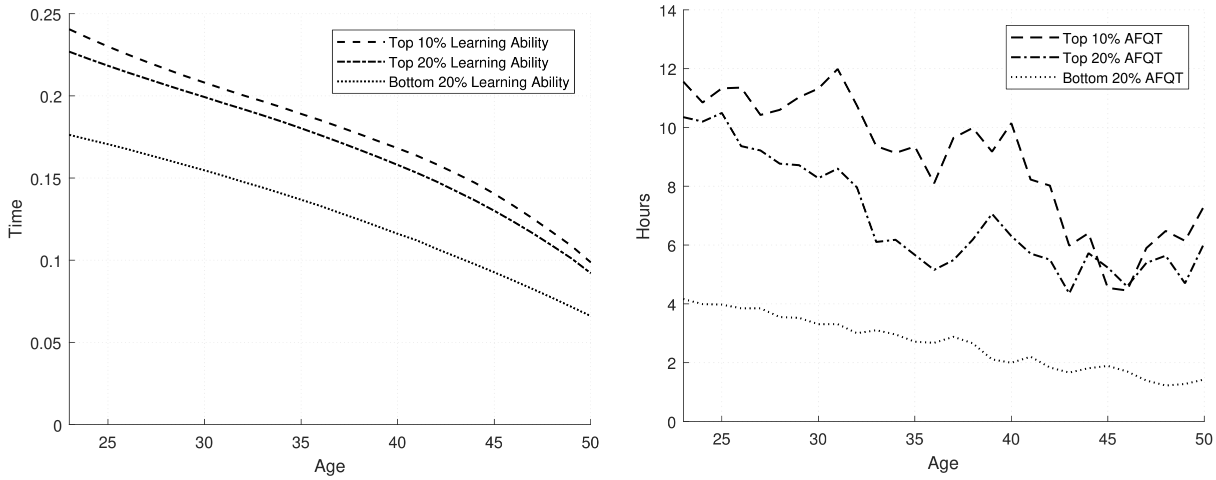
Training time: We use NLSY79 data Round 1 (1979 survey year) to Round 25 (2012 survey year) directly acquired from the BLS website. In each survey there was a series of questions following: *I would now like to ask you about OTHER TYPES of schooling and training you may have had, excluding regular schooling we have already talked about...* Then the respondent was asked about up to 4 training programs they enrolled since the previous interview, with weekly hours spent on each training program provided. We define training time of each respondent as the sum of weekly hours spent on all mentioned training programs.

Measure of learning ability: NLSY79 contains the AFQT (Armed Forces Qualification Test) Percentile

[43] Appendix B documents sample selection and how we create age-year-AFQT cells for males from NLSY data and our methods for estimating age profiles for training hours. It is well known (e.g. Heckman, Lochner and Taber (1998), Kaymak (2014) and Griffy (2017) among others) that male log earnings or log wage rate profiles in NLSY data are steeper for groups with higher AFQT scores.

Figure B4: **Training Time: Model versus NLSY79 Data**

(a) Model Training Time (b) NLSY79 Weekly Hours Trained



Note: The estimated NLSY age profiles are vertically shifted to pass through the mean training time across years for the different AFQT groups at age 45.

Score of each respondent. We use the AFQT Percentile Score as the measure of learning ability.[44]

Sample selection: We keep respondents who are male, have finished formal school education, report positive earnings, and are employed. Respondents are defined as employed if their age ≤ 30 and their annual work hours are between 260 and 5280, or if their age > 30 and their annual work hours are between 520 and 5280.

AFQT-age-year cells: We split the dataset into AFQT-age-year cells and compute the average training time across samples within each cell. We put an NLSY observation in the (q, a, y) cell if the interview was conducted during year y, the respondent's AFQT Percentile Score was in the percentile range q, and the respondent's age in year y was in the interval $[a-2, a+2]$, where ages range from $a = 23$ to $a = 50$.[45] For each percentile range q, we regress average training time on the age and year dummies. The life-cycle profiles we calculate and plot in Figure 9 from section 6.6 correspond to $(\beta_{q,23} + d_q, \beta_{q,24} + d_q, \beta_{q,25} + d_q, ..., \beta_{q,50} + d_q)$, where $\beta_{q,a}$ are the estimated age coefficients for the percentile range q, and d_q is a vertical displacement for the percentile range q selected such that the average training time at age 45 passes through the mean value across panel years at this age.

[44] We use the NLSY79 AFQT Percentile Score revised in 2006 (variable AFQT-3). We use the percentile defined for national population rather than within NLSY79 samples.

[45] We truncate our analysis at age 50 since NLSY79 drew samples aged between 14 and 22 in year 1979, and the number of samples above age 50 are small even in recent surveys. For the same reason, for each percentile range q, the data is an unbalanced panel in age and survey year. To minimize measurement error, we remove any cell that has less than 10 observations. This leaves an average cell size of 284 for the AFQT quintile range, and an average cell size of 59 for the top 10% AFQT range.

References Appendix

Alvaredo, F., Atkinson, A., Piketty, T. and E. Saez, The World Top Incomes Database, http://topincomes.g-mond.parisschoolofeconomics.eu/, accessed10/10/2013.

Attanasio, O. and L. Pistaferri (2014), Consumption Inequality Over the Last Half Century: Some Evidence Using the New PSID Consumption Measure, American Economic Review: Papers and Proceedings, vol. 104, pp. 122-26.

Domeij, D and M. Floden (2006), The labor-supply elasticity and borrowing constraints: Why estimates are biased, Review of Economic Dynamics vol. 9, pp. 242-262.

Griffy, B. (2017), Borrowing Constraints, Search and Life-Cycle Inequality, UCSB manuscript.

Guvenen, F., Ozkan, S. and J. Song (2014), The Nature of Countercyclical Income Risk, Journal of Political Economy, vol. 122, pp. 621-60.

Guvenen, F., Karahan, F., Ozkan, S. and J. Song (2015), What Do Data on Millions of US Workers Say About Labor Income Risk?, manuscript.

Heathcote, J., Perri F. and G. Violante (2010), Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006, Review of Economic Dynamics, vol. 13, pp. 15-51.

Heckman, J, Lochner, L. and C. Taber (1998), Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents, Review of Economic Dynamics, vol. 1, pp. 1-58.

Imai, S. and M. Keane (2004), Intertemporal Labor Supply and Human Capital Accumulation, International Economic Review, vol. 45, pp. 601-41.

Kaymak, B. (2014), Post-Schooling Training Investment and Employer Learning, Journal of Human Capital, vol. 8, pp. 318-49.

Keane, M. and R. Rogerson (2012), Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom, Journal of Economic Literature, vol. 50, pp. 464-76.

MaCurdy, T (1981), An Empirical Model of Labor Supply in a Life-Cycle Setting, Journal of Political Economy, vol. 89, pp. 1059- 85.

Mertens, K. and J. Montiel-Olea (2017), Marginal Tax Rates and Income: New Time Series Evidence, manuscript.

Piketty, T. and E. Saez (2003), Income Inequality in the United States, 1913-1998, Quarterly Journal of Economics, vol. 118, pp. 1-39.

Saez, E. (2004), Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications, Tax Policy and the Economy, vol. 18, pp. 117-72.

Saez, E., Slemrod, J. and S. Giertz (2012), The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review, Journal of Economic Literature, vol. 50(1), pp. 3-50.